Environnemental knowledge extraction from topo-bathymetric lidar : machine learning and deep neural networds for point clouds and waveforms

Mathilde Letard

COLLEGE ECOLOGIE
DOCTORAL GEOSCIENCES
BRETAGNE AGRONOMIE ALIMENTATION

Université de Rennes

THÈSE DE DOCTORAT DE

L'UNIVERSITE DE RENNES

ÉCOLE DOCTORALE N° 600
Écologie, Géosciences, Agronomie, Alimentation
Spécialité : *Statistiques/Modélisation en écologie, géosciences, agronomie et alimentation*

Par
Mathilde LETARD

Environmental knowledge extraction from topo-bathymetric lidar : machine learning and deep neural networks for point clouds and waveforms.

Thèse présentée et soutenue à Rennes, le 20 décembre 2023
Unité de recherche : Géosciences Rennes, CNRS, UMR 6118

Rapporteurs avant soutenance :

Clément MALLET	Senior Researcher, IGN / Université Gustave Eiffel
Jean-Stéphane BAILLY	Senior Researcher, AgroParisTech

Composition du Jury :

Présidente :	Laurence HUBERT-MOY	Professor, Université Rennes 2
Examinateurs :	Jean-Stéphane BAILLY	Senior Researcher, AgroParisTech
	Bodo BOOKHAGEN	Professor, University of Potsdam
	Laurence HUBERT-MOY	Professor, Université Rennes 2
	Lori MAGRUDER	Associate Professor, University of Texas at Austin
	Clément MALLET	Senior Researcher, IGN / Université Gustave Eiffel
	Charlotte PELLETIER	Associate Professor, Université Bretagne Sud
Dir. de thèse :	Dimitri LAGUE	Senior Researcher, Centre National de la Recherche Scientifique
Co-dir. de thèse :	Thomas CORPETTI	Senior Researcher, Centre National de la Recherche Scientifique

Remerciements

Même s'il n'y a que mon nom sur la couverture, ce travail n'aurait pas été possible sans le soutien de celles et ceux qui m'ont accompagnée ces dernières années et que j'aimerais sincèrement remercier.

Pour commencer, mes directeurs de thèse, Dimitri et Thomas. Merci pour votre confiance, votre aide et vos encouragements tout au long du chemin. Vous avez su m'accompagner et me guider à chaque étape tout en me laissant la liberté et l'autonomie d'explorer et d'exprimer mes idées. Vous m'avez montré comment aborder la recherche avec curiosité et passion, mais aussi rigueur et honnêteté. Pour tout cet apprentissage, mais aussi toutes les discussions enrichissantes et votre soutien sans faille, un grand merci. J'ai (re-)découvert ma passion pour la recherche en travaillant avec vous. Merci de m'avoir donné tous les outils dont j'ai besoin pour poursuivre ma carrière, et la confiance pour la construire à mon image.

Les mots manquent pour exprimer ma gratitude pour votre patience et le temps que vous avez consacré à m'aider dans les moments difficiles. Je ne saurai jamais comment vous remercier justement, et c'est le cas également pour Charlotte Pelletier, Marc Robin, Anne-Julia Rollet, Candide Lissak, Hélène Gloria et Ilhem. Merci de m'avoir entendue, d'avoir cru en moi et merci pour votre écoute et votre aide.

Un merci particulier à Clément Mallet et Jean-Stéphane Bailly d'avoir rapporté mon travail de thèse, ainsi qu'à Laurence Hubert-Moy d'avoir présidé le jury. Merci également à Charlotte Pelletier d'avoir partagé son expertise en traitement de séries temporelles et d'avoir fait le déplacement pour ma soutenance. Enfin, merci à Lori Magruder et Bodo Bookhagen d'avoir assisté à distance et malgré le décalage horaire à ma soutenance et d'avoir partagé leurs retours précieux sur ma recherche. J'étais très honorée d'échanger avec vous et je vous remercie tous d'avoir accepté d'examiner ma thèse.

Ce travail n'aurait également pas vu le jour sans le soutien financier du Groupe Saur et sans le programme ARED de la Région Bretagne. Je remercie particulièrement Estelle Grelier et Christophe Tanguy pour la précieuse opportunité qu'ils mont accordée.

Cette thèse a occupé trois années de ma vie, mais elle est l'aboutissement d'un bien plus long chemin. Je ne peux donc m'empêcher d'accorder une pensée particulière à toutes celles et ceux qui (parfois sans s'en douter) m'ont donné la force et la motivation de poursuivre mes objectifs.

Evidemment, merci à ma famille pour son soutien. En particulier, merci à Julie, Marie, et Matthieu : vous avez chacun, dans des moments où j'en avais particulièrement besoin, eu les bons mots pour m'encourager à continuer, et cela m'a beaucoup touchée. Merci à mes parents, par dessus tout - je ne m'étendrai pas ici car ce sont des émotions qui dépassent le pouvoir des mots, mais je suis infiniment reconnaissante et chanceuse de vous avoir comme parents, colocataires, pilliers. Merci pour votre présence toujours juste et le soutien le plus pur et le plus fort. Merci aussi à mon frère Vincent pour les moments précieux partagé et le soutien que je ressens toujours avec émotion, et à ma soeur Eléa, pour avoir été là parfaitement. Ceux qui me connaissent s'attendront également à une mention particulière pour Fifty - notre petit chat européen d'exception - et Ubert - chiot de qualité supérieure - qui ont définitivement égayé et apaisé mes journées et soirées de rédaction ou de programmation.

Merci aussi à mes amis de longue date, du collège à la thèse, en particulier Nawel et Margaux, toujours là des années après, Julie et Nicolas, que j'apprécie toujours autant retrouver, mes copains de prépa (Samy,

Elsa, JB, Victorien, Margot, Thérèse, Lisa, etc), et ma famille et mes amis de l'ENSG (mention spéciale à Marion, Juliette, Hugo, Iris, Geoff, Enola, Maylis, Olivier, et Guillemette que j'ai adoré recroiser depuis). Merci aussi à René et Francine d'avoir été là pour m'épauler pendant mes jeunes années.

Une pensée émue et reconnaissante pour ceux qui ont joué un rôle clé dans mon orientation : Monsieur Guillemot - pour le dévouement pour vos étudiants, et la passion avec laquelle vous enseignez la géographie (ce sont vos cours qui m'ont fait trouver ma voie) -, Monsieur Steegmann - pour les cours passionants, les discours enflammés, les conseils avisés et les encouragements sincères -, et Anna Cristofol et Claire Driessens pour m'avoir conseillée et aidée à construire mon parcours à l'ENSG.

Enfin, merci aux soeurs que la vie m'a données : Elsa - merci d'être mon refuge depuis plus de 10 ans, merci de partager la vie à mes côtés, même de loin -, Marion - merci pour tout ton soutien, les rires, les aventures nordiques et les merveilleuses discussions -, Manu - merci pour les weekends ensoleillés, les appels, et de toujours être à mes côtés -, Roza - merci pour le soleil, les conseils et encouragements, et pour les soirées pijama -, et Louison.

Pendant ces trois années, j'ai rencontré beaucoup de personnes qui m'ont marquée, inspirée, soutenue, consolée, amusée et encouragée. Un grand merci à vous :

- Mes camarades de bureau et d'openspace côté Géosciences. Mentions spéciales à Baptiste (pour l'accueil à mon arrivée et de m'avoir légué Citronelle), Mery (pour les super moments partagés), Thomas (merci de m'avoir aidée quand j'en avais besoin), Quentin (pour les longues discussions), Marine, Benoît, Emma, et Elias, vous qui avez toujours été là quand j'en avais besoin et qui m'avez tant écoutée et encouragée... !

- ...et côté LETG : Paul, Roberta, Gu, Carl, Flo, Quentin, Charlotte, Roza, les Théo, Julien, Marine, Alexandra, Daria, Charlotte, et plus récemment Timothée et Jean Baptiste. En particulier merci à Paul, mon partner in crime, pour tout ce qu'on a partagé. Merci à Carl pour les sessions karaoké sur l'hymne néerlandais et la BO du roi lion et les debriefs de F1. Merci à Gu, pour le réconfort, le temps GPU, les discussions sur la vie... Et merci Roberta de ne pas m'avoir laissé abandonner et de m'avoir inspirée. Merci aussi à Daria et Alexandra, deux superbes rencontres que je n'oublierai pas, et à Marine pour les sessions rédaction au café.

- Globalement, les copains du quotidien dans mes deux labos, notamment pour les Spritz du vendredi soir, les galettes au marché le vendredi midi, les matchs de foot, les soirées jeux de société, les brunchs à la campagne, ou les longues pauses café (coucou à Manu, Marion, Thomas G., Lison, Bastien, Lucas, Alex, Diane, et tous les autres !)

- Les étudiants que j'ai eu la chance d'accompagner, notamment Angéline B. et Pierre L.

- Les équipes de recherche que j'ai cotoyées. Merci pour l'accueil chaleureux et la super ambiance de travail : Caroline, Philippe, Romain pour l'accueil à la Fractory ; l'équipe Rivières dont Alain, Philippe, Laure, Boris, Clément ; les membres de LETG Rennes dont Damien, Samuel, Thomas, Anne Julia, Solène, Jean, Thibaut ; Baptiste et Paul, pour l'accueil au sein de la plateforme lidar et vos encouragements sur tous les plans ; L'équipe de Vannes, que j'ai beaucoup aimé côtoyer pendant ma dernière année.

- Les copains de conférences, dont Zoé, Roza, Iris d. G., Guillemette, Corentin, Paul, Thomas, Valentine, Iris D. Mention spéciale à Barry & Sue, qui m'ont sortie de cette grandiose impasse qu'était le Grand Canyon par 40°C !

- Celles et ceux qui m'ont aidée sur mes campagnes terrain, dont Loïc du MNHN pour les campagnes d'acquisition en mer, Hélène qui n'a jamais rechigné à m'accompagner, ma mère - Véronique - qui s'est déplacée quand j'avais besoin de mains supplémentaires, ou encore les stagiaires et visiteuses de passage au CGEL. De même merci à toutes les personnes avec qui j'ai collaboré - notamment Y. Pastol et l'équipe de AHAB - et à mes co-auteurs.

- Hélène, pour les discussions passionnantes, les balades du midi en bord de mer, les goûters, et d'avoir été une réelle amie et un soutien sans faille.

- Iris, pour nos échanges, tes conseils, les bons moments chez Roza ou en conférence, les dîners et les discussions poterie

- Celles et ceux qui ont égayé ma première année en bord de mer : Clémence et Angéline, Pirta, Allison, les collègues du MNHN, les camarades d'art plastique à l'AMAP et Gaëlle

- Enfin, merci à toutes les femmes qui m'ont inspirées et donné confiance, à celles qui se sont battues pour que ma génération vive mieux, et à celles qui m'ont fait me sentir à ma place, en particulier : Laure Guérit, Charlotte Pelletier, Anne Julia Rollet, Candide Lissak, Chloé Friguet, Marine Le Minor, Roberta Rigo, Iris de Gélis, Marion Holst, ...

J'espère ne pas avoir été trop maladroite, et ne pas avoir oublié trop de monde. Globalement, merci à tous ceux que j'ai croisés et qui ont cru en moi, m'ont soutenue, aidée ou encouragée.

Enfin, souvent, les remerciements de thèse se terminent avec une mention spéciale. La mienne s'adresse à mes soeurs : Elsa, Marion, Manu, Roza, Eléa, Louison ; à mon frère, Vincent, et à vous, Papa et Maman. A chaque moment, dans la douceur ou la tempête, vous avez été mes rocs, mes soleils. Surtout, vous avez été mes points d'appui, ceux qui m'ont rappelé, quand j'en doutais, qu'*il y a en [nous tous] de quoi soulever le monde*[1]. Et parce que je n'ai pas les mots adaptés je conclurai en empruntant quelques lignes à ceux qui savent écrire à leurs proches combien ils tiennent à eux. Papa, Maman, mon frère, mes soeurs, et mes amis les plus proches, *sans vous, [...] une épaisseur manquerait définitivement aux choses. [...] Ce que vous savez peut-être mal c'est à quel point vous êtes un besoin pour ceux qui vous aiment et, qui sans vous, ne vaudraient plus grand chose.[...] Je voudrais bien l'an prochain réduire ma vie à l'essentiel, autant que possible, et vous êtes dans cet essentiel*[1].

1. Albert Camus, René Char, *Correspondance (1946-1959)*.

Table of Contents

GENERAL INTRODUCTION

Importance of land-water interfaces

Land/water interfaces, including lake shores, watercourse banks, and coastal fringes, are a crucial part of Earth. They play key roles in natural equilibriums and cycles maintaining terrestrial systems. Indeed, coastal, estuarine, and fluvial areas shelter many ecosystems and provide ever-growing human communities with key services: food, cultural activities and protection from natural hazards emerging from rising water levels and meteorological events (Barbier et al., 2011). Therefore, they constitute a basis on which entire societies can rely and are the scene of a number of social-ecological processes at different time scales. From semi-daily tides, attracting local strollers or regional fishers, to the development of entire touristic zones over decades through seasonal activities, land-water continuum areas are constantly evolving (Syvitski et al., 2005). Water-level changes bring new landscapes at various time scales (daily, seasonally, yearly), sedimentary processes change the terrain's morphology, and sudden events like storms generate rapid changes in ecosystems (Syvitski et al., 2005).

On both sides of the shoreline, ecosystems are under the influence of marine or fluvial hydrodynamics and terrestrial hydrological processes, making coastal fringes, estuaries, and watercourses a meeting point for spatio-temporal changes. Monitoring these changes is crucial to protect such environments from accelerated natural destructive processes (Leatherman et al., 2000) as well as to ensure that these attractive areas can still sustain anthropic pressure. Indeed, in 2003, 41% of the global population lived within 100 m of the coastlines alone, which also hosted 21 of the 23 world megacities. Moreover, Martínez et al., 2007 showed that the wide diversity of ecosystems located at the interface between the Earth's oceans and continents produced 77% of the estimated economic value of the services and goods provided by ecosystems around the world in 2007, based on the method of Costanza et al., 1997.

Seagrasses, salt marshes, mangroves, macroalgae, sandy dunes, riparian vegetation, and beaches are examples of habitats of the land-water continuum. They continually interact with the water levels, and provide clear examples of how ecosystems sustain the ecological equilibrium of these areas. Seagrasses ensure water quality and are significant carbon sinks, along with salt marshes and mangroves (Barbier et al.,

2011; Turner et al., 2015). Coastal and riparian vegetation also provides protection from hydrological hazards to local communities and infrastructures, and supply many recreational activities such as snorkelling, fishing, swimming, rafting, and land sailing (Barbier et al., 2011; Turner et al., 2015). Finally, they support a wide range of endemic species by offering them nurseries, food, and oxygen (Barbier et al., 2011; Turner et al., 2015).

All of these observations converge to show how ecologically, socially and economically important land-water interface ecosystems are (Barbier et al., 2011; Costanza et al., 1997; Martínez et al., 2007). However, riverine, marine and terrestrial habitats are threatened by climate change and anthropic pressure (Barbier et al., 2011), and numerous studies agree that their evolution is difficult to anticipate and must be monitored to ensure continual support to littoral communities (Barbier et al., 2011; Costanza et al., 1997; Martínez et al., 2007).

Monitoring and protecting land-water interfaces first implies having access to tools providing relevant information to characterize them. Such tools necessarily rely on data acquisition at temporal and spatial resolution resonating with the specifics of lakes, rivers, and coasts. Currently, the uninterrupted observation of the submerged and dry sides of the waterlines remains a methodological challenge due to the presence of water (Gao, 2009, Kutser et al., 2020). The vast diversity of surface covers occurring in these areas also complicates their exploration at a time scale relevant to their fast evolution, and over large, representative extents. Remote sensing, explained below, can adequately address this issue, by providing a means of data collection without requiring direct access to fragile or remote areas and allowing faster surveying of wide portions of terrain.

Remote sensing of land-water interfaces

Remote sensing (RS) refers to an ensemble of techniques used to study from a distance the properties of artificial or natural objects based on how they interact with electromagnetic waves (Rees, 2001). Typically, these measurements are made from aircrafts, satellites or, more recently, uninhabited airborne vehicles (UAVs).

Practically, RS consists of remotely monitoring radiations reflected or emitted by areas, using a wide range of dedicated sensors (Rees, 2001). RS is now widely used for Earth observation (EO), i.e. to gather information about the different systems that constitute planet Earth. Imaging techniques, in particular, constitute a large portion of EO processes. They can be separated into two categories: passive imagery, - which captures the radiations emitted by objects -, and active imagery - which emits radiations towards objects of interest and records the way they reflect it.

RS imagery spans from 2D images to 3D models of the Earth's topography thanks to stereoscopy and Light Detection and Ranging (LiDAR, or lidar). Stereoscopy relies on a high overlap between images to find the 3D coordinates of their common points, thus allowing to derive 3D information from photographs. On the other hand, lidar is an active imaging technique that derives 3D modelling of the environment by emitting laser pulses and observing how they are reflected (Vosselman et al., 2010).

Traditionally, land-water interfaces are mainly observed with sonar or passive multispectral or hyperspectral imagery (Gao, 2009; Kutser et al., 2020), which causes several challenges to arise:

- **The need for an integrated approach:** as explained previously, ecosystems and hydrosystems populating the land-water interface are connected to both hydrological and terrestrial processes. Studying submerged areas and emerged areas along these fringes separately thus discards a large part of informative content. However, submerged and emerged areas have conventionally been studied separately. Typically, submerged areas are surveyed with acoustic waterborne techniques - sonar systems - (Barrell et al., 2015; Komatsu et al., 2003; Pasqualini et al., 1998). On the other hand, landcover remote sensing benefits from a wide range of RS imaging possibilities, from 2D aerial photographs to 3D models of the topography obtained with active imagery. Merging separate underwater and terrestrial campaigns - for example coupling sonar and passive imagery - could be a solution. However, separate surveys often have a thin overlapping area, which can be challenging to sample thoroughly with ground control points to find a coordinate transformation. Indeed, waterborne surveying is not usable everywhere: uninhabited surface vehicles cannot be deployed too far from their operating centre, and boats cannot access unsafe areas. The surveying of extremely shallow waters is thus impossible without risking grounding for both tools. On the other side, passive imagery often does not penetrate water deep enough to bridge the gap (Kutser et al., 2020). The fact that coastal surveys are split between bathymetric and topographic campaigns using different reference systems is a clear illustration of how the will to merge both domains and study them as a common, complex system is still recent.

- **The difficulty of seeing through water:** the few approaches providing seamless land-water continuum observation usually rely on satellite or aerial imagery (McKenzie et al., 2020; Mumby et al., 1997; Topouzelis et al., 2018), which covers both sides of the waterline. In particular, multispectral or hyperspectral sensors gather insightful data on shallow waters (Sandidge et al., 1998; Adler-Golden et al., 2005; Lesser et al., 2007; Klonowski et al., 2007). These data are either studied directly (Kutser et al., 2020) or used to estimate bathymetry by inverting radiative transfer models, which corrects the effects of the water column on the measured seabed reflectance (Stumpf et al., 2003). In particular, hyperspectral imagery is often used for benthic classification tasks in combi-

nation with radiative transfer models inversion (Lesser et al., 2007; Klonowski et al., 2007; Guyot et al., 2019). Satellite and aerial imagery have the advantage of covering large areas with much lower deployment costs than waterborne surveys or field expeditions by foot – especially in the case of spaceborne sensors. However, the main issue with passive imagery is the depth range in which it is usable (Kutser et al., 2020). Due to optical phenomena, past a certain depth threshold that varies with water clarity, passive imagery can no longer give information on what lies beneath the water surface because of light attenuation by water. In turbid rivers or lakes, this maximum depth can be very low, and little to no information may be derived with passive imagery. Except in very clear waters, this option consequently does not really cover the land-water interface.

- **The importance of vertical structure information:** There is a strong need for mapping approaches that provide three-dimensional structural information on ecosystems, allowing the mapping of carbon stocks, biomass production, or flow attenuation capacities (Enríquez et al., 2019; Dubayah et al., 2020; Lu et al., 2023; Mury et al., 2020). In shallow waters, studies based on passive sensors give access to the seabed covers' elevation but not to the seabed's elevation itself (Stumpf et al., 2003; Lesser et al., 2007), which would provide information on the vertical structure of these covers and enable biomass estimation or other structural assessments (Wedding et al., 2008; Lindberg et al., 2012).

- Finally, there is a need for surveying methods that could be deployed rapidly in case of natural disaster assessments.

Topo-bathymetric lidar is an active imagery sensor that effectively addresses these needs. It relies on light detection and ranging, based on laser telemetry, and exploits two sensors simultaneously (Lague et al., 2020). One of them uses near-infrared (NIR) light, while the other operates in the green spectrum, thus penetrating water surfaces (Guenther et al., 2000). Topo-bathymetric lidar is airborne, it provides 3D data at high vertical and horizontal density in shallow waters with reasonable turbidity and is quickly usable on large extents (Lague et al., 2020). Yet, it is still largely underexploited for environmental knowledge extraction about land-water interfaces.

Topo-bathymetric lidar was introduced a few decades ago, shortly after the advent of bathymetric lidar, designed specifically for coastal waters (Lague et al., 2020, Fernandez-Diaz et al., 2014). Its objectives are to provide measurements more adapted to the narrower inland water extents. It delivers one 3D point cloud per wavelength, thus producing two distinct samplings, namely in vegetated or submerged areas. It can also register the complete received energy over time for each laser shot: the full waveforms.

However, topo-bathymetric lidar data are very specific and difficult to process. The presence of water strongly influences the resulting point clouds and waveforms (Guenther, 1985) and requires the devel-

opment of new tools. Furthermore, in topo-bathymetric contexts, both NIR and green wavelength are informative. Managing to accurately process both emerged and submerged areas thus often requires exploiting both lidar datasets simultaneously, which is a consequent processing challenge.

Indeed, independently of the presence of water, **3D point clouds** pose many processing challenges. They are irregularly structured, they present many zones of occlusions and they have irregular densities (Guo et al., 2021b). These characteristics make the automatic identification of the objects they contain challenging. Furthermore, due to their three dimensions, they are harder to process. 3D point cloud processing has largely benefitted from global advances in data processing in the last decade (Guo et al., 2021b). However, these progresses have mainly concerned topographic data processing (Morsy et al., 2022), and have not fully transferred to bathymetric point clouds. The first explanation is the more recent advent of green laser-based lidar and their lower accessibility due to their power requirements. Another major explanation is the persisting separation between marine and terrestrial sciences. Most of the developed point cloud processing approaches focus on terrestrial areas such as cities or forests (Mao et al., 2022b, Huang et al., 2021; Axelsson et al., 2023; Liu et al., 2021a), which are more accessible to RS research. However, marine and terrestrial objects have distinct characteristics in shape, scale, and material. Transferring methods developed for topographic lidar to topo-bathymetric lidar is thus not trivial. Consequently, very few published research producing classified 3D point clouds from topo-bathymetric lidar exist, and even less for topographic and benthic surfaces simultaneously. 3D classifications of bathymetric point clouds including the possibility to identify a large variety of environments both submerged and emerged accurately are thus still expected.

Waveforms are also challenging to process. They are complex series of recorded power (Mallet et al., 2009) and, just as point clouds have numerous empty volumes, they often contain information on less than a third of their length, the rest being filled by noise. Their values can vary by several orders of magnitude over short time ranges due to the wide range of optical conditions encountered over a topo-bathymetric area (Guenther et al., 2000). Lastly, lidar intensity measurements depend strongly on acquisition conditions and require expert knowledge to be processed with limited bias (Kashani et al., 2015). Most of the time, lidar waveforms are consequently only exploited to detect returns from drowned or densely vegetated surfaces, where signals get too weak to be detected in flight. They are then rarely included by the constructors in the delivered datasets. Methods used to produce point clouds from waveforms are also often proprietary software belonging to lidar constructors and are mostly driven by nautical charting accuracy rather than applicative knowledge extraction maximization.

These factors make the exploitation of lidar waveforms still scarce, and even more for topo-bathymetric ecosystem monitoring. The few existing approaches attempting to exploit their knowledge either reduce them to 2D feature rasters (Collin et al., 2012) or only target one side of the water line (Hansen et al.,

2021). They also remain very specific to the area or conditions for which they were designed. For example, processing waveforms from shallow coastal areas with low turbidity is a completely different task than extracting information from turbid river surveys. Yet, while point clouds are an excellent source of geometric information (Hackel et al., 2016), waveforms provide the spectral knowledge to further characterize the surveyed areas (Mallet et al., 2009). They could thus largely benefit from the observation of land-water interfaces if more processing solutions were available.

In the end, methods able to handle the specifics of each side of the waterline exist, but few manage to deal with both simultaneously. For bathymetric waveforms processing, Xing et al., 2019 even suggest that finding one single method applicable to all conditions of depth and turbidity is still out of range for current processing possibilities. A global observation is that at this time, it is still difficult to find topo-bathymetric lidar processing approaches generalizable to the whole land-water continuum.

The main assumption of this thesis is that topo-bathymetric lidar remote sensing could significantly improve our understanding of land-water interfaces. However, methods to fully exploit the data it delivers are still expected. We thus aim to improve knowledge extraction from existing datasets to enhance land-water interface thematic modelling. To this end, we investigate how machine learning can apply to the requirements of topo-bathymetric lidar data.

Processing Earth observation data

Data processing aims to retrieve information from raw data by transforming it using different operations. In practice, it mainly consists of finding mathematical functions to predict the value of a variable depending on the observations of another. Machine learning designates methods that use empirical data to solve these mathematical problems by updating the parameters of model functions (Géron, 2022). They have brought massive progress in lidar point cloud processing, namely in identifying objects present in the data (Guo et al., 2021b). One goal of this work is to explore how existing advances can be extended to bi-spectral datasets modeling land and water without interruption.

For waveforms, we wish to assess the applicability of other processing methods that have revealed more adapted to complex problems: deep neural networks.

The potential of deep neural networks

In parallel with the development of topo-bathymetric lidars, processing methods able to solve more complex models have emerged and made significant advances in RS data processing. In particular, machine learning and deep neural networks have made significant progress thanks to the increase in computing power that characterized the last decades (Goodfellow et al., 2016).

Deep neural networks consist of interconnections of neurons organized in layers (Goodfellow et al., 2016; Lecun et al., 2015). Each neuron performs a linear combination of its inputs associated with a non-linear activation function. The connection of a potentially large number of neurons, organized depending on applications (the so-called architecture), enables the modeling of very complex functions (Cybenko, 1989). The training stage, performed by backpropagation (Lecun et al., 1989; Rumelhart et al., 1986), consists of estimating the parameters of each neuron – the weights and the bias of the linear operation.

Neural networks can adapt and generalize their learning to new inputs, making them powerful tools in scientific research and data analysis. Through this process, features of the data progressively stand out and are used to build task-adapted prediction rules (Rumelhart et al., 1986). While traditional processing methods relying on machine learning implied descriptive feature engineering, neural networks thus learn relevant features adapted to a given problem directly from the data (Goodfellow et al., 2016). They consequently eliminate the need to define descriptive features of the data upstream, contrary to classical machine learning which relies on such transformations.

Deep neural networks have brought new intakes on image processing tasks (Zhao et al., 2019; Khan et al., 2021), time series prediction (Lim et al., 2020; Wen et al., 2022), or even natural language processing (Bahdanau et al., 2014; Vaswani et al., 2017). However, they have nearly never been explored to improve bathymetric waveform processing. It is thus still complicated to evaluate whether they could change our ability to extract information from light's interactions with nature in topo-bathymetric lidar waveforms.

Objectives of this thesis

The goal of this thesis is to provide new insights into knowledge extraction from topo-bathymetric lidar surveys for the observation of coastal areas and inland waters. We wish to determine to what extent new data processing methods can contribute to this task. Our main objective is thus to bridge the gap between a favourable instrumental context - with the increasing popularity and availability of lidar data - and positive methodological circumstances - with unprecedented computing power and processing solutions variety.

Consequently, this thesis addresses the following questions:

- **How to tackle the lack of accessible tools to extract semantic information from bi-spectral lidar point clouds in complex natural areas?** Can we address the unavailability of point clouds classification methods adapted to topo-bathymetric surveys? Is it possible to fill this gap with a method that can be used without expert knowledge in computer science?

- **When waveforms are available, can we exploit them for further distinctions between numerous land and sea covers?** How many surface covers can we distinguish using waveforms only? Does the lack of spatial knowledge negatively impact resulting products? Are waveforms enough to separate surfaces with similar geometries? Is it possible to preserve generalizability without sacrificing accuracy? Can we handle acquisition-related bias without site-dependent pre-processing strategies?

- **To what extent can we deploy waveform processing for environmental knowledge extraction in (topo)-bathymetric environments?** Can they be used to approximate environmental properties even in the absence of synchronous field measurements? Is it possible to use a similar approach for different types of waters (coastal/inland, turbid/clear, very shallow/deep)?

- **How can we deal with the issue of labelled data availability considering the complexity of manually interpreting waveforms?** It is impossible to label waveforms manually without introducing significant bias for some tasks (regression and object detection mainly). Yet, we first start with supervised learning approaches to evaluate the potential of deep neural networks. In this context, how can we propose robust processing methods if no field sampling campaigns were organized during the lidar acquisition?

Organization of the manuscript

The manuscript contains five chapters. **Chapter 1** gives a detailed presentation of lidar remote sensing. It introduces airborne topographic, bathymetric, and topo-bathymetric lidar sensors and recalls the physical principles behind their operation.

Chapter 2 presents existing strategies for lidar data processing. It first recalls how learning-based data processing works and what challenges it faces. Then, it details the existing point cloud classification workflows and how they address our needs. Finally, it also explores waveform processing methods and their remaining challenges in the context of land-water interface observation.

After having introduced more in-depth the challenges and specifics of knowledge extraction from topo-bathymetric lidar data, we propose several methodological contributions to it.

In **Chapter 3**, we introduce 3DMASC, a new workflow for bi-spectral point clouds classification, and its application to topo-bathymetric environments. 3DMASC is the result of our work on the unavailability of accessible tools for complex environmental point cloud processing, namely in the presence of two point clouds. With 3DMASC, we also introduce new point clouds features that could also be used for other purposes than classification in the context of environmental knowledge extraction.

In **Chapter 4** we evaluate new possibilities for the extraction of semantic information from lidar waveforms. We wish to explore how they help to separate surfaces with similar geometry. Addressing surface type identification from waveform components also paves the way to classified data delivery straight from post-flight processing, without the intermediary step of point clouds analysis.

Finally, in **Chapter 5** we explore advanced knowledge extraction from bathymetric lidar waveforms. Namely, we propose a new method to improve lidar backscatter extraction in extremely shallow waters and in deep or turbid waters. We also experiment on the possibility to retrieve water optical parameters in one shot, without pre-processing, and without the need for in-situ measurements. To address the problem of labelled data availability, we train these networks on synthetic datasets, and explore the application of the derived models to real data through domain adaptation.

CHAPTER

1

PRINCIPLES OF TOPO-BATHYMETRIC LIDAR REMOTE SENSING

Contents

In this chapter, we introduce lidar remote sensing (RS). We expose its specificities in detail and discuss its advantages and limits for land-water interface monitoring.

- Section 1.1 is dedicated to the global concepts of lidar.
- In section 1.2, we focus on airborne ones, which are analysed in this document. To better understand why lidar is a key tool for land-water interfaces remote sensing, we present the differences between near-infrared and green lidars respectively associated with near-infrared and green wavelengths.
- Section 1.3 presents bi-spectral topo-bathymetric lidar, which combines the strengths of both wavelengths in a single sensor.
- In section 1.4, we present applications of topo-bathymetric lidar for land-water areas observation.
- Finally, section 1.5 presents the physical principles dictating topographic and bathymetric waveforms, to lay the foundations for chapters 4 and 5, in which waveform processing methods are developed.

As this first chapter solely recalls the processes lying behind lidar remote sensing, specialists of full-waveform topographic and bathymetric lidar acquisitions are invited to continue to chapter 2, reserved for lidar data processing.

1.1 Airborne Light Detection and Ranging

Before entering into details about lidar remote sensing, we recall key concepts about active remote sensing.

1.1.1 Reminders on active and passive remote sensing

Practically, RS consists of remotely monitoring radiations reflected or emitted by areas, using a wide range of dedicated sensors. RS is now widely used for Earth observation (EO), i.e. to gather information about the different systems that constitute planet Earth. Imaging techniques, in particular, constitute a large portion of EO processes, and can be applied to the study of the properties of each Earth system: the geosphere, the cryosphere, the hydrosphere, the biosphere, the atmosphere, and the anthroposphere. Imaging processes can be separated into two categories. Passive imagery, such as traditional photography, captures the radiations emitted by objects. On the other hand, active imagery emits radiations towards objects of interest and records the way they reflect it. Figure 1.1 illustrates both measurement methods.

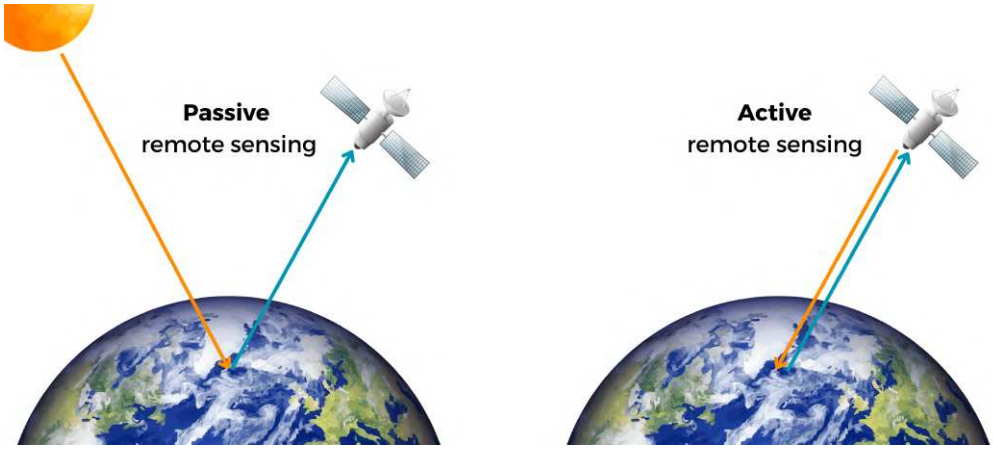


Figure 1.1: Illustration of the difference between passive and active remote sensing.

Electromagnetic (EM) waves are the vectors of this energy. They propagate through space without the need for a medium. EM waves are a form of light, they are sinusoidal and characterized by their frequency – their number of periods per second – and their wavelength – the distance they cover in one period. They span the EM spectrum from small to large wavelengths, most of them being invisible to the human eye, except for the visible part of the spectrum, which includes the colors of the rainbow, each corresponding to specific wavelengths. The amount of energy that these waves transport is inversely proportional to their wavelength, and is carried in discrete packets of energy with no mass, called photons. EM radiations thus have the characteristics of both particles and waves.

The energy transported by EM waves can be absorbed, reflected, transmitted or refracted – transmitted with a change of direction – by materials, as illustrated in Figure 1.2. Depending on the size of the material compared to the incident wave, they can also be diffused inhomogeneously in all directions. The proportion of each depends on the wavelength of the incident radiation and on the material intercepting it. Analysing the interactions of EM waves with the Earth's surface thus informs on its structure and nature.

Both passive and active techniques include sensors operating in the visible spectrum of light or outside of it, thus capturing different interactions. Passive imagery sensors measure continuous wavelength bands defined upstream, while active sensors often operate at a given, fixed wavelength. These sensors can observe different characteristics of EM waves: particle-like or wave-like properties. In modern cameras, for example, both are exploited: received light is filtered to isolate red, green and blue waves, whose photons are converted into electrons, further used to produce pixel information. On the other hand, lidar sensors exploiting only photon counts or radar sensors based on wave interferences exist. In the following paragraphs, we focus on lidar remote sensing, the active imaging technique used for this work.

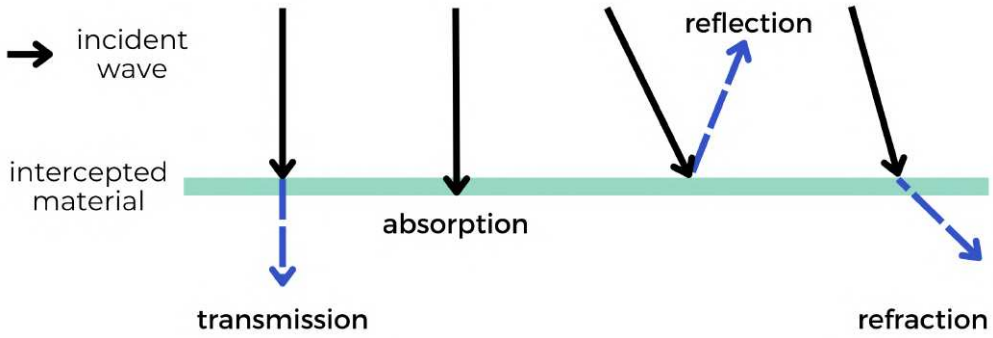


Figure 1.2: Illustration of electromagnetic wave transmission, absorption, reflection, and refraction.

1.1.2 Operating principles of lidar sensors

Light Detection And Ranging (LiDAR, or lidar) is a laser scanning technology that uses the reflection of monochromatic light on objects to derive their position in space and model scenes in 3D (Vosselman et al., 2010). When the emitted laser beam hits a material that reflects a portion of its energy toward the sensor's receiver, this backscatter is registered and used to compute the 3D coordinates of the hit material. In practice, lidars can thus image any object with the capacity to reflect its energy. It must be noted that the receiver records only the portion of the incident laser that is reflected in its direction.

Lidar sensors embed two main optical components: a laser emitter and a laser receptor. The time taken by the emitted wave to travel from the emitter to the object, and from the object back to the receptor is used in combination with the position of the emitter and receptor to derive the 3D position of the object. Two main measurement are thus made by the sensor: range measurement, and digitized received power measurements.

Several surfaces can be intercepted by the laser beam, thus resulting in multiple 3D positions. Laser scanners are able to detect multiple returns originating from a single emitted pulse, up to 15, although the first two returns concentrate around 90% of the total reflected signal (Mallet et al., 2009). Consequently, on ground or concrete buildings, one backscatter is typically received, but in porous covers, several returns may be recorded. All these targets reflect the incident beam differently depending on their geometry and optical properties.

Lidar sensors can be used from a variety of platforms: planes for airborne laser scanning (ALS) (Puech et al., 2014; Yan et al., 2015), tripods for terrestrial laser scanning (TLS) (Liang et al., 2016), or mobile vehicles or surveyors (Kukko et al., 2012), resulting in mobile laser scanning (MLS). Satellite lidar sensors also exist: NASA's ICESat (Duncanson et al., 2020; Zwally et al., 2002) and GEDI (Duncanson et al., 2020; Lang et al., 2022) are two missions embedding green and infrared laser scanners, respectively. More

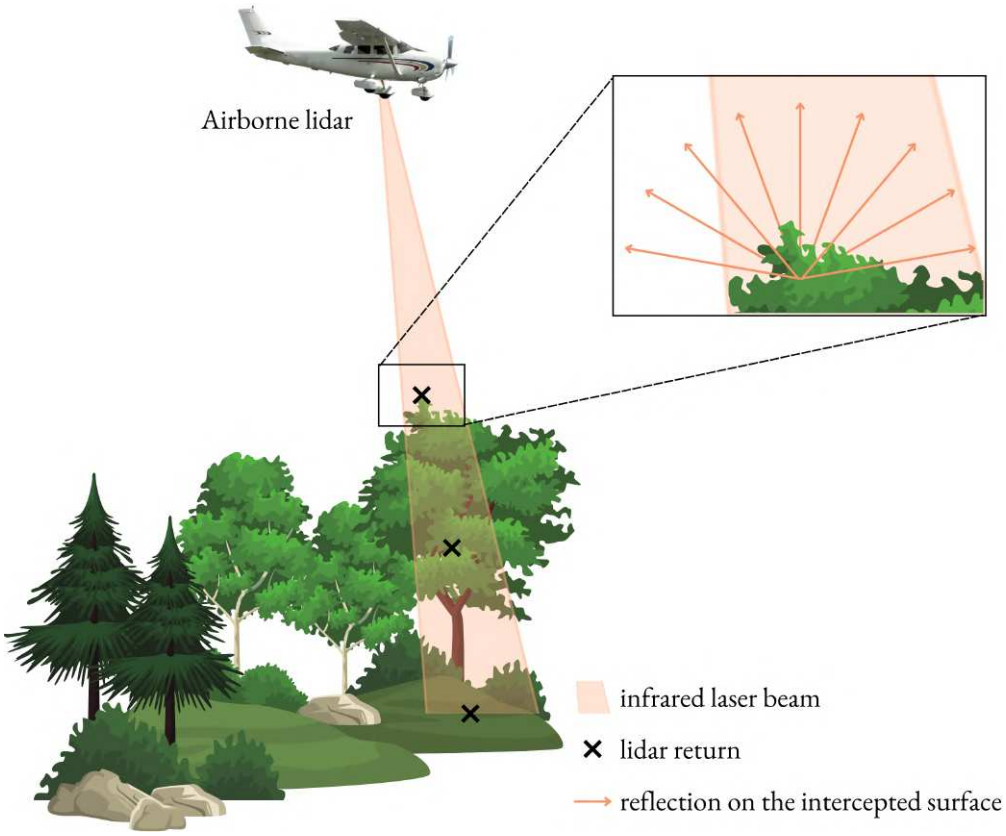


Figure 1.3: Working principle of Light Detection and Ranging.

recently, lightweight lidar sensors have become available for uninhabited aircraft, increasing the accessibility of ALS (Mandlburger et al., 2020, Mandlburger et al., 2016, Mano et al., 2020).

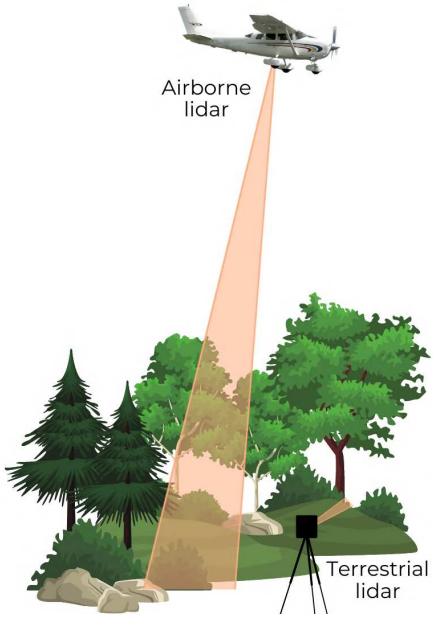


Figure 1.4: Illustration of terrestrial and airborne lidar surveys.

Most lidar sensors use pulsed light with a high emission frequency to maximize the amount of objects reached. These laser pulses have a pre-defined length, typically of a few nanoseconds (Guenther et al., 2000; Vosselman et al., 2010), and are directed towards the objects to study by rotating mirrors, that allow lidar sensors to scan complete surfaces from a unique point of view by redirecting the emitted light in varying directions (Vallet, 2011; Vosselman et al., 2010; Baltsavias, 1999). Other systems rely on continuous light waves and their phase shift to derive target positions. However, they can only handle shorter distances because of the weaker laser power.

A laser beam tends to diverge as it travels through space, thus forming a cone (Mallet et al., 2009; Vosselman et al., 2010; Guenther, 1985). The larger the diameter of the cone, the more diluted the energy across the section of surface intercepted. This diameter depends on the divergence angle of the laser pulse, which can be enlarged for safety reasons or to maximize the surface cross sections, or kept narrow (Vosselman et al., 2010; Guenther et al., 2000; Baltsavias, 1999).

Lidar sensors based on pulsed lasers are also referred to as time-of-flight laser scanners, as they use the time taken by the punctual laser pulses to travel back to the sensor to compute the position of the hit target (Vosselman et al., 2010). The time elapsed is measured with a precise clock. This method provides measurements with accuracies ranging from millimeter to decimeter scales, depending on the distance between the sensor and the target, and the sensor specificities. The largest uncertainty and factor of impact

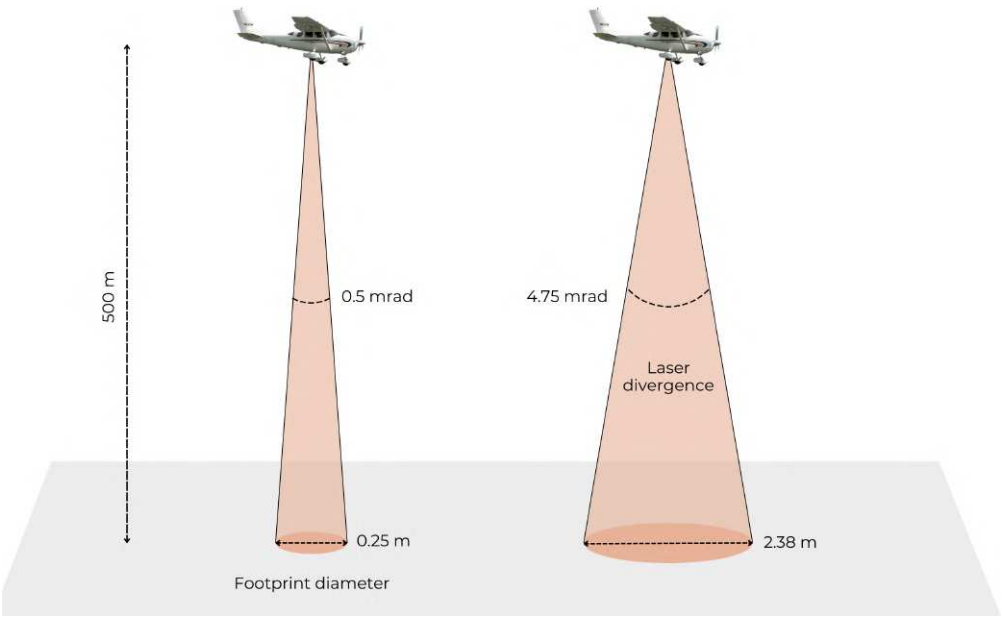


Figure 1.5: The difference between small and large laser divergence, and the resulting laser footprints.

on the measurement is the precision of the sensor's measured position (Guenther et al., 2000; Vosselman et al., 2010). More details on measurement uncertainty sources can be found in Baltsavias, 1999.

To determine the position of the hit targets, the shooting angle of the laser pulse and the position of the sensor must be known for each emitted pulse (Vosselman et al., 2010; Guenther et al., 2000; Baltsavias, 1999). The position of the sensor is measured using the position of its support, determined using a global positioning system (GPS). When the sensor is mounted on a vehicle, an inertial measurement unit (IMU) is used in addition to the GPS to measure its attitude. The traveling direction of the laser pulse is then obtained through precise measurements of the rotating mirror's angle at the time of firing of the laser shot (Guenther et al., 2000; Vallet, 2011; Vosselman et al., 2010; Baltsavias, 1999). This results in two vectors: the vehicle's displacement vector, and the laser pulse's direction vector. Using these and the GPS position, the coordinates of the intercepted surfaces can be determined using the time range and the speed of light in the considered medium (Vosselman et al., 2010; Vallet, 2011; Baltsavias, 1999).

Two procedures exist to detect the interception of a surface by the laser beam (Kashani et al., 2015). The first consists of monitoring the amount of energy received in real time and interpreting a sudden increase as a backscatter. The second consists of recording the complete amount of energy captured by the receiver over time. This time series of digitized received power is the full waveform. Each intercepted surface forms a peak in the waveform, as it causes an energy increase in the direction of the receiver. The post-flight re-analysis of waveforms is another option to determine when objects are intercepted. Both

procedures allow to deduce the range using the following equation:

$$d = c \times \frac{T}{2} \tag{1.1}$$

with d the range, c the speed of light in the considered medium, and T the time elapsed between the emission of the laser pulse and the reception of the backscatter.

Apart from the technical procedure used to detect the presence of an object in the illuminated cone, the physical possibility of sensing them is linked to two factors: the characteristics of the emitted laser pulse, and the **field of view** (FOV) of the sensor's receiver (Baltsavias, 1999). The FOV refers to the portion of space seen by the receiver. The amount of energy backscattered in the direction of the sensor and in the FOV determines the possibility of capturing the object reflecting it. The **emitted laser pulse**, on the other hand, mostly impacts the vertical discrimination abilities of the sensor, i.e. the minimum distance at which two objects can be separated (Baltsavias, 1999). In practice, the collected backscatter is the result of a convolution operation between the emitted pulse and the intercepted surface's response (Abdallah et al., 2012; Guenther et al., 2000; Jutzi et al., 2006; Baltsavias, 1999). The pulse is more or less asymmetrical and characterized by a specific width and amplitude. When it is convolved with two surface responses, its width determines if both responses will be mixed in a single peak or if they will remain separable (Jutzi et al., 2006). Though more details are given about this in Section 1.5, it is important to note that all lidar sensors have a minimum separation ability, that is physically impossible to reduce to 0 (Guenther et al., 2000; Baltsavias, 1999). As an illustration, Leica Geosystems sensors have a mean vertical separation of 50 cm.

Additionally to the X, Y, and Z coordinates of all backscatters, the GPS Time at which they were generated is recorded, along with the total number of returns captured with each pulse, and the rank among those of the point considered (Vosselman et al., 2010). The scan angle used to fire the laser beam, a potential classification flag, and **the received intensity** are saved too. However, caution is required regarding the interpretation of this attribute, as it is not an absolute measurement of the reflectance of the target (Kashani et al., 2015). Indeed, although intensity depends on the optical parameters of the surface intercepted, it is also greatly influenced by the sensor's parameters, and sometimes also by the processing used to generate its value, namely its measuring method (Kashani et al., 2015). For example, the recorded value is not necessarily the amount of energy sensed by the receiver at the exact same time position of the target as detailed in (Kashani et al., 2015). Precisions are given in Sections 1.1.3, 1.2, and 1.3.

1.1.3 Airborne lidar

Further on, ALS is presented in depth, as this thesis analyses mainly ALS data.

ALS sensors are characterized by their brooming mechanisms, which have specific ranges of rotation angles, and thus varying swaths (Vosselman et al., 2010; Gatziolis et al., 2008; Baltsavias, 1999). They can also have different **scanning patterns** depending on how the mirrors move to redirect the laser beams (Baltsavias, 1999). Typical scanning patterns include oblique parallel lines, ellipsoids, and roses-like shapes, and impact the repartition of the laser shots on the ground (Gatziolis et al., 2008; Vosselman et al., 2010). Lidars also differ in their **laser footprint**, which depends on the **divergence angle** and the above-ground elevation of the flight (Gatziolis et al., 2008; Vosselman et al., 2010; Baltsavias, 1999). A large footprint sensor benefits from a higher probability of penetrating through dense canopies, but has a smaller point density and resolution on the ground (Jakubowski et al., 2013). These differences result in surface samplings that differ between sensors, and even between surveys. The combination of the laser divergence, the scanning pattern and the laser power results in uneven point densities over space and surveys.

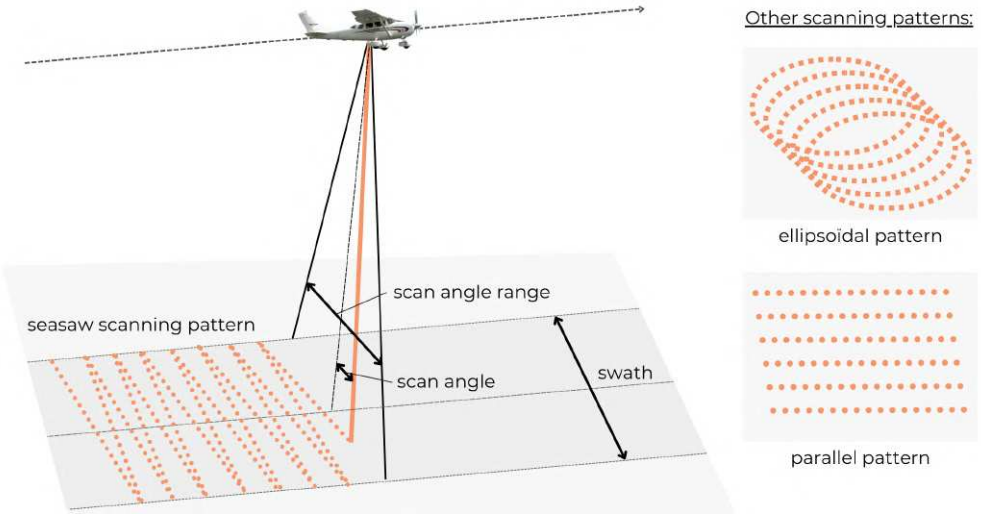


Figure 1.6: The impact of the scanning pattern and the scan angle range on the surveying possibilities.

In airborne settings, surveys are made following a flight plan incorporating multiple flight lines that overlap. Overlapping is necessary to ensure that there will be no holes in the survey's coverage even if the attitude of the plane impacts the sensor's swath punctually (Gatziolis et al., 2008; Guenther et al., 2000; Vosselman et al., 2010; Baltsavias, 1999). It also allows the possibility to correctly calibrate point clouds

obtained with different flight lines to ensure low positioning errors and compensate for bias linked to higher incidence angles (Vosselman et al., 2010; Guenther et al., 2000). Indeed, range measurement is less precise at a higher incidence angle, due to the lower portion of energy backscattered towards the receiver. Large incidence angles also impact the point pattern: density is often increased due to a slowdown of the rotation mechanism at higher angles, causing more laser shots to be emitted in these directions. Some scanning patterns also natively imply higher densities at larger incidence angles.

1.1.4 Lidar data: point clouds and waveforms

Lidar sensors can deliver data in two forms:

- full-waveforms: time series of received power for each laser shot;
- 3D point clouds, generated in flight, or from waveform analysis;

"**Full-waveform**", "**backscattered signal**", and "**time series** of received power" can all be used to refer to the lidar waveform. Indeed, in practice, waveforms are a specific type of signal, and take the form of series of values varying over time.

3D point clouds are obtained by grouping the 3D points captured by all the laser shots fired during a survey. These points correspond to peaks generated in the waveforms by intercepted surfaces. As one single shot can produce up to 15 returns, and hundreds of pulses are emitted in one second, dozens of points per m^2 can be obtained, resulting in dense point clouds describing the relief. On the other hand, one full waveform, i.e. hundreds of intensity samples, is recorded for each emitted pulse, or for every one or two emitted pulses depending on data volume storage capacities. Figure 1.7 illustrates the products obtained with one single shot, and a dozen of emitted pulses, respectively.

In the case of **point clouds**, each point is characterized by its cartesian coordinates, the additional lidar attributes (number of returns, return number, intensity, GPS Time, scan angle, classification flag), and any other attribute generated independently, resulting in an n-dimensional representation. In these clouds, the distances between the points vary, which causes assumptions that are valid about gridded data not to be applicable to point clouds.

In airborne lidar point clouds, the order of the points depends on varying factors such as the aircraft's flight plan, the brooming mechanism, the characteristics of the laser beam emitted, and the surface geometry (Vallet, 2011; Baltsavias, 1999). The local point repartition and point density depend on the targets hit, but also on the ability of the laser beam at the time to intercept them (Baltsavias, 1999). Consequently, point repartition and point density do not only depend on the physical characteristics of the scene, but

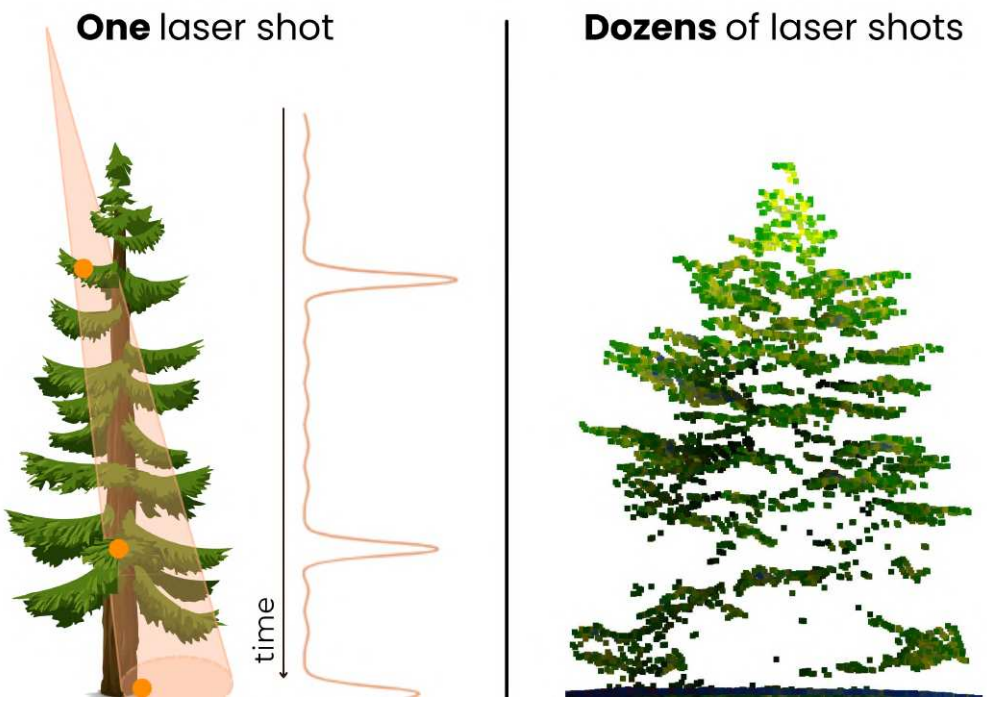


Figure 1.7: The different data delivered by lidar sensors: waveforms (left) and point clouds (right).

also greatly on the sensor's shot frequency, range determination procedure, mirror rotation mechanism, receptor field of view, size of the laser beam on the intercepted cross-section, and shot incidence angle (Baltsavias, 1999). Because of the combination of all of these factors, it is impossible to have two points at exactly identical positions in space. Surveying the same scene simultaneously with two identical sensors will thus never result in strictly similar samplings of the 3D scene, also because even if the laser shots are fired simultaneously, it is impossible to replicate exactly the same sensor trajectory.

Lidar point clouds also have the particularity of not covering the complete volume of the objects they represent. Although they have 3D coordinates, they are sometimes referred to as 2.5D data, since they do not capture all the faces of the 3D polygons that they model, and tend to represent the upward face of the 3D volumes, their sides – at best – but not their base, in the case of airborne lidar.

The **full-waveforms** can have a fixed, constant length, or an adaptative duration, automatically optimized in flight depending on the energy received to reduce the volume of data to save. Waveforms are characterized by their time sampling interval, which is different depending on the sensors, and by their received power units, which can be artificially offset to be constantly positive. Most of the time, when full waveforms are recorded, the emitted laser impulses associated with each of them are also saved separately.

Although lidar data consist of 3D point clouds, potentially with associated full-waveforms, they can

be converted to gridded data through a rasterization process. This is a popular way of visualizing or processing lidar surveys more easily than by analyzing the point clouds, which require adapted methods, as explained in Chapter 2. **Rasterization** consists in aggregating spatially the characteristics of the points into squares of fixed size that will constitute the pixels of an image. Depending on the goal, the minimum or average elevation or intensity can be used, resulting in different products. Other rasterization methods derive more knowledge from the survey by analyzing the geometrical characteristics of the point clouds before aggregating them into pixels, rather than simply using the raw point cloud characteristics (Guiotte et al., 2020). Overall, although rasterization is a popular way to process lidar data, it leads to unavoidable loss of information, as the inherent structure of point clouds contributes to information about the surveyed surfaces.

Another possibility is to convert uneven, irregular point clouds into **voxels**, by aggregating information into 3D cubes of fixed volume (Popescu et al., 2008), which is a compromise between 2D rasters – and the implied loss of spatial information – and 3D point clouds – and the implied computation complexity linked to their sparsity –, although voxelization also alters the native point repartition of point clouds.

Globally, lidar offers the possibility to image Earth's surfaces in 3D, providing additional information than those offered by traditional imagery, whether passive or active. The ability of lidar to penetrate through several layers of surface covers – for example in vegetated or submerged areas – allows to study these different layers, their spatial repartition and vertical structures (Nayegandhi et al., 2009), whereas passive imagery only gives access to the top of the surfaces. In the case of land-water interfaces, lidar offers significant advantages by allowing to image the seafloor. In practice, the use of lidar sensors in geosciences has been widely explored since their introduction. Examples of application of this technology to natural environment observation include:

- Monitoring and mapping of **landslides** (Jaboyedoff et al., 2012, Bernard et al., 2021) and **Earth surface processes** - erosion, active volcanoes monitoring, glacier mass balance estimation - (Eitel et al., 2016)

- **Forest extent and structure** (crown or trunk diameters and heights), canopy structure, tree species, tree health, and forest biomass mapping (Axelsson et al., 2023; Bye et al., 2017; Cao et al., 2016; Dai et al., 2018; Hamraz et al., 2019; Liu et al., 2021a; Nie et al., 2017; Noordermeer et al., 2023; Reitberger et al., 2009; Richardson et al., 2011; Scheeres et al., 2023)

- **Structure and infrastructure** (railways, dikes, roads, electrical networks) **detection and mapping**, including archaeological remains (paleostructures, ancient fisheries)

- **Benthic habitats or coastal zones mapping** (dunes, salt marshes, coral reefs, mangroves, macroalgae) (Grande et al., 2009; Janowski et al., 2022; Launeau et al., 2018; Nayegandhi et al., 2006; Tulldahl et al., 2012; Wang et al., 2007; Zavalas et al., 2014)

- **Fluvial areas monitoring** (Kinzel et al., 2013; Lague et al., 2020; Laslier et al., 2019; Mandlburger et al., 2015; McKean et al., 2009; Pan et al., 2015)

- **Land-cover classification** and map updating (Ekhtari et al., 2018; Matikainen et al., 2017)

1.2 Single wavelength topographic and bathymetric lidars

Most lidar sensors only use a unique laser wavelength, and are thus referred to as **single-wavelength lidars**. Common wavelengths are within the infrared, ultraviolet or green ranges. In the following sections, only infrared and green lidars are discussed, as ultra-violet lidars are mostly used to study the atmosphere. The laser wavelengths used are chosen in order to maximise the potential backscattered energy and to minimize absorption or reflection by elements that are not the user's subject of study.

1.2.1 Airborne topographic lidar

Near Infrared (NIR) lidars are commonly referred to as **topographic** lidars since they were designed to improve topographic surveys. Depending on the use cases, they may use different wavelengths, among which 1064 nm is a common value, along with 1550 nm and 1560 nm (Vosselman et al., 2010). Topographic lidars typically have low divergence – and thus small footprints – and high shot densities, leading to high spatial resolutions, and point densities above 30 pts/m^2 (Vosselman et al., 2010; Lague et al., 2020).

These sensors were the first laser scanning tools developed, and thus benefit from a longer experience of use, leading to improved capacities, a great variety of settings and use cases, smaller and cheaper options, and extensive processing possibilities. They are the most commonly used lidar and have recently been popular as high-tech additions to vehicles, phones or tablets, and are now easy enough to transport to be left on location for real-time geoscientific monitoring (Anders et al., 2020; Anders et al., 2019) or UAV-borne sensing (Scheeres et al., 2023).

As infrared light is fully reflected and absorbed by water, infrared lidars are only suitable for emerged areas (Vosselman et al., 2010; Guenther et al., 2000; Lague et al., 2020). However, in some cases, it remains impossible for airborne lidars to reach the ground. This is particularly the case in rugged relief, and in densely vegetated areas (Jakubowski et al., 2013; Vallet, 2011). In the latter, most of the incident light is reflected by the canopy, leaving too little energy to penetrate further downward. In uneven relief, due

to the incidence angles, some areas of the terrain remain inaccessible to ALS, as the illustration below, inspired from (Vallet, 2011) shows.

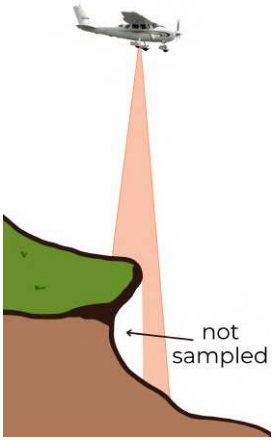


Figure 1.8: The problem of unreachable areas in uneven terrain surveyed with airborne lidar.

Topographic lidars are very popular for forestry applications (Gatziolis et al., 2008), as vegetation interactions with NIR light give information about the phenological state, chlorophyll concentration, and water content of the plants (Korpela et al., 2023). The analysis of the backscattered intensities can thus be very useful, in combination with the geometrical assessments enabled by lidar point clouds, to perform assessments of forest compositions and health. However, lidar intensity is a very site- and acquisition-dependent measurement. Kashani et al., 2015 investigated the parameters influencing the received intensity and separated the factors into four main categories:

- Target surface characteristics
- Data acquisition geometry
- Instrumental effects
- Environmental effects

Table 1.1, reproduced from (Kashani et al., 2015) gives a summary of the different ways in which these elements influence lidar intensity. For more details and references, the reader is invited to refer to (Kashani et al., 2015).

An example of topographic lidar and its characteristics is given in Section 1.3.2.

Category	Factor	Description
Target surface characteristics	Reflectance	By definition, surfaces of higher reflectance will reflect a greater portion of the incident laser radiation, thereby increasing the received signal power. In radiometric calibration, this is typically the parameter of interest.
	Roughness	Surface roughness dictates the type of reflection.
Acquisition geometry	Range	The emitted pulse energy decays as a function of range.
	Angle of incidence	Greater angles of incidence typically result in less of the incident laser energy being backscattered in the direction of the receiver, thereby reducing received optical power. Additionally, when the laser beam strikes a surface obliquely, it increases the backscattering cross-section.
	Multiple returns	When a single laser pulse reflects from objects, a correction can be applied to compensate for the energy split between objects.
Instrumental effects	Transmitted energy	The amount of energy backscattered from targets is related to the amount of energy transmitted with every pulse. Transmitted pulse energy is related to peak transmitted power (which varies with pulse repetition frequency in many systems) and transmit pulse width.
	Intensity bit-depth and scaling	Different scanners use varying bit depth (e.g., 8-bit, 12-bit or 16-bit) when digitizing the return signal. Recorded digital numbers are typically scaled to fit the available dynamic range.
	Amplifier for low reflective surfaces	Some scanners amplify the intensity values measured on low reflective surfaces.
	Automatic gain control	Some systems employ automatic gain control, which increases the dynamic range that can be accommodated but can also result in discontinuities in the intensity signal.
	Brightness reducer for near distances	Some scanners reduce the intensity values measured on close objects (e.g., less than 10 m distance).
	Aperture size	A larger aperture admits more light, increasing received signal strength.
Environmental effects	Atmospheric transmittance	Radiant energy attenuates in propagating through the atmosphere, as a function of humidity, temperature, pressure and other variables.
	Wetness	Wet surfaces also absorb more energy from the pulse, resulting in weaker returns.

Table 1.1: Factors influencing the intensity parameter in lidar data.

1.2.2 Airborne bathymetric lidar

Bathymetric lidars are the result of a later development of laser scanners for hydrographic charting (Guenther et al., 2000; Pastol, 2011; Philpot, 2019; Wozencraft et al., 2005). Contrary to topographic lidars, they rely on **green lasers**, exploiting their lower absorption by water - illustrated in the absorption spectrum of Figure 1.9 -, allowing them to eventually reach the bottom. Bathymetric lidars use lasers at 532 nm or 513 nm. Since they were created for large-scale hydrographic surveys, they are for now solely airborne, carried by satellites or planes, though UAV prototypes are being designed (Mano et al., 2020).

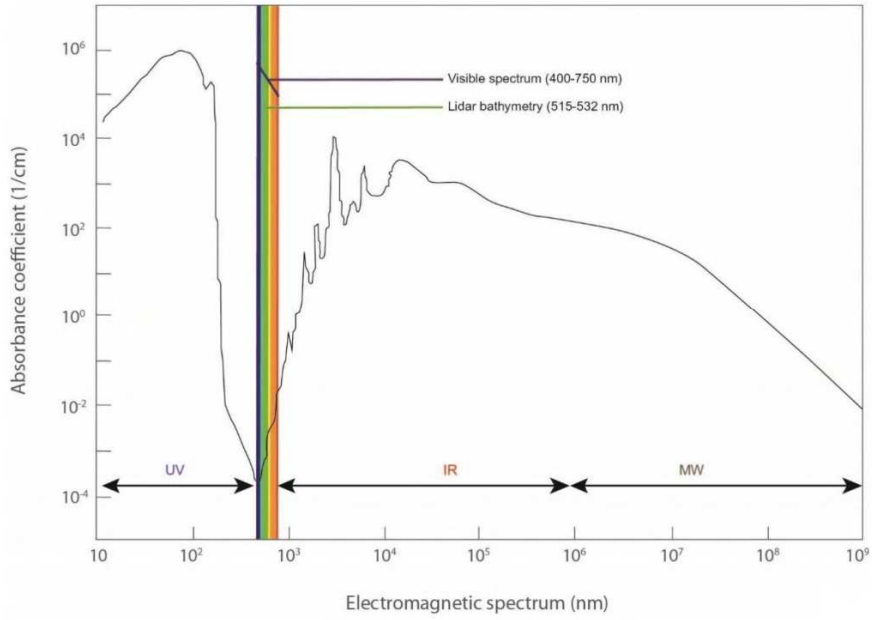


Figure 1.9: Absorption of electromagnetic waves by pure freshwater depending on their wavelength (source: Saylam et al., 2020).

When it encounters water, the green laser beam is partially reflected, absorbed, and transmitted. If the water surface has a plane, mirror-like structure, specular reflection dominates, and few energy reaches the receiver. However, if the surface is more rugous, the beam is typically mostly refracted and transmitted to the water column, where it is impacted by optical properties, before eventually being reflected by the seabed. To measure bathymetry, bathymetric lidar sensors thus exploit the time taken by the beam to travel back and forth between the water surface and the seabed.

Bathymetric lidars have to overcome several challenges linked to the effects of water on light propagation in order to produce accurate samplings of seabeds and inland water floors. The main factors dictating airborne lidar bathymetry are the following (Guenther et al., 2000; Philpot, 2019):

- **Light attenuation in water:** yellow particles, phytoplankton and sediments in suspension in the water column attenuate exponentially the amount of energy transmitted with depth, through absorption and scattering of light. The possibility to image the seabed thus depends on water turbidity, which is measured using the Diffuse Attenuation Coefficient, K, detailed in Section 1.5.

- **Beam elongation:** water also exponentially elongates – both in time and space – the laser pulse with depth, resulting in a significant increase of the laser beam's footprint when it reaches the seabed. This results in lower spatial resolution and loss of accuracy if inhomogeneous geometries are hit simultaneously. Energy is also backscattered in a wider range of directions and less concentrated. The FOV must thus be large to maximize the captured energy.

- **Light refraction:** at the surface, water deviates the laser beam's trajectory due to refraction. All ranges must thus be corrected according to the new direction and speed of the beam. This requires to locate precisely the water surface.

- **Water surface uncertainty:** the first return generated when surveying submerged areas could be mistaken for a surface backscatter. Located very close to the surface, it is the result of energy reflected from the water surface and from particles of the water column below it. The combination of both results in a single peak with a vertical offset of up to dozens of centimetres (Guenther et al., 2000; Mandlburger et al., 2013) below the actual surface depending on water clarity. As bathymetry is measured relatively to mean water levels or ellipsoids, an additional NIR channel is essential to locate the water surface and respect the standards of hydrographic charting, as it does not suffer from these effects.

- **Strongly varying intensities:** bathymetric lidars have to incorporate a receiver able to deal with variations of several orders of magnitude of received energy, considering the differences between strong water surface returns and exponentially attenuated seabed backscatters. If the dynamic range of the receiver is not adapted, echoes can be saturated – thus unexploitable for range measurements – or not detected.

As a result of these factors, airborne lidar bathymetry (ALB) always combines adapted receivers and two lasers: a NIR laser to locate the water surface and compute refraction corrections and depths, and a powerful green laser to image below water (Guenther et al., 2000; Philpot, 2019). Due to eye safety restrictions, the green laser has a higher beam divergence angle, resulting in a large footprint diameter.

ALB is thus not usable for small object detection and is limited by water clarity and specular reflection effects. Ideally, ALB acquisitions should be planned precisely to optimize the possibility of imaging the seabed depending on the presence of waves, vegetation, the tide level, or phytoplankton blooms. More

Category	Factor	Description
Acquisition geometry	Water depth	In bathymetric lidar, pulse power decays exponentially with the product of water depth and the diffuse attenuation coefficient.
	Off-nadir transmit angle	Affects the signal return due to pulse stretching and retro-reflectance of the surface material.
	Receiver field of view loss factor	Loss factor due to a receiver FOV insufficient to accommodate the spreading of the pulse in the water column.
	Altitude, refracted beam angle, effective area of receiver optics	Other acquisition geometry factors which have an effect on the return power.
	Diffuse attenuation coefficient	Light travelling through the water column is exponentially attenuated, due to absorption and scattering by particles in the water.
	Pulse stretching factor	Stretching of the pulse due to acquisition geometry and scattering properties of the water.

Table 1.2: Factors influencing the lidar intensity in bathymetric surveys.

details are given in (Guenther, 1985; Guenther et al., 2000; Lague et al., 2020; Philpot, 2019). Figure 1.10 illustrates the difference between a topographic and a bathymetric lidar survey over a coastal area.

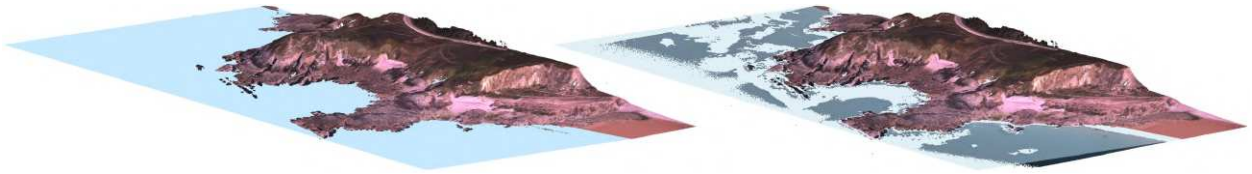


Figure 1.10: Illustration of the difference between topographic (left) and bathymetric (right) lidar surveys.

Over land, bathymetric lidars behave similarly to topographic lidars, but provide a different sampling as they have different footprints and interactions with vegetation. Similarly to topographic lidars, several parameters influence bathymetric lidar received intensity. The factors presented in Table 1.2, reused from (Kashani et al., 2015) are specific to ALB and add up to those presented in the previous section for lidars in general.

An example of bathymetric lidar and its associated characteristics is presented in Section 1.3.2.

1.3 Multi-spectral airborne lidar

Although single-wavelength lidar sensors are more common, multi-spectral sensors also exist. **Multi-spectral lidars** have the particularity to embed multiple lasers using **multiple wavelengths**. In practice, they often combine single wavelength sensors that operate simultaneously. This has the advantage of benefiting from different RS technologies in one single survey. The different wavelengths may be generated by a single power source and separated into rays of proportional wavelengths, or generated each by a dedicated source. In both cases, there are at least as many receptors and recording systems as there are lasers (Liang et al., 2016), meaning that each wavelength will produce its own individual dataset (Lague et al., 2020). Whether generated by one or multiple sources, the emitted laser beams are never co-focal. They consequently never hit the same surfaces in the exact same places, even on surfaces that react similarly. In cases where they are generated by separate sources, they also may have different parameters (in terms of footprint, divergence, power, etc). Multispectral surveys thus produce a rich sampling of the surveyed areas: they not only reach the surfaces differently but they are also reflected by the surfaces differently, resulting in several unique viewpoints of the same area.

The most common multispectral lidar sensors embed two to three lasers. These may be different infrared wavelengths for example, or infrared and green lights, as in topo-bathymetric lidars (Lague et al., 2020, Fernandez-Diaz et al., 2014). Since topo-bathymetric lidar is the main instrument used in this thesis, along with bathymetric lidar, it will be the only example of multi-spectral lidar detailed here.

1.3.1 Topo-bathymetric lidar

Topo-bathymetric lidars were specifically introduced to satisfy the needs of shallow waters and land-water interfaces surveying (Lague et al., 2020). Although bathymetric lidars embed NIR and green lasers and can thus survey both emerged and submerged zones (Guenther et al., 2000; Philpot, 2019), their green lasers are not adapted to rivers or shallow water parts of coastal bands and lakes. Indeed, to generate the higher power needed to penetrate through the water column, they use a longer laser impulse, with a larger width that results in highly overlapping echoes in shallow waters (Guenther et al., 2000; Philpot, 2019), making it impossible to distinguish between water surface/column and water bottom returns. Their large divergence resulting in wide footprints and low spatial resolution is also not adapted to the survey of narrow inland water extents (Lague et al., 2020). Topo-bathymetric lidars thus embed a topographic lidar and a green lidar with a smaller aperture – generating a smaller footprint – and sufficient power to survey shallow waters while having a short enough impulse to avoid overlapping as much as possible.

Topo-bathymetric lidars operate both lasers simultaneously and register collected backscatter separately, resulting in two individual NIR and green point clouds. As NIR and green wavelengths have

drastically different interactions with water and vegetation, the amount of information on the relief is duplicated, and particularly useful to study vegetation, land-water transitions, and shallow waters.

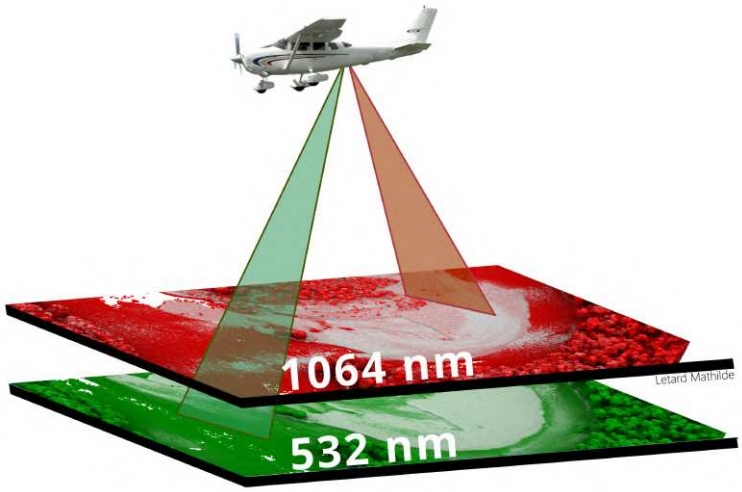


Figure 1.11: The principle of topo-bathymetric lidar.

The following section illustrates the differences between the sensor specifications of topographic, bathymetric, and topo-bathymetric lidars built by the same manufacturer.

1.3.2 Comparison of airborne lidar sensors

In this section, three lidar sensors are presented to illustrate how topographic, bathymetric, and topo-bathymetric lidars differ. We chose to use Leica Geosystems' instruments, as the company produces the three types of airborne sensors discussed. However, in this thesis, data originating from different sensors are exploited.

The parameters describing each sensor are summarized in Table 1.3. Aside from the operating wavelengths, the main distinct parameters are the scan speed and the laser divergence of the topo-bathymetric and bathymetric sensors, which result in a larger footprint size, and a lower point density, both contributing to their coarser spatial resolution. The laser divergence of 4.75 mrad of the ALB sensor illustrates the problem of energy dilution necessary for eye safety mentioned in Section 1.2.2. The characteristics of Leica's airborne topo-bathymetric system – the Chiroptera – illustrate well why such systems were introduced: although it has both green and NIR wavelengths, the Chiroptera exploits a green laser with a 4.75 mrad divergence, in between the TerrainMapper's 0.25 mrad, and the HawkEye's 7.5 mrad. In practice, at a 500 m flying height, this produces laser footprints radiuses of 1.2 m and 2.4 m for the topo-bathymetric and bathymetric lidars, respectively, and of 0.125 m for a topographic system flown at 1000

m above ground level (AGL). This provides explanations as to why ALB was not suited to observe rivers or inland water extents with small scales, as a narrow watercourse may be less large than 5 m, and thus not observable with a bathymetric lidar.

Another important difference is the operating altitude of the three lidars: topographic lidars can be used at much higher altitudes, while bathymetric and topo-bathymetric lidars are limited by the amount of power needed for the green laser, and by its high divergence, resulting in an increasingly larger footprint with flying height. Since the swath width – detailed in the above table – depends on the operating height above the ground, this means that with a similar flight plan, topo-bathymetric and bathymetric lidars cover a smaller surface than topographic sensors. As a result, the operating cost of bathymetric and topo-bathymetric lidars is significantly higher, because more flight lines are needed to cover the same areas with a reasonable overlap (Lague et al., 2020). This also explains why the increased availability of lidar data is for now still largely limited to topographic surveys.

1.4 Lidar remote sensing of land-water interfaces

Airborne topo-bathymetric lidar constitutes a reliable remote sensing approach to survey land-water interfaces, as the data it delivers have no interruption in transition areas (Guenther, 1985; Lague et al., 2020). It also allows to cover large areas quicker than what sonars allow and can survey shallow waters as well as deeper waters in adapted survey conditions (Guenther et al., 2000). Lidar provides adequate spatial resolution and precision, too, and contrary to imagery-based bathymetry, it provides the elevations of the seabed covers and of the seabed itself, making seabed covers 3D analyses accessible. Additionally, when it penetrates through multiple layers of surface covers, it opens up to vertical structure analyses, inaccessible to passive imagery. The authors of (Guenther et al., 2000) summarize the benefits of this technology as five key points:

- Possibility to perform surveys quickly over large or small areas
- Ability to survey dangerous areas impossible to access by boat
- Simultaneous collection of data both above and below the waterline
- High mobility, allowing rapid assessments of seasonal changes or storm damage
- Capacity to complete surveys in short windows of operability conditions in areas in which traditional surveys could not (ice cover for example)

In the literature, uses of lidar sensors to observe the land-water continuum include:

Sensor	Leica TerrainMapper	Leica Chiroptera-5		Leica HawkEye-5		
Type	Airborne topo. lidar	Airbone TB lidar		Airborne bathy. lidar		
Channels	Topo.	Topo.	Bathy.	Topo.	Shallow bathy.	Deep bathy.
Laser wavelength (nm)	1064	1064	515	1064	515	515
Laser divergence (mrad)	0.25	0.5	4.75	0.5	4.75	7.5
Intensity digitisation	14 bits	14 bits		14 bits		
Scan pattern	Oblique with even distribution	Oblique front-back palmer		Oblique front-back palmer		
Max. scan speed (scans/s)	300	140		170	100	
Scan angle range	20° – 40°	±14° front/back, ±20° left/right		±14° front/back, ±20° left/right		
Operation altitude	300 – 5500 m AGL	400 – 600 m AGL		400 – 600 m AGL		
Swath width	70% of AGL	70% of AGL		70% of AGL		
Min. Vertical separation	50 cm	Below 50 cm		Below 50 cm		
Vertical accuracy	Below 5 cm	Below 5 cm	Relative to depth *	Below 5 cm	Relative to depth *	
Horizontal accuracy	Below 13 cm	Below 15 cm	Relative to depth *	Below 15 cm	Relative to depth *	

Table 1.3: Sensor specifications of three Leica Geosystems lidars

- Evaluation of **habitat complexity** (Wedding et al., 2008), **morphology** (Wilson et al., 2019), or **biomass** (Webster et al., 2020)
- Seabed types and **underwater** vegetation **mapping** (Schmidt et al., 2014; Eren et al., 2018; Kumpumäki et al., 2015; Wang et al., 2007; Tulldahl et al., 2012; Zavalas et al., 2014; Pe'eri et al., 2007)
- **Land-water separation** (Hu et al., 2019; Morsy et al., 2016; Shaker et al., 2019)
- Coastal **hydrographic and nautical charting** (Pastol, 2011; Wozencraft et al., 2005)
- Intertidal and/or subtidal **ecosystems mapping** (Chust et al., 2010; Grande et al., 2009; Hansen et al., 2021; Launeau et al., 2018; Nayegandhi et al., 2006)
- **River geomorphology, bathymetry, and vegetation monitoring** (Mandlburger et al., 2015; Hilldale et al., 2008; Kinzel et al., 2013; Lague et al., 2020; Pan et al., 2015; McKean et al., 2009; Laslier et al., 2019)
- Assessment of the **consequences of natural events** (Parrish et al., 2016)
- Coastal engineering structures surveying (Irish et al., 1998)
- Tidal inlets and beaches **erosion monitoring** (Irish et al., 1998)

However, although this shows that lidar can address multiple needs in environmental sciences, it is still clear that integrated approaches of both emerged and submerged surface covers are still scarce, as most of the applications still focus on one side of the waterline. Another limit is the actual use of the 3D data: many of the aforementioned lidar-based studies exploit 2D products, or rasterize the data or the results, thus not yet realizing real 3D mappings of surface layer structures. Lastly, none of these methods exploit the information provided by the two distinct samplings offered by topo-bathymetric and bathymetric lidar – as it is always used with a NIR laser too. There is thus still a lack of processing methods allowing a global analysis, in 3D, of domains below and above the waterline, and fully exploiting the knowledge contained in lidar surveys.

1.5 Lidar waveforms

Although lidar data are more commonly studied as collections of punctual range measurements, it should be kept in mind that lidar originally consists of continuous light signals digitizing and processing. 3D coordinates obtained through lidar surveys are often the result of a reanalysis of the reflection of the light

wave emitted by the sensor (Guenther et al., 2000; Philpot, 2019; Kashani et al., 2015), as explained in Section 1.1.2. Technical progress has allowed this transformation to be made in flight in order to provide 3D measurements more directly. In such settings, the receiver only triggers recording when the number of received photons exceeds a certain threshold. However, there are obvious limitations in using a fixed threshold: some backscatters may be missed, or, on the contrary, some false positives may be recorded as useful information. A generation of lidar sensors with the capacity to digitize and record the complete light signals received after laser emission has thus emerged, along with attempts to make more of each lidar survey planned (Chauve et al., 2007; Mallet et al., 2009).

Full-waveform lidar systems record the complete backscattered signal in addition to the discrete returns. A **lidar waveform** consists of a **time series of received laser power**, that practically contains the portions of the laser beam that were reflected against various targets. Optical and mechanical tools ensure that each laser return is recorded separately and linked to its initial impulse, even though current sensors have very high measurement frequencies. The signals are digitized at constant time intervals and are either triggered and stopped at fixed time windows after photon emission or depending on the received energy. The length of a lidar waveform thus depends on the sampling capacities of the digitizer and the recording settings. The signals are expressed as digitizer photon counts as a function of time (Pirotti et al., 2011), and are relatively vertical (depending on the incidence angle) altimetric profiles.

Since the receiver cannot be completely desensitized to ambient light as it has to stay receptive to the backscatters, the received energy is never zero (Guenther et al., 2000; Abdallah et al., 2012). It is however higher when reflective targets cross the path of the emitted beam, causing the reflection of some of it towards the receiver. Thus, waveforms contain the recording of the different objects laying within the illuminated cone, noticeable in the signal as **peaks** characterized by a **range**, **width** and **amplitude** (Pirotti et al., 2011).

In practice, full-waveform lidar data contain more information than the lidar point clouds detected during acquisition. First, adapted post-flight analysis allows the retrieval of more reflection points and thus more knowledge of the scene's topography. This is particularly the case in complex areas such as urban (Mallet et al., 2011), vegetated mountainous (Chauve et al., 2009) or submerged zones (Kogut et al., 2019). Pirotti et al., 2011 quantified this advantage and wrote that full-waveform processing increased the point density by a factor of two relative to the conventional discrete-return lidar at equal mission time in forestry applications. Secondly, lidar waveforms also contain finer information on the origin of each backscatter, whether on its precise position - Kogut et al., 2019 show that waveform signal analysis allows getting more accurate seabed points altitudes – or on its nature, since the different peaks are shaped by the optical and geometrical characteristics of the hit targets (Jutzi et al., 2006; Wagner et al., 2006; Wagner, 2010). The next section explores the characteristics of lidar waveforms in detail to show to what extent

they can give indications of the physical traits of the hit surfaces.

1.5.1 Shaping factors of lidar waveforms

Mathematically, lidar waveforms can be treated as a sum of N components, with N being the number of surface elements that were intercepted by the laser pulse (Abdallah et al., 2012; Chauve et al., 2007; Pirotti et al., 2011), and a noise component corresponding to noise internal to the receiver and to the ambient light that is still captured in the background (Abdallah et al., 2012; Guenther, 1985; Guenther et al., 2000; Philpot, 2019). While studying the satellite lidar GEDI, Zwally et al., 2002 showed that if the elements on the beam path have a vertical height distribution that follows a Gaussian law, the waveform obtained can be considered as a sum of Gaussian components sometimes with an added noise bias, because the transmitted laser pulse has an approximately Gaussian distribution in time. They analyze a series of waveforms acquired over oceans, sea ice, ice sheets or land and show that plane areas result in single Gaussian waveform components, while more geometrically complex areas can produce multiple distinct or overlapping Gaussian-shaped peaks. Similarly, Wagner, 2010 states that small footprint lidar waveforms can also be well modelled with a sum of Gaussian pulses, as waveform impulses can have shapes similar to an ideal Gaussian function. A major conclusion of such analyses is that the **shape of the echoes** in the waveforms is mainly influenced by the **transmitted pulse shape** and the **height distribution** within the laser footprint, namely the roughness and slope of the surface hit by the footprint (Zwally et al., 2002). In practice, it has been formalized that the returned echoes in the waveforms are the result of a **convolution between the transmitted laser pulse function and the corresponding response function of the surface encountered**. The surface response is a function of its geometric and radiometric properties. In the following paragraphs, we present the global factors that shape lidar waveforms. Formal definitions of these dynamics are introduced in section 1.5.3.

Influence of the intercepted surface geometry

The main factors of influence of the surface geometry on the backscattered signal are **roughness, slope, and the vertical repartition of targets**. Surface slope impacts the asymmetry of the peak, as the part of the surface closest to the sensor will start reflecting earlier than the furthest (Jutzi et al., 2006). Photons will be reflected to the receiver for a longer time range, but in a progressively diminishing amount. This process is called **pulse stretching**. The impacting slope can be the actual slope of the surface, but also the relative slope induced by the incidence angle of the laser beam, which causes the same pulse stretching effect. Surface roughness influences the angle of incidence of the laser rays, the slope over the hit area and the range by generating multiple scattering (Jutzi et al., 2006). The waveform peak generated by a surface

with high rugosity will thus have a different skewness, kurtosis and number of inflexion points than a perfectly smooth and plane material, as Figure 1.12 illustrates.

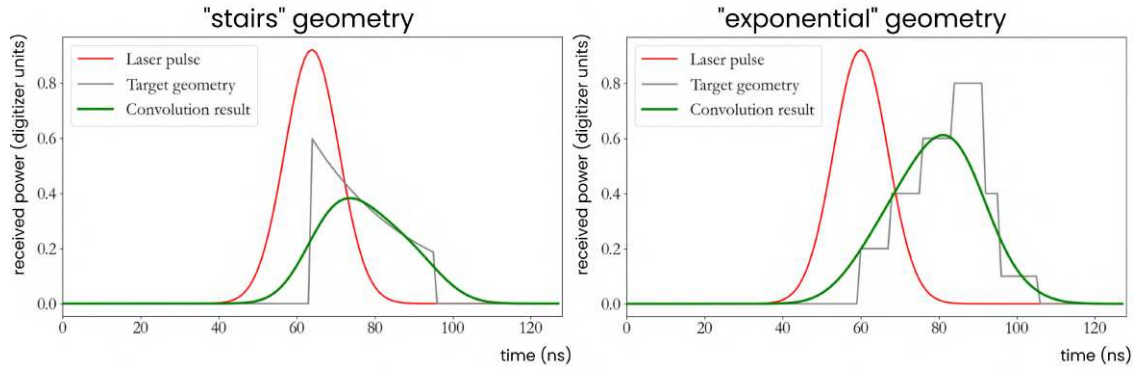


Figure 1.12: Impact of the intercepted surface's geometry on the resulting waveform shape.

A typical example of this phenomenon is vegetation. Contrary to land, trees produce less symmetric returns in waveforms, which is mostly attributed to multiple scattering and to the distribution of the scattering elements depending on the canopy shapes, foliage densities and on the slope of the ground (Chauve et al., 2007; Chauve et al., 2009).

Jutzi et al., 2006 illustrated typical waveform shapes depending on the geometry of the surface portion falling in the laser footprint, as featured in Figure 1.13:

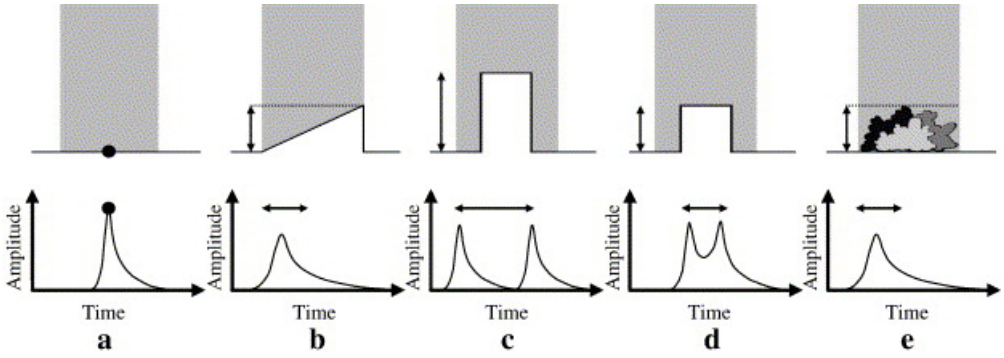


Figure 1.13: Impact of the intercepted surface's geometry on the resulting waveform shape, extract of Jutzi et al., 2006: a) Plane surface, b) sloped surface, c) two significantly different elevated areas, d) two slightly different elevated areas, e) randomly distributed small objects.

Finally, the vertical repartition of objects in the illuminated cone also influences the shape of the received waveform. As mentioned previously, each object intercepted by the laser beam reflects it in its own way, generating peaks of backscattered energy. However, if both objects are too close in reality, they

can reflect energy during overlapping time intervals, and thus be difficult to separate in the signal. Figure 1.14 illustrates how, when they span over intersecting time intervals, surface backscatters can overlap in the signal.

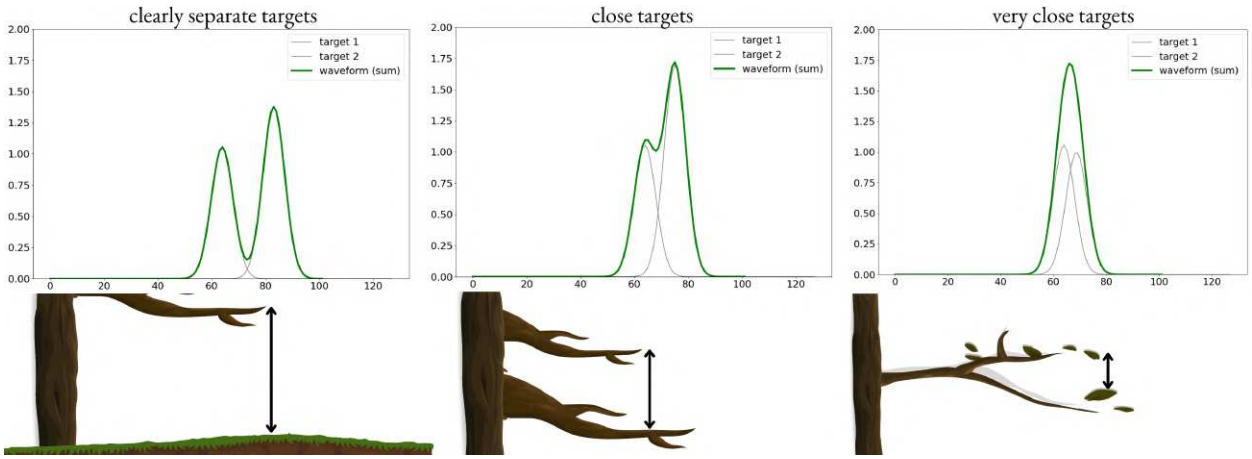


Figure 1.14: Impact of the vertical repartition of multiple intercepted targets on the resulting waveform.

It is important to note that peak shapes can be analyzed as a function of surface geometry because the time sampling remains constant in the full waveform. However, this time sampling interval may translate into a different space interval depending on the medium crossed, mostly in the case of bathymetric or topo-bathymetric lidar. Thus, depending on the signal sampling and recording frequency, the vertical resolution of waveforms captured in different surveys can vary. For example, Teledyne Optech's Titan samples once every nanosecond (a frequency of 1 GHz), which is 15 cm in air and 12 cm in water, while Hexagon's HawkEye samples once every 556 picoseconds, doubling the vertical resolution.

Influence of the transmitted laser pulse

Considering the physics of lidar remote sensing, it is globally assumed that the main factor governing waveform peak shape independently of any geometrical or spectral specificity is the sensor's **laser impulse**. Initially, it was assumed that this emitted pulse was shaped as a Gaussian function with a fixed width that could be estimated through calibration procedures (Chauve et al., 2007). Jutzi et al., 2006 thus defined the waveform as a time-delayed Gaussian modulated by multiplicative noise following a Gaussian distribution too. In this setting, the length of the pulse corresponds to the half of the pulse's maximum amplitude.

Some sensors have distorted impulses, i.e. pulses with more peaked, flattened or asymmetric shapes and may thus not be accurately depicted by Gaussians (Chauve et al., 2007; Hofton et al., 2000; Jutzi

et al., 2006). For example, Chauve et al., 2007 explain that RIEGL's LMS-Q560 has a laser pulse with a steeper ascending part and a longer but weaker descending one than a Gaussian.

In all cases, the shape of the emitted pulse contributes to the shape of the backscattered pulse, principally through two aspects: the amplitude of the impulse, and its width. As the **width of the impulse** is linked to its duration, it influences the duration during which the intercepted surfaces are illuminated, and thus for how long they reflect energy. The overlapping of backscatters explained in Section 1.5.1 is thus linked to the distance separating the targets, but also to the duration of the emitted laser pulse (Guenther et al., 2000). The larger the impulse, the more likely the targets are to overlap, as there are more chances that by the time the first target stops reflecting light, a second, close one has already been hit too. The **amplitude of the impulse**, on the other side, influences the amplitude of the backscattered echoes, as it dictates the amount of energy travelling to the surfaces and available for reflection. Both the width and amplitude of the impulse are thus somewhat transmitted to the backscatters. Mathematically, this is formalized through a convolution between the impulse function and the surface response function, depending on its geometry, as depicted previously, and its physical properties.

Influence of the physical properties of the intercepted surface

Another factor influencing the shape of the backscatter in the waveform is the optical behaviour of the surface intercepted (Chauve et al., 2009; Chauve et al., 2007; Kashani et al., 2015; Wagner, 2010). In particular, **surface reflectance** impacts the amplitude of the recorded peak, as it dictates the amount of energy that is backscattered towards the sensor. Reflectance depends on the physical and chemical characteristics of the objects, such as their water content or their chlorophyll concentration. It is thus informative on the nature of the surfaces hit. In the case of multiple returns, different amounts of energy can be reflected by the hit surfaces, resulting in a more complex waveform shape. Similarly to geometry, vegetation gives practical examples of these radiometric factors: Chauve et al., 2007; Chauve et al., 2009 show how the reflectance of each species, of the foliage, of the branches and of the ground shape the obtained peaks. Figure 1.15 illustrates the impact of surface reflectance on the resulting waveform component for identical incident pulse amplitudes.

However, the obtained amplitude values do not constitute accurate reflectance measurements, even when discarding the noise, as they greatly depend on other phenomenons, such as meteorological conditions on the day of the flight, sensor settings including receiver sensitivity and amplification, and emitted laser pulse intensity, that is not constant (Kashani et al., 2015).

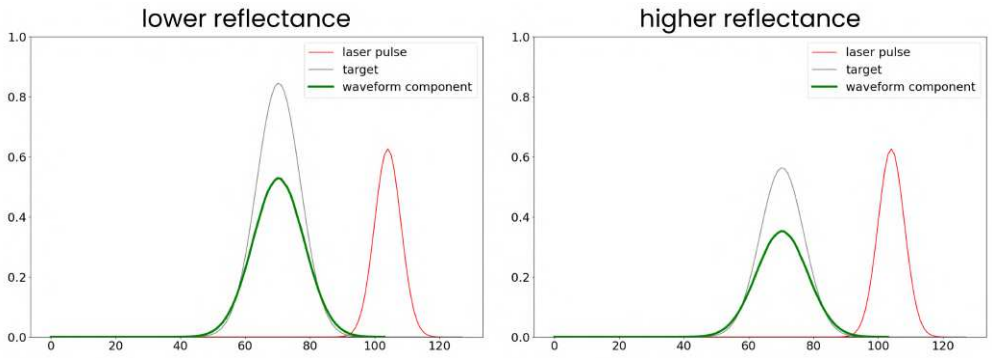


Figure 1.15: Impact of the intercepted surface's reflectance on the resulting waveform.

Instrumental effects: digitizer, FOV, ambient light (noise)

As presented in Kashani et al., 2015, the intensities measured by lidar sensors depend greatly on instrumental parameters (see Section 1.3.2). Waveform values are thus also dependent on sensor parameters such as radiometric resolution, amplification or gain control. Among instrumental effects, two can significantly impact the interpretability of waveform data: noise and FOV (Liang et al., 2022). Waveform noise depends on the **receiver's sensitivity** to ambient light and causes variations in the power values stored. When backscattered echoes are weak, they are thus harder to identify, as they can be of similar amplitude than noise. The **FOV** parameter is thus important, as it determines the amount of backscattered energy that can be intercepted by the receiver, and consequently the strength of the peak in the waveform. A larger FOV increases the amount of light that reaches the receiver, resulting in stronger backscatters, easier to separate from low amplitude noise. On the contrary, when a small FOV is used, there are greater chances that only a small portion of the backscattered energy reaches the surface of the receiver, and that surfaces with lower reflectance, or second and later returns, generate weak returns lost in noise. Figure 1.16 illustrates the FOV in bathymetric contexts (inspired from Philpot, 2019).

1.5.2 Specificities of bathymetric waveforms

Due to the different propagation of light in water, some specificities apply to waveform components obtained in the presence of water, to which we will refer in this section as **bathymetric waveforms**.

Bathymetric waveforms include three main components: the **air-water interface** component, the **water column** component, and the **water bottom** one (Guenther, 1985; Guenther et al., 2000; Philpot, 2019). They follow the same principles as classical waveform components but have specificities induced by water. Abdallah et al., 2012 formulate mathematically the mechanisms behind the shape of a simple

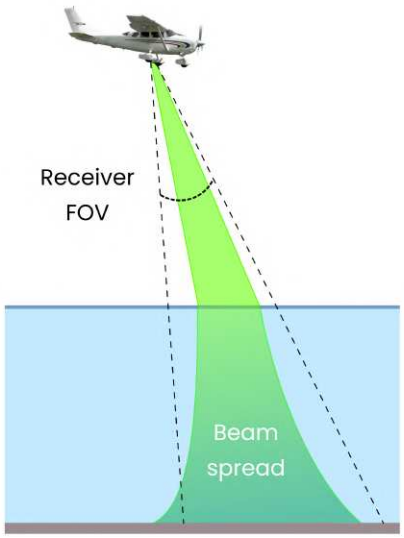


Figure 1.16: Illustration of the importance of the field of view in airborne lidar bathymetry.

bathymetric waveform, which is the sum of different components: roughly Gaussian noise, convolutions of Gaussian functions with estimated backscattered energy values, and an exponential decay of the signal after the water surface. Figure 1.17 illustrates the typical structure of a bathymetric waveform, after Guenther, 1985 and Guenther et al., 2000.

The air/water interface component

As explained by (Guenther et al., 2000) and introduced in Section 1.2.2, the first peak detectable in a bathymetric waveform cannot be considered a pure representation of the water surface. It is rather a **superposition** of energy reflected by the **air/water interface**, and particles from the **higher part of the water column**, located just below the surface (Guenther et al., 2000; Mandlburger et al., 2013). The beginnings of both components are simultaneous – when the pulse starts to hit the water – which explains why they mix in the observed peak. However, they originate from different sources and thus have different characteristics. Since the air/water interface's shape is mostly influenced by the transmitted pulse shape and the incidence angle, it has a traditional lidar backscatter shape, i.e. more or less Gaussian with a skewed character (Guenther et al., 2000). The water column component, on the other side, has a completely unique exponential signature, as detailed in the next section. The water column component has a much longer rising time than the air/water interface and thus causes the global maximum of the resulting mixed peak to be attained later than the actual maximum of the air/water interface (Guenther et al., 2000). As a matter of consequence, the resulting peak cannot be used to derive the position of the water

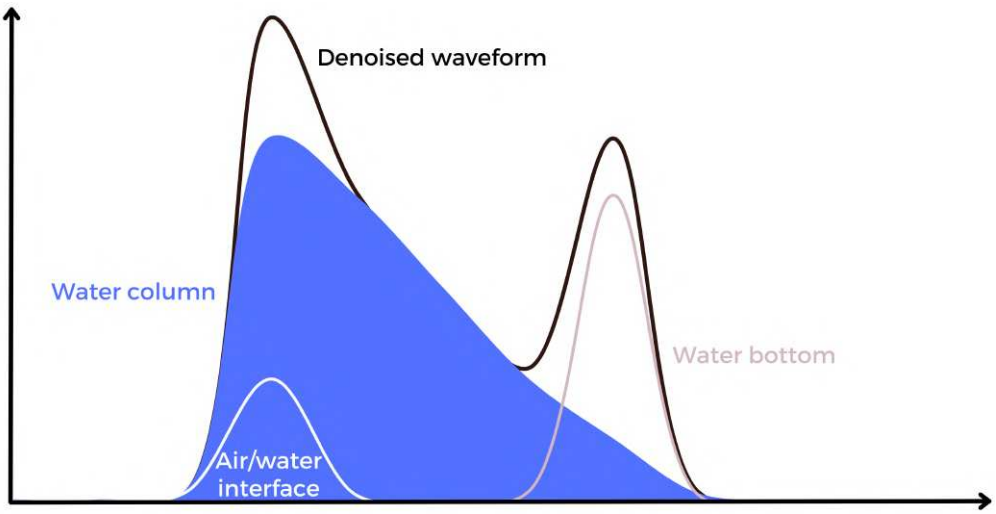


Figure 1.17: The three components of the bathymetric waveform.

surface, as it will be delayed in time, and may cause an error up to dozens of centimetres in the estimated position (Guenther et al., 2000; Lague et al., 2020; Mandlburger et al., 2013). This delay depends on the water turbidity, which is the most influencing factor on the rising time of the water column component. If water is very turbid, the water column part of the global return can be the most contributing of the two, and the resulting error will be higher (Mandlburger et al., 2013). In clearer water, on the contrary, the interface dominates. It is however very difficult to clearly identify if the first peak is dominated by interface or water volume character (Guenther et al., 2000). If not post-processed and decomposed, the first peak of bathymetric waveforms thus reflects "a certain level of penetration into the water column" as explained by (Zhao et al., 2017), which partially depends on water parameters, but also on the laser incidence angle and divergence. Later in the signal, both components end up forming an elongated peak with a longer decreasing time than traditional surface signatures. The global effect of water on bathymetric lidar waveforms is summarized in Figure 1.18.

On its own, the air/water surface component has the following characteristics:

- Its **amplitude** is mostly determined by the **incidence angle** on the water surface, which is also influenced by the illuminated waves and their slopes, and the wind speed. Air/water interface peak amplitudes thus have a high standard deviation spatially (Guenther et al., 2000), contrary to water column components.

- Its **shape** mostly reflects the **transmitted pulse**, but is stretched by geometric effects of off-nadir incidence angles

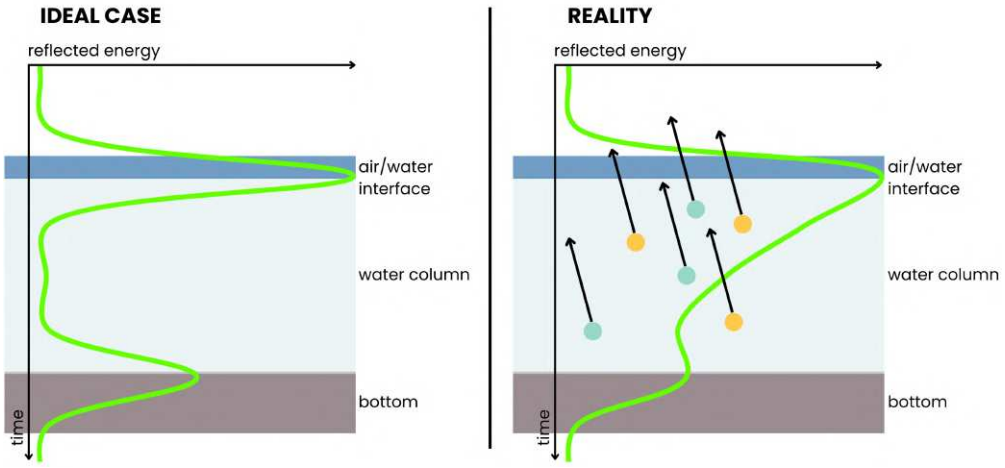


Figure 1.18: The impact of water on bathymetric lidar waveforms. Yellow and blue dots represent yellow substances and phytoplanktons, respectively.

The water column component

As it travels through the **water column**, the green laser beam is **exponentially backscattered and absorbed** by particles with depth, and partially transmitted (Guenther, 1985; Philpot, 2019). Optically, the components that attenuate the most green light in water are **yellow substances**, **phytoplanktons** and **sediments**. Their concentration adds up to the optical behaviour of clear water to impact the penetration of light in the water column. The water column is consequently visible in bathymetric waveforms as an exponentially diminishing shaped peak, characterized by a width related to the depth of the water body.

The attenuative power of water is approximated by the **diffuse attenuation coefficient** specific to each wavelength, K_d. As stated in Guenther et al., 2000, K_d "is the exponential factor by which the downwelling vector irradiance of the incident light field, at a given wavelength, decreases with increasing depth". The decay rate of light in water induced by K_d follows an exponential law, that can be formalized as:

$$P_r = P_i \times \exp\left(-2K_d D\right)$$

with K_d the diffuse attenuation coefficient of the considered water body, P_i the incident power and P_r the received power at the depth D.

K_d can be measured locally by monitoring irradiance depending on depth (Guenther et al., 2000). This coefficient is specific to each wavelength and depends on the scattering-to-absorption ratio, which is related to the amount and nature of **dissolved organic and inorganic material**. These water properties vary with tides, weather, and in time and space. In waters where absorption is dominant, bathymetric

lidars penetrate less deep than in scattering-dominated settings, which redistribute more of the energy. The **absorption** and **scattering coefficients** are **inherent optical properties** (IOP) of the water (Abdallah et al., 2012; Guenther, 1985; Guenther et al., 2000). They are expressed as the sum of the scattering and absorption characteristics of the different elements that constitute turbid water: yellow substances, phytoplankton and sediments.

The shape of the water column component in the waveform is therefore dictated partly by the local value of K_d, and consequently by the concentrations of the water column in yellow substances, phytoplankton and sediments. It also depends on the sensor's geometry, namely for narrow FOV lidars, which do not capture all the diffused energy and thus integrate only a portion of K_d in their waveform, denoted as K_{lidar}. Practically, K_d or K_{lidar} (depending on the FOV) can be approximated through the shape of the water column component in the waveform, using the linear relationship existing between received power and depth (Guenther, 1985; Guenther et al., 2000; Lague et al., 2020).

The seabed/riverbed/lakebed component

If there is enough remaining energy despite the effects of water turbidity on the laser beam, light reaches the water bottom and is reflected back to the receiver. The amplitude of the backscatter depends partially on the transmitted power and the **seabed reflectance**, similarly to any waveform component (Jutzi et al., 2006; Wagner, 2010), but also on the effects of the **water column**, through the **depth** and K_d parameters (Guenther et al., 2000; Kashani et al., 2015), since the power arriving at the seabed depends on the $\exp(-2K_dD)$ factor. The shape – and in particular the width and asymmetry – of this component is linked to the **geometry** of the surface and to sensor-related factors just as topographic components (see Section 1.5.1), but it must be remembered that the bathymetric lidar's footprint on the seabed is much larger due to its originally higher divergence necessary to emit a powerful wave, and to the scattering effects of water that exponentially spread the laser beam (Guenther et al., 2000). This **spreading** is both spatial and temporal, causing the incident pulse to have a larger width, and increasing the probability to have overlapping returns in shallow areas where air/water interface and water bottom returns occur nearly simultaneously (Guenther et al., 2000). This wider footprint may also influence the incidence angle on the seabed if it has a significant dip or roughness, potentially increasing the number of inflexion points of the resulting peak due to multiple backscatters across the footprint. Finally, the **refraction** of the incident beam impacts its path in water and thus the location in space of the backscatter.

Considering the specificities of lidar waveforms, simulators can be useful to formalize the numerous interconnected parameters shaping them. In the next section, we thus summarize the interactions of the laser beam on its travelling path in the form of radiative transfer equations.

1.5.3 Simulation of airborne lidar waveforms

Due to the complex physical interactions that dictate the generation of lidar waveforms in bathymetric environments, the development of processing methods adapted to such data often requires a calibration phase on simulated data. In this section, we develop the mathematical formulations that enable the simulation of realistic bathymetric waveforms.

Globally, each waveform component can be seen as an **environmental response function** (ERF) resulting from the interactions of the **laser source pulse** (LSP) and the **surface impulse response function** (IRF) (Guenther, 1985; Jutzi et al., 2006; Jutzi et al., 2003; Jutzi et al., 2007), so that:

$$\text{ERF} = \text{LSP} * \text{IRF} \tag{1.2}$$

The waveform (WF) is then defined as the **sum** of a **noise component** and as many **ERFs** as there were surfaces intercepted by the laser beam (Guenther, 1985; Jutzi et al., 2006; Abdallah et al., 2012; Jutzi et al., 2003; Tulldahl et al., 1999):

$$\text{WF} = P_{noise} + \sum_{i=1}^{N} \text{ERF}_i \tag{1.3}$$

where P_{noise} is a noise component and N is the number of intercepted surfaces. To formalize the energy received by the sensor after backscattering over a target, the interactions of the laser pulse as it travels towards and backwards the target must be considered. At each laser pulse emission, a surface A_i is illuminated. Considering the range R to this surface, the divergence of the beam γ and the aperture diameter of the laser emitter Da, A_i corresponds to the solid angle *viewed* by the beam (Baltsavias, 1999; Höfle et al., 2007; Jutzi et al., 2003; Babushka et al., 2014):

$$A_i = \frac{\pi}{4R^2}(D_a + R\gamma)$$

The power P_{in} received by A_i is (Baltsavias, 1999; Höfle et al., 2007; Jutzi et al., 2003; Babushka et al., 2014):

$$P_{in} = \frac{P_{em} \times \eta_e \times T_{atm}}{A_i}$$

with P_{em} the emitted power, η_e the optical efficiency of the emitter and T_{atm} the transmittance of the atmosphere. The illuminated surface A_i then reflects a portion of this energy back to the receiver. If this surface is Lambertian, it reflects energy uniformly in all directions with a cosine weighting depending on the angle of incidence of the received radiation (Baltsavias, 1999; Höfle et al., 2007; Babushka et al., 2014).

Thus, the portion of radiation directed at the sensor is

$$P_{out} = \frac{\rho \times P_{in}}{\pi} \times A_i \times \cos\theta$$

with ρ the reflectance of A_i and θ the angle of incidence of the laser beam on A_i. Finally, the power arriving at a receiver characterized by a surface A_r and an optical efficiency η_r is (Baltsavias, 1999; Höfle et al., 2007; Jutzi et al., 2003; Babushka et al., 2014):

$$P_{rec} = P_{out} \times \frac{T_{atm} \times \eta_r \times A_r}{R^2}$$

since $\frac{A_r}{R^2}$ is the solid angle *viewed* by the reflected radiation, that is again affected by the transmittance properties of the atmosphere (Baltsavias, 1999; Höfle et al., 2007; Babushka et al., 2014; Jutzi et al., 2003). In the end, under the assumptions that the illuminated surface is the sole intercepted target and provided it has Lambertian properties, the power reflected back to the receiver is

$$P_{rec} = \frac{\rho}{\pi} \times \frac{P_{em} \times \eta_e \times T_{atm}}{A_i} \times A_i \times \frac{T_{atm} \times \eta_r \times A_r}{R^2} \times \cos\theta$$

which simplifies as

$$P_{rec} = \frac{\rho P_{em} T_{atm}^2 A_r \eta_e \eta_r \cos\theta}{\pi R^2} \tag{1.4}$$

If several targets are hit successively, the received energy for the first return is described by Equation 1.4 (Baltsavias, 1999; Höfle et al., 2007; Jutzi et al., 2003; Babushka et al., 2014), but the returned energy of a second target of reflectance ρ_{t_2} hit with an incidence angle θ_{t_2} is

$$P_{rec_{t_2}} = \frac{\rho_{t_2} P_{trans} T_{atm} A_r \eta_r \cos\theta_{t_2}}{\pi R^2} \tag{1.5}$$

with

$$P_{trans} = 1 - P_{out_{t_1}} = 1 - \frac{\rho_{t_1} P_{em} \eta_e T_{atm} \cos\theta_{t_1}}{\pi}$$

These simplified equations are valid under the assumption that each intercepted target has a surface corresponding to the illuminated surface. Otherwise, both surfaces do not compensate for each other. In the specific context of bathymetric waveform simulation, three targets are intercepted: the air/water interface, the water volume, and the water bottom. Consequently, a bathymetric lidar waveform can be expressed (Tulldahl et al., 1999; Feigels, 1992; Guenther, 1985; Abdallah et al., 2012):

$$\mathrm{WF} = P_s + P_c + P_b + N$$

with P_s the return from the water surface, P_c the return from the water column, P_b the return from the water bottom, and N the noise.

Water surface backscatter When the illuminated surface is water, the same principles apply, but considerations linked to the specific interactions of light and water intervene. In particular, the reflectance of a water surface depends on its roughness. When it is completely smooth like a mirror, specular reflection occurs (Babushka et al., 2014; Feigels, 1992), and little to no energy comes back to the receiver. Water bodies are thus always surveyed with waves, which can be seen as a group of finite facets with Lambertian reflection (Abdallah et al., 2012; Guenther, 1985). However, these facets and their characteristics depend on the wind and the incidence angle (Guenther, 1985; Feigels, 1992). In the end, if we consider the water surface to be the first surface intercepted, equation 1.4 can be adapted to the backscatter P_s of a water surface (Guenther, 1985; Abdallah et al., 2012; Feigels, 1992; Kopilevich et al., 2008; Walker et al., 1999; Tulldahl et al., 1999):

$$P_s = \frac{\rho(w)N(\theta,\omega)P_{em}T_{atm}^2 A_r \eta_e \eta_r}{\pi R^2}$$

In practice, this represents the peak pulse power of the waveform return generated by a water surface. This equation thus incorporates the different components influencing the pulse's shape, as explained before. The influence of the intercepted surface's material is present through $\rho(w)/\pi$, which is the surface reflectivity per unit solid angle (Guenther, 1985). The impact of its geometry is conveyed by $N(\theta,\omega)$, the normalized Cox-Munk wave slope distribution that depends on the incidence angle θ and on the wind speed ω (Guenther, 1985). Finally, the parameters of acquisition are represented throug P_{em}, T_{atm}, A_r, η_e, η_r and the scanning angle θ, that impacts the measure of R. In the end, P_s is simply the portion of the emitted pulse power P_{em} reflected backwards after attenuation by a loss factor L_s that depends on water surface geometry and reflectance. It must however be noted that the function used to take water surface geometry into account in the above formalization of P_s - i.e. the Cox-Munk wave slope distribution - is merely a model of waves in coastal waters. It is thus not generalizable directly to all types of water: although lakes have waves generated by wind too, rivers do not follow similar patterns. Consequently, the loss factor should be modified for fluvial waters simulation. In the end, a global formula for P_s is:

$$P_s = \frac{L_s P_{em} T_{atm}^2 A_r \eta_e \eta_r}{\pi R^2} \tag{1.6}$$

Water column backscatter After having penetrated through the air/water interface, the beam is refracted and transmitted to the water column. The interactions of the laser pulse with the water column can be seen as a succession of interceptions of very small surfaces: the particles suspended in water. As a matter of consequence, the water column component of a bathymetric waveform is affected by sur-

face material and geometry in a similar way that other components are. However, considering the size of the particles hit relatively to the incident beam, the interactions are of a different nature - scattering and absorption, as opposed to reflection (Guenther, 1985) - and occur repetitively until the light reaches the bottom, thus involving a temporal component. In the end, the water column backscatter P_c is expressed using Equation 1.5 (Guenther, 1985; Abdallah et al., 2012; Feigels, 1992; Kopilevich et al., 2008; Walker et al., 1999; Tulldahl et al., 1999; Kim et al., 2016):

$$P_c = \frac{\beta_\pi \exp\left(-2kr_w\right)(1 - L_s)P_{em}T_{atm}^2 A_r \eta_e \eta_r}{n_w^2 R^2} \tag{1.7}$$

The $\frac{1}{\pi}$ and $\cos\theta_w$ weighting disappear as water column particles do not generate Lambertian reflection but diffusion and absorption. The incident energy is absorbed exponentially with depth - denoted here by the range in water r_w - depending on a coefficient k. Since this absorption occurs both during downwelling and upwelling, a factor of 2 is applied in the exponential (Guenther, 1985). k is a coefficient reflecting the attenuation due to water turbidity and the influence of the FOV - thus not strictly similar to the diffuse attenuation coefficient (Guenther, 1985; Feigels, 1992; Tulldahl et al., 1999). For large FOVs, k can be approximated by the diffuse attenuation coefficient K_d (Guenther, 1985; Feigels, 1992). However, for narrower FOVs, a loss factor linked to the sensor's receiving optics must be incorporated to get k (Guenther, 1985). Additionally, the incoming radiation is diffused by suspended particles of the water volume. This diffusion occurs over 2π, but the receiver only captures the portion of energy diffused parallelly to the incidence - i.e. at $180°$ -, which is β_π. A $\frac{1}{n_w^2}$ factor appears because the solid angle illuminated by the incident beam is enlarged in water due to light refraction, both when it enters and leaves the water column (Guenther, 1985). In the end, contrary to P_s, P_c is computed for the complete time interval during which light travels through water, and not only at one time step.

Water bottom backscatter Eventually, light reaches the water bottom. To model the backscatter generated by the water bottom, equation 1.5 must be modified to account for light attenuation (Guenther, 1985; Abdallah et al., 2012; Feigels, 1992; Kopilevich et al., 2008; Walker et al., 1999; Tulldahl et al., 1999; Kim et al., 2016):

$$P_b = \frac{\rho_b \exp\left(-2K_d r_w\right)(1 - L_s)P_{em}T_{atm}^2 A_r \eta_e \eta_r}{\pi n_w^2 R^2} \tag{1.8}$$

Although the seabed is assumed to have Lambertian reflection, the weighting in $\cos\theta_w$ of its reflectivity is not present in Equation 1.8 because it is, by definition, included in K_d (Guenther, 1985).

Noise component The last parameter to take into account to simulate realistic lidar waveforms is noise. The noise of lidar waveforms is often considered Gaussian (Abdallah et al., 2012; Guenther, 1985; Tulldahl

et al., 1999). Two types of noises can be considered in lidar waveforms: a background solar noise, causing the signal to never be truly zero, and a noise relative to the receiver electronics (Abdallah et al., 2012; Guenther, 1985; Tulldahl et al., 1999; Feigels, 1992). The solar noise is linked to the background light that cannot be suppressed and is thus also captured by the receiver. In the bathymetric case, for daytime acquisitions - which are the most common as explained in Guenther, 1985 - it can be expressed as (Guenther, 1985; Feigels, 1992)

$$P_{N,D} = \frac{I_S A_S \Delta B A_R \eta_R}{R^2} \tag{1.9}$$

with I_S the solar irradiance reflected by the water column, A_S the area within the receiver FOV on the water surface, ΔB the bandwidth of the interference filter of the receiver, and A_R the effective aperture area of the receiver telescope. To simulate the solar noise, a Gaussian white noise with a mean of 0 and a standard deviation of 1 can be convolved with $P_{N,D}$. On the other hand, the electronic noise is a Gaussian noise with a mean of 0 and a standard deviation varying with respect to the signal level (Abdallah et al., 2012).

In the literature, multiple simulators of bathymetric lidar waveforms relying on these equations can be found. They sometimes involve Monte Carlo processes (Guenther, 1985; Guenther et al., 2000), namely for ray tracing (Gastellu-Etchegorry et al., 2016). Most of the time, they are built to better understand lidar principles (Guenther, 1985; Feigels, 1992; Tulldahl et al., 1999; Walker et al., 1999; Kopilevich et al., 2008) or to develop waveform processing approaches and better anticipate the results of lidar surveys (Wu et al., 2011; Abdallah et al., 2012; Kim et al., 2016).

1.6 Conclusion

Topo-bathymetric lidar remote sensing has the potential to address the issue of land-water continuum observation: it allows the simultaneous, seamless modelling of surfaces below and above the waterline at large scales while preserving their physical integrity. However, this sensor produces very specific data that are difficult to interpret straightforwardly. Similarly to other remote sensing imagery data, they thus require the development of adapted processing methods, whether for visualization or information extraction purposes.

3D point clouds are irregularly spaced and characterized by a great sparsity, which makes them inadapted to existing image processing tools without rasterization and the consequent loss of information. Their third dimension also makes them heavier to process, which means that developing adequate methods is not trivial. On the other hand, lidar waveforms are particularly noisy and site- and acquisition-dependant. The physical processes they originate from also make them difficult to accurately analyse. For example, overlapping or weak returns require signal processing tools designed specifically to take into

account the physics behind waveforms.

Though these data were still particularly challenging to deal with a few decades ago, the development of advanced data processing techniques based on machine learning, combined with computing hardware progresses, has resulted in major advances for lidar data processing. In chapter 2, we review methods adapted to lidar data processing in complex environments and discuss the remaining challenges and opportunities for the objective of land-water areas observation.

CHAPTER

2

LIDAR DATA PROCESSING FOR ENVIRONMENTAL APPLICATIONS: A REVIEW

Contents

Remote sensing (RS) measurements - including lidar data - rarely constitute straightforward information that can directly be used to draw conclusions on natural phenomena. Most RS data require processing to be visualized in understandable forms and analyzed. The quantity of information contained in the raw RS imagery data, their ever-increasing amount, and their complexity also necessitate automated processing chains to be used to derive knowledge about Earth. The emitted or reflected radiations measured can be a mere proxy of a target phenomenon or the studied subject. In both cases, processing methods are required, to formalize intuitive analyses of imagery such as how reflectance informs on an object, or to automate the conversion of raw recordings into knowledgeable information.

In this chapter, processing lidar data is discussed. We first introduce learning-based methods in general in section 2.1. The processing of 3D point clouds and lidar waveforms is then discussed in sections 2.2 and 2.3, respectively. In the conclusion, we point out open challenges that arise from this chapter and that will be developed in the next chapters.

2.1 Learning-based data processing methods

2.1.1 General notations

Data processing consists of extracting information Y from data X using a mathematical function linking them so that $f(X) = Y$.

X can be an image - a group of pixels - or a point cloud - a group of 3D coordinates. Y can be a numerical or a categorical variable. If Y is categorical, the task of deriving Y from X is **classification**. If Y is numerical, this task is **regression**. The relation $Y = f(X)$ can be known - for example, extracting the slope from a digital terrain model is computing a gradient - or unknown. In this latter case, it can be estimated based on examples of X and Y, using machine learning.

Learning f directly on samples of X is a delicate task since the relation from raw data X to interpretable variables Y can be complex. To simplify this mathematical problem, X can be transformed into **features** $g(X)$ defined by the user. Learning $Y = f(g(X))$ is then a simpler problem. The features $g(X)$ are designed to formalize the descriptive characteristics of the processed data and increase the amount of information available to derive the prediction rule.

In practice, to approximate f, a function $f_\theta(g(X))$, where θ is a set of parameters of f, is used. Parameters θ then have to be optimized to make $f_\theta(g(X))$ as close to Y as possible. For example, a simple model can be linear regression, in which case

$$f_\theta(g(X)) = \alpha g(X) + \beta, \text{ and } \theta = (\alpha, \beta)$$

Efficient methods to optimize the value of θ depending on $g(X)$ exist for various models. In the last decades, methods to optimize more complex functions have emerged, lifting the need to simplify the mathematical problem by computing features $g(X)$ and directly working on approximating $f_\theta(X)$. A typical example is **neural networks**. Such methods have quickly outperformed the simpler models used to estimate $f_\theta(g(X))$.

Estimation of θ

In most cases, finding the value of θ is **solving an optimization problem**. Since θ is unknown, its value is estimated **based on data**. The relation linking X and Y is learned on N couples (X_i, Y_i), for example by finding θ such that

$$\forall i \in \{0, 1, \dots, N\}, f_\theta(X_i) \approx Y_i$$

The optimal value of θ, θ_{optim} can be formalized as

$$\theta_{optim} = \underset{\theta \in \mathbb{R}}{\operatorname{argmin}} \mathcal{L}(Y, f_\theta(X))$$

with $\mathcal{L}$ a **loss function**, defining the error between Y and $f_\theta(X)$. $\mathcal{L}$ depends on the application and must be defined by the user. Common loss functions are the Mean Squared Error (MSE) or the Mean Absolute Error (MAE) for regression problems, and the Cross Entropy or the Focal Loss for classification tasks.

To solve the optimization problem and find θ_{optim}, numerous techniques exist. Among them, gradient descent is commonly used.

2.1.2 Main families of machine learning models

In this section, the most common models for f relying on transformations $g(X)$ are reviewed. Among them, one can list:

- **Probabilistic methods** involve modelling probability distributions for Y based on the uncertainty and variability of input data X. These approaches offer a more comprehensive and robust perspective for decision-making, especially in domains with noisy or incomplete data. Final decisions regarding the values of Y are made by considering the underlying probability distributions, and associated loss functions often aim to maximize probabilities.

- **Margin-Maximization models**, such as SVM (Support Vector Machines) for classification and

SVR (Support Vector Regression) for regression, seek to either find the hyperplane that maximizes the margin between different data classes (in the case of SVM) or maximize the margin around data points (in the case of SVR) (Cortes et al., 1995). Leveraging the kernel trick, which allows data to be projected into high-dimensional spaces where linear models are effective, these methods are well-known for their efficiency in modelling non-linear relationships between features and target values within a supervised learning framework.

- **Hierarchical tree-based methods**, such as decision trees, are widely employed for classification and regression tasks (Breiman et al., 1984). They operate by recursively dividing the dataset into smaller subgroups based on features, creating a tree-like structure. In classification, each leaf of the tree represents a class, while in regression, each leaf contains a numerical value. These trees facilitate decision-making by following a path from the root to a leaf, making them well-suited for modelling complex relationships between features and target outputs.

It is worth noting that most machine learning algorithms are likely to be unstable to bias in the data, such as the presence of outliers or the redundancy of information in the input. To reduce this sensitivity and build robust learners, ensemble methods use multiple algorithms trained on random subsamples of the training datasets and use the mean prediction of the group of algorithms obtained as the final decision (Sagi et al., 2018).

Ensemble learning

Bagging is a part of ensemble learning that designates the method used to resample the training dataset before building each algorithm of the ensemble. The different subsets are obtained randomly and may thus contain similar samples. The advantage of bagging is to stabilize algorithms with high variance and dependence on the input distribution.

Boosting is another strategy to make a machine learner more robust. When using ensemble learning, each model is built independently, regardless of the strengths or weaknesses of its companions and its potential to improve the global result. This means that the ensemble of models will put on the same level the highly and badly performing models obtained with different subsets of training data. Boosting aims at weighing the individual models that populate the ensemble depending on their performances, to try to compensate for the errors made in the group (Chen et al., 2016; Hastie et al., 2009).

Random forests (Breiman, 2001) are an example of the use of ensemble learning. They are among the most popular, robust and easy-to-deploy machine learning algorithms. They are the result of

applying ensemble learning to hierarchical trees. Though they do not natively involve boosting, they can be coupled with boosting methods such as AdaBoost (Hastie et al., 2009) or XGBoost (Chen et al., 2016), that update the forest based on the individual trees. Random forests use a subsampling of input samples and input features to build trees that perform regression or classification (Breiman, 2001). Since they rely on ensemble learning, they produce their prediction in the form of a proportion of classifiers that voted for a given output, thus resulting in prediction probability information (Breiman, 2001). They are also popular for another statistic that they provide on the features used to make the prediction: using the impact of a feature on accuracy or node impurity when it is used, they compute an indicator on each feature's importance in the obtained result (Breiman, 2001; Breiman et al., 1984). It should be noted that ensemble learning can be used with models based on pre-defined features and with feature learning approaches.

Feature selection for classical machine learning

When using machine learning algorithms based on pre-defined features, introducing a wide range of information to represent the data may not always be optimal. Some of the multiple descriptive features may be irrelevant regarding the target variable, or redundant. Although irrelevant or redundant information should intuitively not impact the prediction, it can impair the learning process by increasing the number of parameters involved and complexifying the optimization process. Similarly to the scales exploited, an optimized feature set should thus incorporate the most information possible, while also limiting redundancy between attributes. Considering the variety of information derivable from 3D data, empirically selecting the attributes to integrate into a classification is not only time-consuming but also hazardous, as it might impair classification performances. Feature selection methods allow the automation of a great part of the feature vector construction. They are mainly based on the estimation of an attribute's relevance relative to the predicted variables and on the minimization of correlation between relevant parameters. As explained in Dash et al., 1997, feature selection methods can be split into three categories. Filter-based or univariate methods aim at maximizing the relevance of the predictors used. They use relevance score functions and rankings of the scores to only keep a subset of the most informative features for classification. Popular score functions include Fisher's index or Information Gain index, but adapted metrics that take multiple aspects of feature relevance into account also exist (Weinmann et al., 2013). Multivariate methods try to minimize feature redundancy among the relevant attributes, often by combining score functions with correlation assessments (Dong et al., 2017; Weinmann et al., 2015). Both univariate and multivariate approaches are independent of the classifier used, and its settings, which is sometimes seen as a generalization advantage (Weinmann et al., 2015), but also do not account for inter-feature synergies, and may evict highly correlated but still informative features (Guyon et al., 2003). Wrapper methods

and embedded feature selection consist of exploiting classifier outputs to select features. They either use classification accuracy obtained using each feature separately as a score to prune the input vector (Dong et al., 2017) through backwards or forward selection, or rely on feature importance information provided by algorithms, to evict least important predictors and improve accuracy (Guan et al., 2012). Random forest-based metrics are among the most common embedded selection strategies.

Recently, neural networks have significantly developed due to their higher performances and abilities to solve more complex problems. In the following sections, we detail the principles of neural networks.

2.1.3 Deep neural networks

Deep neural networks belong to probabilistic machine learning methods. They are used to model complex probability distributions.

General principles

A neural network is a stack of neurons, also called perceptrons. A **perceptron** P is a system that takes an input X of dimension N, $X = \{X_1, \ldots, X_N\}$ and models the output Y as:

$$Y = \sigma\left(w_0 + \sum_{i=1}^{N} w_i X_i\right)$$

Here, the parameters to optimize are the weights $w_0, \ldots, w_N$. The function σ is called an **activation function**. Its role is to incorporate non-linearity into the system, which is otherwise a mere linear regression. Figure 2.1 illustrates a perceptron.

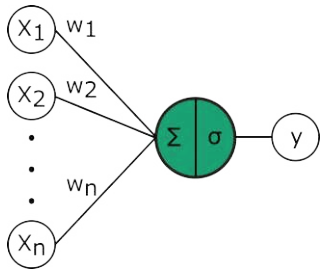


Figure 2.1: Schema of a perceptron.

A perceptron forms an elementary function that can be combined multiple times in an interconnected manner to form **neural networks**.

It has been shown that no matter its complexity, any mathematical relation f can be approximated with a limited number of stacked perceptrons (Cybenko, 1989). As a consequence, using a large number of perceptrons to approximate a complex function f linking outputs of interest Y to observed data X can be done with a combination of a (potentially large) number of perceptrons.

Figure 2.2 illustrates a neural network formed by stacking a finite number of perceptrons to estimate a more complex function.

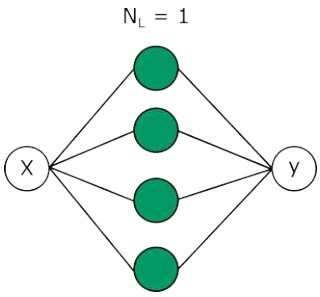


Figure 2.2: A neural network with one layer of stacked perceptrons. N_L refers to the number of layers in the network.

Empirically, it reveals more efficient to successively use smaller stacks of perceptrons than to use a single, very large, stack of perceptrons. Each stack of perceptrons then forms one **layer** of a neural network, also called in this setting a **multi-layer perceptron** (MLP). Figure 2.3 illustrates the resulting neural network.

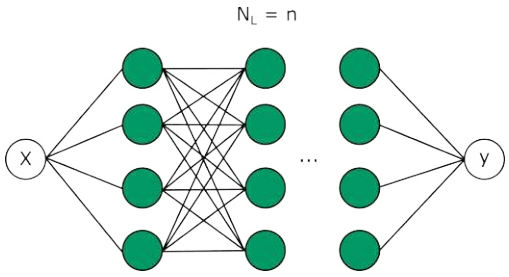


Figure 2.3: A multi-layer perceptron. N_L refers to the number of layers in the network.

The number of perceptrons per layer, the number of layers, and their interconnections are what we call the **architecture** of the neural network. An architecture with multiple layers (typically a large number of layers) results in what we call a **deep neural network**. In such cases, a closer look at the last layer of the network shows that it works similarly to a single layer perceptron applied to transformed data, similarly to approaching f based on features of the input X. At this stage of the network, the task is to approach a simpler function on well-transformed data. When taking this point of view, the first layers of the network

are in charge of performing the transformation of the inputs and thus learning features of the inputs. The goal is then to **learn the best transformation of** X. This is one of the main strengths of deep neural networks, which directly learn and compute features of the input data instead of requiring predefined and pre-computed features as input.

The optimization of deep neural networks can be delicate, since the cost function associated with a complex function $f(X)$ can be very complicated and a large number of parameters w must be estimated. This problem is dealt with thanks to:

1. **Gradient backpropagation**, a technique that allows efficient computation of the gradient in complex networks (Lecun et al., 1998);

2. **Access to large datasets**: approximating thousands of parameters requires using a lot of observations;

3. **Access to powerful computers** with the capacity to perform heavy computations.

In this section, we presented deep neural networks based on perceptrons. Networks relying on other operations also exist for different types of inputs and tasks. In the end, deep neural networks provide a flexible framework to work with very complex relations.

Deep neural networks for 2D and 3D structured data

In EO, many different types of structured data can be collected. Structured data are organized around neighbourhood relationships, with no independence between neighbouring measurements[1]. It can be interesting to exploit this **spatial structure** when extracting information from the data. Several deep neural network architectures are specifically designed to capture the relations between neighbour data instances.

An option is to perform spatial convolutions to capture the patterns characterizing neighbourhood dependencies. Such architectures are **convolutional neural networks (CNNs)**. They follow the same principles as MLP but simply replace the linear regressions with convolution operations.

1. Sometimes, point clouds are considered "unstructured data" although they have a spatial structure. This is simply because they do not follow a regularly gridded structure.

2D Convolution

In 2D, the convolution operation involves two elements: a **kernel**, and an input matrix. The kernel is a square matrix containing **weights** and defined by its **size**. The size of the kernel is systematically smaller than that of the input. During the operation, an **element-wise multiplication** is performed between the kernel and the input. The values of this multiplication are then **summed**. In practice, in a convolutional layer, the kernel is used as a **sliding window** to compute a **feature map** with values corresponding to the convolution operation obtained by centring the kernel on each of the input matrix elements successively. The weights of the kernel are learned during optimization.

Convolutions use a sliding window - the kernel - of a given size to consider the whole input. To capture long-range relations, a solution can be to use large-size windows. However, increasing the size of the kernel results in a drastic increase of the dimension of θ, and thus of the number of parameters to estimate. To limit the corresponding complexity and computation cost increases, another option is to reduce the size of the input of the convolution, so that the same filter size virtually *sees* a wider portion of the input. **Pooling** operations are used to reduce the size of the input to a layer, by applying mean or maximum filters of size two to it, for example.

Different types of neural network architectures exist, mainly defined by the number and organisation of the neurons composing them, and by the type of linear operations they embed. Among the deep neural networks family, convolutional ones (Lecun et al., 1995) are among the most performant thanks to their native ability to consider the spatial context of the considered element in their prediction, and thus learn motives in the data. Based on alternations of convolution and pooling operations, these networks progressively learn more and more abstract features and consider increasingly larger portions of data to incorporate multiple scales.

Convolutional networks use the same weights and operators across the whole input, contrary to MLPs, which require the repetition of as many linear operators as desired outputs. This reduces the number of parameters to optimize, as this number depends on the size of the convolution kernel, the convolution stride and padding used, and the number of convolution filters performed, and not on the size of the input and the output. The key to their performance is the use of such shared weights involving strong local connections in multiple layers (Lecun et al., 1995; Zhou et al., 2020)

When processing structured data such as an image, a possibility is to produce a single output for the whole image (image classification) or an output per pixel of the image (semantic segmentation).

The **CNN** is the typical architecture used to output one value per image. To produce per-pixel predictions, another typical architecture exists: U-Net. **U-Net** (Ronneberger et al., 2015) is named after its U-shaped architecture, coupling two parts: one consisting of convolutions and downsampling operations, and the other made of upsampling operations and convolutions. The first is called the **encoder**,

and the latter is the **decoder**. U-Net adds to each decoder layer's input the encoder layer's output of the corresponding size to maximize the amount of learned features included in the prediction. The resulting architecture is schematized in Figure 2.4, extracted from the original paper (Ronneberger et al., 2015).

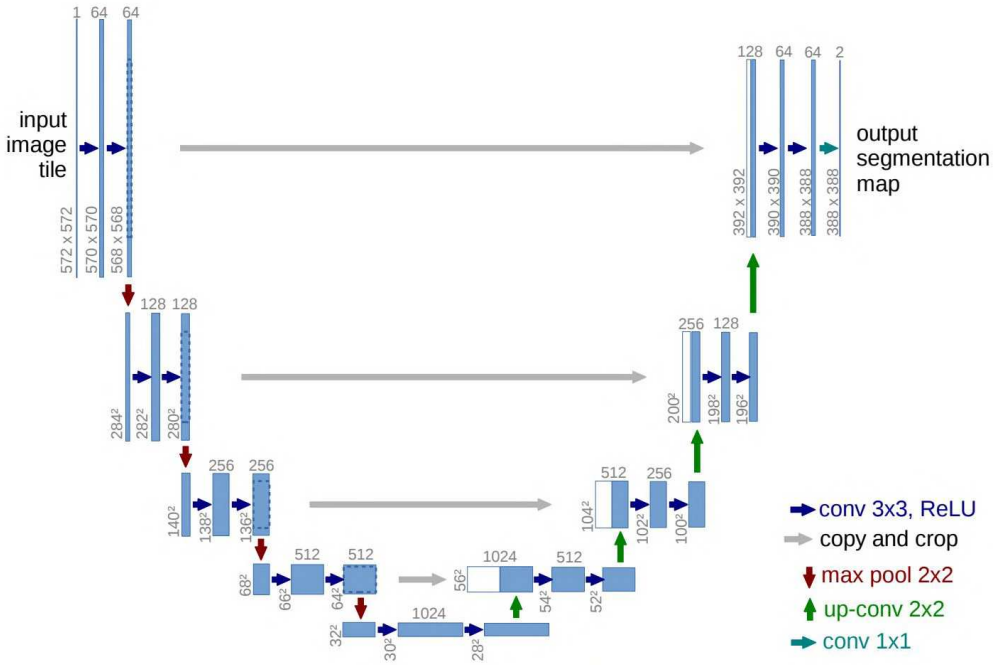


Figure 2.4: The U-Net architecture, as illustrated in the original paper of Ronneberger et al., 2015.

Both of these typical architectures exist with multiple variants allowing them to solve a variety of problems and tackle typical limitations identified by their predecessors.

Several typical variants are:

- **Residual networks** – introduced in (He et al., 2016) – connect convolution layers' outputs to their inputs before feeding them to the forward layers. Residual connections mitigate the problem of vanishing gradients and thus allow deeper networks to be trained.

- **Autoencoders** (Rumelhart et al., 1986) are networks used to reconstruct their input. Their most popular form incorporates two parts: an encoder in which the size of the feature maps is progressively reduced to allow convolutions with constant kernel sizes to cover bigger portions of the data, and a decoder in which feature maps are progressively upsampled to recover the size of the output. Such networks are often used to learn and extract meaningful representations of unlabeled data.

- **Inception networks** – introduced in (Szegedy et al., 2015) – are CNNs that learn convolutions

with different kernel sizes in parallel to make the receptive field of each layer vary, and thus learn more representations of the data

- **Siamese networks** – a concept originally introduced in (Bromley et al., 1993) – combine two twin networks that join at their outputs. They exist in different configurations, mostly used to fuse several input sources (Hazirbas et al., 2017), or perform change detection (Caye Daudt et al., 2018).

- **Region-based neural networks** are a family of networks built to perform object detection and classification. Among them, YOLO (Redmon et al., 2016) is the most famous, as it performs object detection and classification simultaneously, and in real time.

Initially designed for 2D image processing, CNNs can originally only operate on regular Euclidean data such as images and sequences (Lecun et al., 1995). Though they have been adapted to graphs using adjacencies relationships to propagate weights (Zhou et al., 2020), their development for sparse 3D point clouds was slowed down by their complex neighbourhood relations, causing specific architectures to be developed, such as PointNet (Qi et al., 2017b; Qi et al., 2017a) and KPConv (Thomas et al., 2019), which will be better introduced later in this Chapter.

1D Deep neural networks

Although all of the previously cited architectures were developed for image processing purposes, they can all be adapted to other types of data on which convolutions can be performed – including graphs and time series. Good examples are **time-delayed CNNs** (Lang, 1988), **TempCNN** (Pelletier et al., 2019) and **U-Time** (Perslev et al., 2019), 1D versions of CNNs and UNet, or GNNs (Wu et al., 2021), **graph convolution networks**, that exist in various configurations. However, sequential inputs with temporal relations also benefit from specifically designed methods:

- **Recurrent neural networks** (RNN) (Rumelhart et al., 1986) include specific modules that memorize information about previous timesteps when processing a given sequence sample.

- **Long Short Term Memory** (LSTM) (Hochreiter et al., 1997) gates and **Gated Recurrent Units** (GRU) (Chung et al., 2014) are specific types of RNN modules, with different ways of selecting the information to keep in memory.

- **Transformers**, introduced in (Vaswani et al., 2017) combine feed-forward MLPs with **attention mechanisms** – introduced in (Bahdanau et al., 2014). Wen et al., 2022 sum up the self-attention modules of the Transformer as "a fully connected layer with the weights that are dynamically generated based on the pairwise similarity of input patterns". One of the main traits of Transformers

is that they can attend to complete sequences thanks to self-attention, which gives them abilities to model long-range dependencies and interactions in sequences that convolutions do not capture (Wen et al., 2022; Khan et al., 2021). By not relying on sequential computation, they are also optimized for parallelisation. Finally, they experimentally appear to be scalable to very large models, high-complexity tasks, and large-scale datasets (Khan et al., 2021). However, these advantages come at the expense of an increased need for training data and computationally intensive training processes.

2.1.4 Current challenges

Below, we list some of the current limitations and challenges of deep neural networks.

Making more of available labelled datasets A first possibility to produce networks with a lower training cost can be to exploit better the amount of labelled data available, even if it is not massive. Datasets can be artificially augmented using **data augmentation**, which consists of artificially rotating, cropping, or changing reflectance parameters of the labelled samples to simulate different conditions of observation of the objects they represent (Perez et al., 2017).

Data can also be **simulated**, either using physics-based simulators (Gélis et al., 2021) or by relying on generative networks (Goodfellow et al., 2014) to create new samples to train on.

Another way to reduce the training cost by finding data workarounds is to use **domain adaptation** or other types of **transfer learning** in order to reuse another existing labelled dataset, and potentially even another already trained model (Weiss et al., 2016). Domain adaptation consists of adapting a machine learning model originally trained on data from one domain - the "source" domain - to generalize and perform on data from another related domain - the "target" domain. For example, it can imply finding a transformation between two data ensembles and transferring data from one domain to another, so the features learned on the first dataset are also relevant to the other.

Weakly supervised and unsupervised learning The second possibility is to change the training paradigm. Instead of relying on supervised learning, it is possible to use **weak supervision** (Zhou, 2018), or even **no supervision** to learn the network's parameters. This reduces, if not cancels, the labelling effort. Weakly supervised learning has two main configurations:

- Using a **limited amount of labelled samples**. This includes few-shot learning (Wang et al., 2020b) – generalizing to new classes with few examples of each –, zero-shot learning (Wang et al., 2019b) – identifying objects never seen during training –, and active learning (Settles, 2009) – in which the model queries the user to label a subset of samples.

- Using **incomplete annotations**, which is also referred to as inexact supervision or weak supervision: for example using per-image annotations to derive per-pixel predictions (Xu et al., 2020).

Unsupervised learning (Barlow, 1989; Ghahramani, 2004) mostly consists of learning patterns on unlabelled data samples. Clustering is then performed in the feature space to identify categories of items with similar features. In the context of deep learning, the goal can be to learn descriptive features of the data by using an encoder-decoder (Kingma et al., 2013; Rumelhart et al., 1986) trained to reproduce the inputs, or to exploit contrastive learning (Khosla et al., 2020) to learn attributes common to data samples, and attributes that separate them.

However, globally, these developments are still recent and remain methodological experiments rather than solutions applicable to a large set of problems. Supervised learning is thus still present, mostly in applied contexts.

Explainability

Increasing attention is also given to the **explainability** of results obtained with deep neural networks, which are often pointed out as black boxes, which sometimes hinder their deployment.

According to Roscher et al., 2020, explainable machine learning in the natural sciences should incorporate transparency, interpretability, and explainability.

Several frameworks exist to address these objectives, such as Local Interpretable Model-agnostic Explanations (**LIME**) (Ribeiro et al., 2016), or the use of **Shapley** values (Lundberg et al., 2017; Shapley, 1953). Other methods more specific to deep neural networks are also progressing, for example, **Axiomatic Attribution** - or Integrated Gradients - analysis (Sundararajan et al., 2017) or **Grad-CAM** (Selvaraju et al., 2020) which use gradients evolution or activation maps to identify the key elements behind a network's decision.

Computational cost

Another current challenge of deep learning is the **computational complexity** of most models. There are more and more tasks requiring almost real-time solving, which is sometimes incompatible with complex architectures.

A global trend in the field of feature learning at the time of this manuscript is thus to turn towards **lighter models** while trying not to sacrifice performance. To this end, the **Transformer** architecture (Vaswani et al., 2017) is increasingly adapted to other tasks than natural language processing (Wen et al., 2022; Robert et al., 2023; Dosovitskiy et al., 2020; Khan et al., 2021) with modifications allowing to reduce the computation cost initially associated to self-attention. Indeed, Transformers initially require a large

amount of training data and a computation-intensive training phase to reach the high performances they are praised for (Khan et al., 2021). However, they have the major advantage of being able to learn both short-range and long-range dependencies, contrary to convolutional neural networks, thanks to their self-attention mechanism (Wen et al., 2022; Khan et al., 2021). Coupling them with sampling strategies or new self-attention modules designed to reduce the computation cost thus enables models to exploit their strengths while limiting the computation expense. For example, the adaptation of the 3D PC processing network Superpoint Graph (Landrieu et al., 2018) with a Transformer-inspired architecture results in a network as accurate but 200 times more compact than its state-of-the-art equivalents, which also makes its optimization much faster (Robert et al., 2023).

Current challenges of machine learning and lidar remote sensing

Machine learning has enabled great progress in EO and in the interpretation of RS imagery. However, several factors caused lidar to be left out of the original development of machine learning algorithms, which mostly focused on images, times series, and language:

- Lidar emerged in the 1970s, after the advent of the laser, while language processing and image processing were investigated already in the late 1940s. This historically longer period of application makes image processing more advanced than 3D data interpretation.
- Technical constraints long limited the analysis of lidar data in 3D, as they come with additional complexity, requiring adapted visualization software and processing tools that are able to support the heaviness of handling a third dimension.
- Lastly, the availability of lidar instruments and lidar data, as well as their cost has limited their accessibility and thus their use.

As a consequence, lidar originally did not benefit from state-of-the-art data processing developments. Lidar data were thus long reduced to *2D rasters*, losing the spatial structure information of the point clouds, in order to fit in the existing processing options and exploit the highly powerful 2D processing methods available. As the computational power of computers increased, machine learning became possible on 3D data, but is still a very recent field compared to image analysis. Similarly, lidar waveforms benefitted from the available machine learning algorithms such as random forests. However, their analysis with deep learning methods, lifting the need to extract features, is still scarce.

Fewer data benchmarks, pre-trained models, or user-oriented software exist for 3D point clouds and full waveforms than for other RS data. As a consequence, **deep learning approaches**, that allow a larger automation of data analysis and have proven to outperform shallow learning in recent years, **are still less**

accessible to lidar data. Limitations specific to 3D point clouds are discussed more deeply in Section 2.2, while those related to full waveforms are presented in Section 2.3.

2.2 3D point clouds classification in natural environments

Classification of 3D data is a challenge, as 3D data are unstructured, irregular, and unordered. Their characterization is made harder by the local density variations, and the complex objects they contain. In this section, we review methods for producing supervised 3D classifications. Clustering methods and approaches relying on rasterized lidar data are not reviewed. Existing supervised 3D point cloud classification methods can be organized into two categories: handcrafted features with conventional classifiers and learned features with deep neural networks. Before entering into the details of each paradigm, let us recall the specifics and motivations of lidar point clouds processing.

2.2.1 Motivations and specifics of 3D lidar point clouds processing

3D point clouds are **unordered** and **irregular** data. Practically, this means that the points are not registered in a logical order and can thus be interchangeable, that their spatial repartition does not follow any a priori structure, and that they are not regularly spaced, which results in point clouds being a particularly **sparse** type of data.

The main information provided by airborne lidar point clouds is a model of the **relief of the surveyed area**. Lidar point clouds are thus often used to characterize the geomorphological structure of sites, either by analyzing the surface topography – thus including any object located over the ground – or by considering only the terrain's elevation - which implies removing all points resulting from interceptions of overhanging objects. To produce such Digital Terrain Models (DTMs) or Digital Surface Models (DSMs), point clouds are often pre-processed and regularly gridded to output 2D rasters of the desired information. Airborne lidar point clouds can provide information on the **nature of the scene** too, which allows the generation of land use/land cover 3D maps, or the production of models analyzing specific types of objects, such as Digital vegetation Height Models (DHMs) or canopy models. In particular, topo-bathymetric lidar data classification is necessary to respond to the following challenges:

- identification of bathymetric points to derive seabed topography;
- accurate classification of the water surface;
- correction of the effects of refraction on the delivered 3D point coordinates;
- detection of variations in seabed types.

Additionally, point clouds can be used to study **characteristics of the scene** that are not the nature of the objects intercepted. For example, the dip, roughness or mean elevation can be studied directly to characterize the terrain, or used to approximate related parameters about the environment. **Extracting knowledge from point clouds can thus be an end in itself or simply a step towards other goals**. To illustrate, automatically detecting the nature of the surfaces sampled in the data is the finality for 3D mapping purposes, but a mere pre-processing task in biomass assessment studies.

To access all the knowledge contained in lidar point clouds without manually estimating the necessary parameters – point nature, dip of the surface, etc – data processing methods specifically designed for point cloud characteristics extraction exist. In particular, many methods allow the extraction of descriptive parameters of the point cloud that allow automatic interpretation, such as automatic classification of the point's nature.

When similar surveys are conducted on multiple dates, the changes occurring in the scene can be studied using lidar point clouds too (De Gélis, 2023). However, here we will not mention methods tackling the issue of point clouds comparison, as it is out of the scope of this thesis. We will only describe how descriptive information can be extracted from point clouds to allow automatic knowledge extraction, namely to characterize the nature of the objects at the origin of each point. Similarly, though the methods described in the following sections can be used to perform regression and approach physical parameters or any other desired numerical output, regression methods on lidar point clouds will not be detailed hereafter, as solely the classification of categorical variables is developed in the thesis.

2.2.2 3D point clouds classification using handcrafted features

Semantic classification of point clouds consists of attributing a label to each point, which can also be seen as point classification. To perform this task with supervised machine learning classifiers, an input vector must be fed to the classifier. This vector often consists of a group of handcrafted data attributes – also called features, or descriptors – that encode characteristics of the points and their context. The following sections detail how a 3D point cloud can be automatically described using mathematically formalized characteristics, and how the resulting attributes can be used to obtain a classified point cloud.

Features definition

The **spatial repartition of the points** in a PC and, for multiple return lidar, the echoes' number, ordering and characteristics depend on a combination of sensor physics and surface geometry. The reflected **intensity** is also linked to the albedo of the surveyed object and to the sensor. They thus act as proxies of the actual surface characteristics. PC classifications consequently exploit the **geometry** of the PCs

(Hackel et al., 2016) and their spectral dynamics (Chehata et al., 2009) or their **local dimensionality** (Brodu et al., 2012; Vandapel et al., 2004). For example, the eigenvectors of each point's neighbourhood covariance matrix are popular attributes in identifying isolated points, lines, planes, volumes, contours and edges in PCs (Gross et al., 2006). Using ratios of these eigenvalues allows the assessment of the linearity, planarity, sphericity, anisotropy, eigenentropy, omnivariance, scattering or change of curvature of the 3D shape (Chehata et al., 2009; Gross et al., 2006; Pauly, 2003). Estimates of the PC's **local point density** (Weinmann et al., 2013) or the **verticality** (Demantké et al., 2012) are other helpful parameters to classify points. **Multiple returns characteristics** associated with airborne lidar data also constitute information on the objects surveyed: the number of returns, return number or ratio of both are useful for identifying ground, buildings or vegetation (Chehata et al., 2009). **Height-derived features** such as elevation variations between points of a neighbourhood or point distribution kurtosis or skewness are also used for classification purposes (Antonarakis et al., 2008; Chehata et al., 2009; Guan et al., 2012; Yan et al., 2015). They can be combined with **distance-to-ground features**, corresponding to the distance between the points to classify and the ground, through the analysis of a digital terrain model, for example (Blomley et al., 2017; Chehata et al., 2009; Niemeyer et al., 2012). Another possibility to assess the geometrical characteristics of point clouds locally is to use **histogram-based features** that analyse the variations of geometrical features around the point through histograms, whose bins are used as descriptors (Blomley et al., 2016; Blomley et al., 2017; Himmelsbach et al., 2009; Osada et al., 2002; Rusu et al., 2009; Tombari et al., 2010; Wohlkinger et al., 2011). However, computing these features experimentally demands a higher computation time (Garstka et al., 2016).

Though some studies solely exploit PC geometry to identify 3D objects (West et al., 2004), the **radiometric information** contained in lidar data can further improve PC interpretation where objects have similar geometries (Yan et al., 2015). Radiometric information is rarely used on its own (Song et al., 2002) and is often integrated as a complement to previously mentioned geometrical features. It is often among the most contributive features to improve segmentation (Dai et al., 2018) and classification results (Im et al., 2008). The most popular attribute is the **mean value of the backscattered intensity** over a neighbourhood or between the first and last returns (Antonarakis et al., 2008). Combining **multispectral radiometric measurements** provides even more reliable information than single wavelength data (Morsy et al., 2017b). Using multispectral lidar systems allows to incorporate **intensity ratios** – for example, vegetation indexes – to classification predictors (Chen et al., 2017; Morsy et al., 2017b; Wichmann et al., 2015) or to compare surface reflectances in different optical domains (Chen et al., 2017; Gong et al., 2015) and even create colour composites with different channel combinations (Wichmann et al., 2015), thus refining point identification (Im et al., 2008). However, these existing methods perform nearest neighbour interpolations of different intensities. Among the reviewed features, no solution to formalize

the sampling differences of multi-spectral point clouds exists.

Features extraction

3D point clouds are unordered and have varying densities. Also, points are 0-dimensional and thus do not contain any meaningful geometrical information other than their Z coordinate. For these reasons, **descriptors are computed on the neighbourhood of each point**, which describes the local geometry. The **spherical neighbourhood** is the most common for PC processing, defined by its radius or diameter and comprising each point's nearest neighbours with respect to 3D Euclidean distance. **Cylindrical neighbourhoods** are also exploited by Niemeyer et al., 2012, and **cubic** or **cuboid** ones are explored by Dong et al., 2017. Overall, spherical neighbourhoods are considered the most helpful, based on the observations of Thomas et al., 2018 and Hermosilla et al., 2018, which compared the use of **nearest neighbours** (NN) and spherical searches to describe PCs. They are considered more stable than NN to the variations of density (Hermosilla et al., 2018), surface slope or orientation and point pattern that occurs in PCs, and more efficient for handcrafted feature extraction (Thomas et al., 2018). Thomas et al., 2019 additionally states that a consistent spherical domain helps classifiers learn more meaningful representations of the local aspect of the PCs during training.

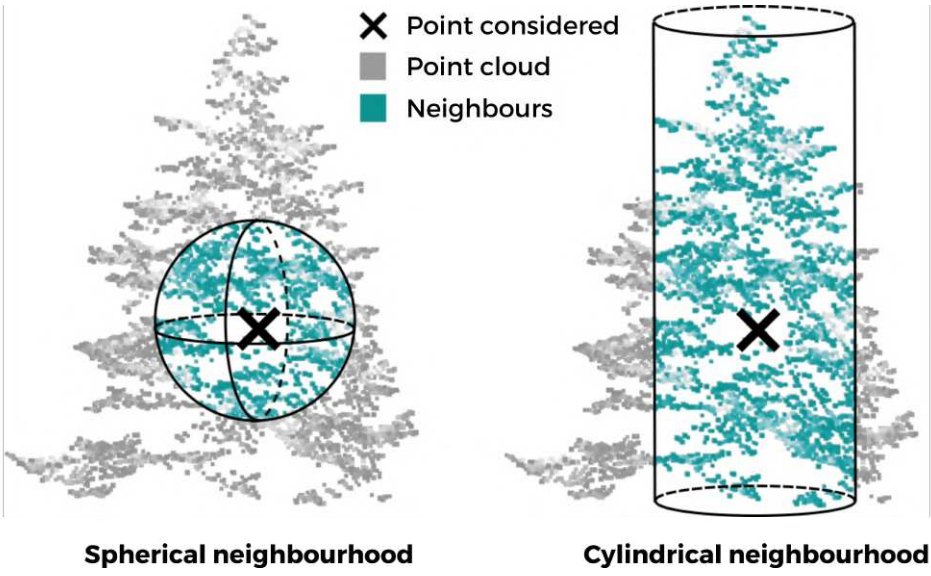


Figure 2.5: Illustration of spherical and cylindrical neighbourhoods for 3D point cloud processing.

Independently from the type of neighbourhood implemented, descriptive features of 3D data can be computed at a **single constant scale** (Chehata et al., 2009) or **multiple scales** (Blomley et al., 2017; Brodu et al., 2012; Hackel et al., 2016; Hackel et al., 2017; Niemeyer et al., 2012).

Multiple scales successively applied to each point have proven to have greater descriptive power than a single constant scale since they can better capture scene elements of different sizes (e.g., vegetation) and the variations of object geometry with scale (Brodu et al., 2012; Hackel et al., 2016; Hackel et al., 2017; Thomas et al., 2018). Considering the diversity of objects in PCs, the neighbourhood type, the number of scales used, and their values impact the classification of the data and thus require careful parameterization.

Automatic **optimal scale identification** has been investigated to avoid empiric selection. It mainly relies on minimizing information redundancy – through correlation or entropy estimates – and maximizing relevance in terms of classification accuracy. For single-scale classification, Niemeyer et al., 2011 advised an optimal scale of 7 NN in terms of classification accuracy when classifying urban scenes with lidar data. Rather than defining a fixed set of multiple scales, Demantke et al., 2012 try to identify automatically the most relevant scale to describe each point's neighbourhood by using its dimensionality. Similarly, Weinmann et al., 2015 select each point's individual optimal scale before extracting and selecting descriptive features. These approaches combine the use of multiple scales across the PC and the computation of features at a single scale for each point. Dong et al., 2017 propose to select an optimal neighbourhood type and its scale for each feature rather than optimizing the scale for each point, thus combining the advantages of different types of neighbourhoods, multiple scales and uncorrelated features.

3D points classification

Many classification algorithms have been developed to classify 3D PCs. The most common ones **classify each point individually** without considering the relationships between the point's label and its neighbour's assigned labels. They include **instance-based techniques** such as NN classification, rule-based predictions as applied by decision trees, probabilistic learners like Maximum Likelihood, max-margin learners as Support Vector Machines, and ensemble learning (Kotsiantis et al., 2007; Sagi et al., 2018). **Ensemble learning** (Sagi et al., 2018) is the most popular among individual point classification strategies. It relies on bagging, which consists of assembling several independent weak learners and combining them into a single strong learner using a voting mechanism. Random Forest (RF) models implement ensemble learning. Their ease of use, efficiency, robustness to overfitting, generalization abilities and production of a feature importance metric (Breiman, 2001; Pal, 2005) explain their frequent use for 3D data classification. They have been used for point-based classifications of both topographic and TB lidar (Chehata et al., 2009; Hansen et al., 2021; Letard et al., 2022b; Letard et al., 2022a). In RF, since the decision trees are independent, one cannot compensate for the potential weaknesses of another to improve the global performance of the forest. Algorithms like AdaBoost (Hastie et al., 2009) and XGBoost (Chen et al., 2016) overcome this limitation by incorporating **boosting**, which consists of training each weak learner to correct their predecessor's errors; however, they require setting additional parameters, such as the rate

at which the weak learner's parameters are updated, or when to stop optimizing. Most of the time, the multiple iterations needed to optimize the decision trees can also not be performed in parallel, since they rely on previous training iterations, thus increasing the global computation time.

Individual point classifiers can only consider the spatial context of each point by encrypting it into the feature vector. However, they ignore that neighbour points' labels tend to be linked. Some algorithms thus implement **contextual classification**, which involves estimating the relationships between 3D points from a neighbourhood – often different from the one used for feature extraction – in the training data. Additionally to the classification performance objective, they aim to produce spatially consistent classifications of 3D PCs, avoiding the noisy output that individual point classifiers can produce. Thus, they tend to reach higher accuracies. Examples of such approaches are applications of Associative (Munoz et al., 2008; Triebel et al., 2006) and Non-associative Markov Networks (Najafi et al., 2014), Conditional Random Fields (Lim et al., 2009; Niemeyer et al., 2012; Niemeyer et al., 2011; Vosselman et al., 2017), and Markov Random Fields (Lu et al., 2012) to 3D data. However, modelling 3D spatial relationships is computationally intensive and thus challenging to apply to large 3D datasets. These approaches also depend on the relationships observable in the training data, which makes exact inference of correlations between labels unattainable. In Landrieu et al., 2017, a structured regularization framework allowing the conservation of a probabilistic approach and relying on a computationally lighter optimization is thus proposed, allowing the regularisation of any point-based classification with contextual information while keeping a form of precision information.

2.2.3 Supervised classification of learned features

In general, neural networks are mainly based on linear combinations or convolution operators. When 3D data emerged, their processing with neural networks posed many conceptual and computational issues. One first issue is the **heavy computations** needed to load and process data in three dimensions, which is progressively solved by technical advances in computer hardware. A second, major issue, was to adapt proven methods to **unstructured, irregular 3D data**. As a result, although convolutional neural networks (CNN) are among the most performant deep neural networks, partly thanks to their ability to extract high-level features while considering spatial context, their **application to 3D PCs is not straightforward**.

To process 3D PCs, the first methods thus converted them into structures that could be processed by other proven 3D networks. Typical examples are the so-called **projection networks**, like SnapNet (Boulch et al., 2018), that convert point clouds into multiple 2D snapshots that are segmented individually and then reprojected in 3D using the principles of stereoscopy. The first deep learning methods on unordered 3D data were PointNet and PointNet++ - an enhanced version of PointNet -, two architec-

tures relying on **multi-layer perceptrons** published in 2017 (Qi et al., 2017b; Qi et al., 2017a).

The development of **3D convolutional networks** is more recent and includes a wide range of variations to optimize performances and complexity. Some networks perform convolution using transformed points (Atzmon et al., 2018; Hua et al., 2018; Li et al., 2018; Xu et al., 2018), voxelized PCs (Graham et al., 2018; Tchapmi et al., 2018), or **graphs**.

Graphs can be derived from adjacencies between points (Mao et al., 2022c; Wang et al., 2019a; Wang et al., 2019c; Wei et al., 2023; Wen et al., 2021) or groups of points (Hui et al., 2021; Landrieu et al., 2019; Landrieu et al., 2018) – called **superpoints**. The main example of such superpoints-based networks is SPG (Landrieu et al., 2018) – for SuperPoint Graph – a network that performs graph convolution on a graph whose edges are partitions of geometrically similar elements of the input point clouds. The obtained partitions are embedded using PointNet, resulting in edges of an oriented graph with attributes that is then segmented with graph convolution networks.

The last type of 3D convolutional networks relies on **kernel points**, introduced in KPConv (Thomas et al., 2019), that replicate in 3D the principle of the 2D convolution operation, that uses a weighted kernel. Currently, KPConv (Thomas et al., 2019) and SPG (Landrieu et al., 2018) are among the state-of-the-art architectures for 3D points classification.

However, research around 3D deep learning is very active and working towards different improvements of existing solutions. Examples of recent experiments include the use of spatially sparse convolutions (Graham et al., 2018; Schmohl et al., 2019), that are specifically designed to handle the sparsity of PCs. Some architectures also incorporate state of the art 2D deep learning solutions to 3D networks, such as **attention mechanisms** (Deng et al., 2021; Huang et al., 2021; Wang et al., 2019a; Zeng et al., 2023; Zhang et al., 2022) or residual connections (Huang et al., 2018; Ye et al., 2018; Zeng et al., 2023). In a similar attempt to adapt 2D principles to 3D, multiple projects combine several receptive fields as in inception networks to improve 3D processing (Mao et al., 2022a; Mao et al., 2022c).

3D deep neural networks have also started being applied to airborne lidar surveys (Huang et al., 2021; Lin et al., 2021; Mao et al., 2022a; Schmohl et al., 2019; Wen et al., 2021; Yang et al., 2018; Zeng et al., 2023; Zhang et al., 2022; Zhao et al., 2018a). However, existing work on airborne lidar processing with 3D deep neural networks concerns urban areas or forests exclusively and thus **does not address the issue of complex and diverse natural or semi-urban environments**.

Current evolutions also tackle the biggest challenges faced by 3D deep learning: **computational complexity, applicability to large datasets, and limitations linked to the amount of labelled training data** required. The first motivation for the development of SPG is the constatation of the authors that "Deep learning architectures specifically designed for 3D point clouds [...] display good results, but are limited by the size of inputs they can handle at once". They thus introduce SPG to propose a method

with lighter computation able to perform on large datasets, since the superpoints, similarly to superpixels, are orders of magnitude less numerous than the number of points of the input point clouds. SPG thus offers a method that is applicable to real, large datasets.

More recently, **adaptations of the Transformer architecture** on 3D data have been investigated to build on the performances offered by the attention mechanism while designing lightweight architectures (Hu et al., 2020; Zhao et al., 2021a; Guo et al., 2021a; Lai et al., 2022; Park et al., 2022; Cheng et al., 2023; Robert et al., 2023). For example, the Transformer-based version of SPG: SuperPoint Transformer, reaches state-of-the-art performance while drastically reducing the number of parameters to optimize compared to other networks (Robert et al., 2023). Such solutions - often combining high-performance architecture modules based on attention with sampling techniques that counterbalance the computational cost associated with the original Transformer - allow the reduction of the computational cost of 3D neural networks and thus make them more usable in practice. Another possibility is to use networks **performing on features rather than points** for faster, lighter alternatives (Gao et al., 2023).

To simplify the use of 3D deep neural networks, the development of approaches to train the algorithms on a **limited amount of labelled data** also currently arises. An option is to rely on **few-shot learning**, which consists of training models using a small amount of labelled training samples, thus drastically reducing the number of labelled samples required to train (Dong et al., 2018; Feng et al., 2022; Garcia et al., 2018; He et al., 2023; Li et al., 2022; Mao et al., 2022b; Xu et al., 2020; Zhao et al., 2021b). **Domain adaptation** is also a possible strategy to reuse previously learned weights to process new, unlabelled data (Cazorla et al., 2022; Jaritz et al., 2023; Yuan et al., 2023). Experiments of **unsupervised representation learning** on 3D point clouds also exist, using the encoding of contrastive transformations of the input point clouds (Jiang et al., 2021), or optimal transport-based clustering and contrasting processes (Mei et al., 2022).

2.2.4 The specific case of bispectral lidar point clouds

TB lidars have the specificity of embedding two different non-co-focal lasers, thus **producing two distinct samplings of the same scene** under the form of two separate PCs whose points systematically have different positions. Figure 2.6 illustrates the sampling difference in submerged areas that can be used to characterize them in point clouds.

At this time, no research on the application of 3D deep neural networks to TB areas surveyed with airborne bispectral lidar has yet been published. Existing approaches rely exclusively on handcrafted features extracted on full-waveforms (Launeau et al., 2018; Leigh et al., 2016), rasters (Wedding et al., 2008; Laslier et al., 2019), or PCs (Hansen et al., 2021).

There is, however, more research on the classification of multispectral airborne lidar data over terres-

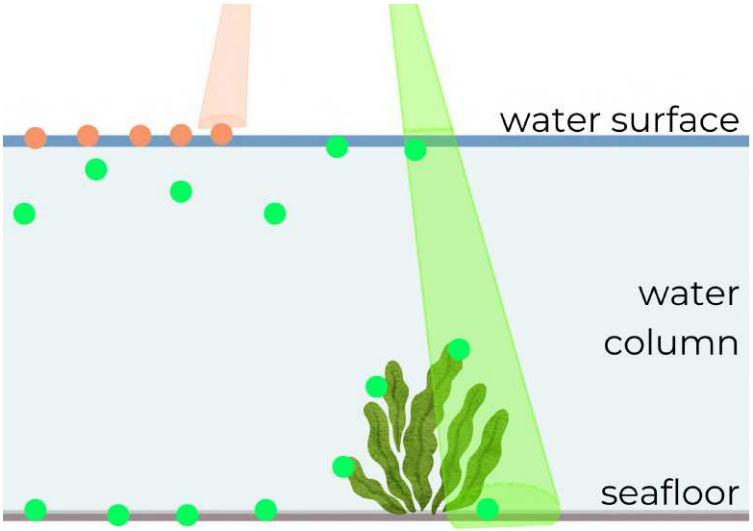


Figure 2.6: The sampling differences of infrared and green lidars over submerged areas.

trial areas, using 3D PCs (Morsy et al., 2017b; Morsy et al., 2017a; Morsy et al., 2022; Morsy et al., 2016; Wang et al., 2020a), including an application of PointNet++ over mixed urban areas (Jing et al., 2021). In this paper, however, the multispectral character of the data is summed up in one single point cloud with several intensity attributes, obtained using **nearest neighbour interpolations** over the three available spectral channels.

There is thus no direct processing of multiple PCs simultaneously, as in Ekhtari et al., 2018, in which the authors create a **synthetic individual multispectral point cloud** from the three PCs of their multispectral survey and use it for classification. Other approaches exploit the information contained in each point cloud of each wavelength individually and use it contiguously to perform classification (Wang et al., 2020a).

Another existing possibility consists in computing **spectral ratios** between the different spectral channels to exploit the multiple wavelengths of the sensors, as in (Matikainen et al., 2017; Morsy et al., 2022; Morsy et al., 2017b; Morsy et al., 2017a; Morsy et al., 2016; Shaker et al., 2019).

However, existing papers often do not include a wide range of possible classes, sometimes only tackling the problem of land/water distinction (Shaker et al., 2019).

2.2.5 Limitations of existing classification methods for environmental lidar point clouds

Although 3D point cloud processing has drastically progressed over the last fifteen years, some limitations remain, hindering the exploitation of airborne lidar point clouds at their full potential.

Regarding supervised classification of handcrafted features, there is **still a lack of methods involving sufficiently varying and exhaustive features to describe the many facets of complex environmental point clouds**. Urban point clouds or forest point clouds are widely studied, but scenes mixing a great variety of classes with very different characteristic scales and shapes still lack representation in existing research and remain difficult to automatically analyse. This results in many environmental science studies either not exploiting existing lidar surveys, or reducing them to rasters, thus losing the precious information of spatial structure and point repartition.

Although descriptive features adapted to a great variety of environments and shapes exist separately in multiple explorative studies, **an accessible and applicable framework unifying the different perspectives of urban, forestry, other topographic environments, and bathymetric objects is needed** to provide better tools for environmental studies to be improved with 3D knowledge.

Deep learning can be a solution to the issue of representation of the many aspects of environmental point clouds. However, its **large-scale applicability is still limited** by several factors. The computational requirements of 3D deep learning necessitate specific hardware settings, which limits their accessibility to a wide range of environmental researchers. Their increased complexity in handling, optimising and parameterising without extended knowledge about machine learning, scientific programming and mathematics is another factor limiting their deployments for 3D analyses by thematic users. This complexity also limits their interpretability and explainability, which is a key aspect of understanding uncertainties in environmental studies. Finally, the difficulty of training networks to perform the desired tasks is a major drawback.

There is indeed a great **lack of foundation models to analyse 3D data in natural environments**. Most of the available models and experiments for airborne lidar are made with benchmark data acquired in urban environments. Few data in forested areas are available, and none in mixed, natural environments including built-up areas, densely vegetated zones, and submerged surfaces. However, supervised classification implies the construction of a labelled dataset to be used for training and testing the classifiers. The complexity of PC processing makes point labelling fastidious and time-consuming. A solution to train supervised methods despite the lack of benchmark data can be to simulate lidar PCs, as featured in (Gélis et al., 2021) for change detection and classification, but it requires yet again the availability of simulators adapted to natural environments and does not really solve the problem of classifying real lidar data acquired over complex natural areas.

Developments in that direction are currently numerous, and this area of deep learning research is very active, as illustrated in Section 2.1.4. Nevertheless, these developments are still recent and, although highly performing, 3D deep neural networks are still experimental and less accessible to non-specialist users than more classical machine learning algorithms. The recent advances towards lower supervision and smaller amounts of labelled data still show lower performances than fully-supervised approaches and remain experimental, and thus not deployable to operational data processing at a large scale, and by non-specialist users. Finally, there are, to our knowledge, no methods capable of considering two point clouds derived from two different wavelengths.

Overall, there is a **lack of available methods exploiting both the spectral and the geometrical differences between the different point clouds obtained through multispectral lidar surveys** for the classification of land-water interface areas. In these areas, the sampling of the environment obtained with NIR or green wavelength is extremely distinct both spectrally and geometrically, providing key information about the environment. Yet, currently, existing classification methods do not exploit it. There is thus a need for a framework that would allow the classification of 3D point clouds representing a variety of land use/land covers, and supporting their cohabitation with seabed use/seabed covers.

In the next section, we review existing lidar waveform processing methods, their specifics, and the challenges they still have to face.

2.3 Lidar waveforms processing in topo-bathymetric areas

2.3.1 Motivations and specifics of lidar waveforms processing

Lidar waveform processing first consists of extracting the targets present in the illuminated cone by detecting variations in the amount of energy received (Chauve et al., 2007). Then, the information contained by waveforms about the intercepted surfaces' geometry and reflectance can be used to study them more deeply.

Waveform processing is however a complicated process. **The size of the datasets** obtained is significant: for example, the Leica HawkEye-5 presented in Chapter 1 shoots up to 170 times per second, thus producing 170 waveforms containing between 512 (shallow bathymetry channel) and 1064 (deep bathymetry channel) samples at each second. Waveforms are also **particularly noisy** due to receiver internal noise and ambient light, which can be complicated to deal with. Lastly, as lidar **intensity measurements are sensor and acquisition-dependent**, specific measures need to be applied to compare waveforms. It is also worth noting that additional precautions such as **time and space matching** of waveforms recorded with different sensors, wavelengths or during different acquisitions are necessary since it is im-

possible to have strictly identical interceptions. In the following paragraphs we review existing methods to tackle normalization and denoising of lidar waveforms.

Normalization

Similarly to discrete lidar return intensities, lidar waveform values can be influenced by sensor characteristics and acquisition conditions (see Kashani et al., 2015 for details). Thus, in order to be able to compare waveforms and limit the bias of varying emitted intensities and different illumination conditions, waveform data must be normalized in pre-processing.

Contrary to typical machine learning tasks that involve data normalization among the complete training or test dataset, waveforms are **mostly normalized individually** to decrease or delete a maximum of elements that could cause bias (Lang et al., 2022; Pirotti et al., 2011). By normalizing each signal individually, signal patterns, pulse shapes and the relative amplitude differences between different waveform elements can be compared, and overfitting on specific cases with specific intensity ranges is reduced. Pirotti et al., 2011 and Lang et al., 2022 perform normalization by dividing each sample's returned energy from the total received energy, so that the integral of the waveform is equal to one, thus converting it into a distribution. Another option is to normalize each waveform between 0 and 1, which also allows the comparison of pulse shapes and reduces varying backscattered intensity bias.

Denoising

Currently, **noise is problematic for most of the existing waveform processing methods**. Denoising is even considered "a prerequisite" in Zhou et al., 2021. In the literature, several methods to reduce noise can be found. Many of them rely on **thresholding** the waveform amplitudes (Drake et al., 2002; Lang et al., 2022; Pirotti et al., 2011). The threshold is often determined using an estimation of the noise level by analyzing waveform amplitudes in sections of the signal that precede the first reflective element (Drake et al., 2002; Lang et al., 2022; Zhao et al., 2022). Some sensors record this noise level for each waveform, as NASA's Land Vegetation and Ice Surveyor (LVIS), so it can be deducted from waveform amplitudes (Pirotti et al., 2011; Zhou et al., 2021). For bathymetric waveforms, Yang et al., 2022 use an adaptative threshold depending on the return-to-noise ratio of the last return in the signal. Most of the time, threshold methods are based on a unique value for each waveform, as a uniform threshold to fit all signals is impossible to find, and a wrong value can cause over- or under-smoothing (Yang et al., 2022).

The other category of waveform denoising methods relies on **smoothing** the whole signal to reduce the noise-induced content, based for example on Wiener filters (Jutzi et al., 2006).

2.3.2 Detecting and locating objects in the signal

Waveform processing serves two objectives: the first and most popular one is to **detect intercepted targets** in the signal, and potentially to **increase the number of returns extracted** (Chauve et al., 2007; Mallet et al., 2009). The secondary aim is to **extract information on the environment surveyed** (Kumpumäki et al., 2015). Overall, it allows the extraction of more information from the lidar survey than range measurements and their associated intensity.

Waveforms are still rarely considered as a whole signal; they are mostly processed to separate the backscatters identified, which are then studied individually. To detect and locate each object hit by the laser, three approaches exist: signal peak detection, waveform decomposition, and waveform deconvolution. All three approaches allow the production of point clouds, but some also provide additional information on the objects behind the peaks.

Classical peak detection

Peak detection in denoised waveforms is often made using thresholding of their second derivative, or by monitoring their **first-derivative zero-crossings**. Another possibility is to **analyse the centre of gravity** of the second derivative or use a Ramer-Douglas-Peucker curve fitting algorithm to **detect inflexion points** (Qin et al., 2012). Peak detection is made on denoised signals to avoid overdetection of peaks because of high-frequency noise. This procedure is used to detect inflexion points in waveforms and estimate the number of surfaces that were intercepted by the laser. It is sometimes used as a starting point for more advanced procedures used to locate targets.

Waveform decomposition

Waveform decomposition consists of **modelling the signal** using different **functions to separate it into a sum of components**. Once decomposed, it becomes possible to extract statistical parameters from the obtained elements to characterize the objects behind them (Chauve et al., 2007). In practice, this consists of **estimating the parameters of a finite mixture model** (Mallet et al., 2010). The process implies three phases:

- **Components detection**, which consists of performing peak detection to identify the number of fittings to perform;
- **Initialisation**, which consists of estimating the initial parameters of the used model for each of the detected peaks;

- **Optimization**, which consists of finding the considered mathematical function's parameters that best fit the portion of the signal analysed, for each component. Non-linear least squares with the Levenberg-Marquardt algorithm or maximum likelihood estimates with the expectation-maximization algorithm are among the most popular optimization methods for waveform decomposition. Both depend on the initial values of the parameters and can be falsely induced into local optimums if they are faulty.

Decomposition of topographic waveforms Different mathematical assumptions can be used to describe the peaks, depending on the transmitted waveform (Jutzi et al., 2006). The most common one is to consider the backscatters to be **Gaussians**, and thus perform Gaussian decomposition on the denoised waveform (Pirotti et al., 2011; Qin et al., 2012; Yang et al., 2022; Zhu et al., 2012; Ma et al., 2019). In this context, a finite number of Gaussian components are fitted to the signal until their sum is close to the initial waveform. However, since **waveform components are not strictly Gaussian** (see Chapter 1), other more complex parametric models have been investigated to improve waveform processing possibilities. Standard extensions are skewed normal, lognormal, generalized Gaussian functions, or Weibull distributions (Bruggisser et al., 2017; Chauve et al., 2007; Fang et al., 2022; Montes-Hugo et al., 2016).

Gaussians can also be used in combination with point processes (Mallet et al., 2010), or mixed with other of the above examples to fit components with different shapes, for example in bathymetric cases where different shapes are represented (Kogut et al., 2019). More complex models, like B-splines (Liu et al., 2022a; Shen et al., 2017) or wavelet extraction, are also used to allow for non-Gaussian-shaped pulses.

The fitting functions are chosen **depending on the applicative context**, namely the transmitted pulse shape, and the size of the footprint (Mallet et al., 2010). Their parameters also constitute descriptive features of their characteristics, which vary with the chosen model. However, the more complex the model, the more difficult the optimization, as there are more parameters to update and a greater degree of freedom. In any case, the model has to be adapted to the shape of the transmitted waveform for the decomposition procedure to be successful.

Bathymetric waveform decomposition For bathymetric waveforms, **specific models** have to be used, as the global shape of the signal is significantly impacted by the **water column component**, which has an exponential shape (Wang et al., 2015). Consequently, obtaining an accurate decomposition requires the use of mixture models including adapted functions. In most proposed approaches, the air/water interface and the water bottom components are approximated with the same function, which is different from the one used for the water column.

For the water column, one of the first propositions was to use **triangle functions** (Abdallah et al., 2013). Models were then progressively improved to increase decomposition performance, and thus

bathymetry detection. **Quadrilateral functions**, for example, allowed to improve the detection when combined with Gaussians for the seabed and water surface (Abady et al., 2014; Ding et al., 2018). Another option is to build on the inherent exponential shape of the signal decay and use **exponentials** to perform decomposition. For example, Schwarz et al., 2017 use multiple concatenated exponentials to decompose bathymetric waveforms. In Schwarz et al., 2019, they propose to combine two rectangle functions with an exponential model to isolate the air/water interface, the seabed, and the water column component, respectively, as illustrated in Figure 2.7.

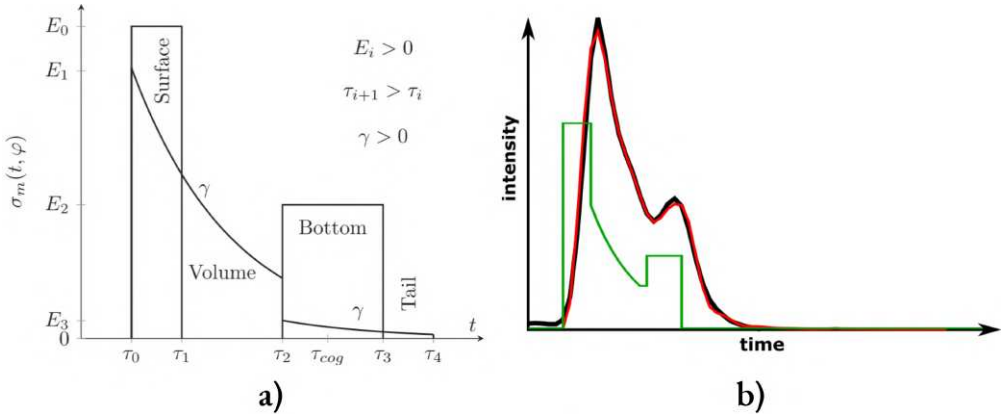


Figure 2.7: Illustration of the waveform decomposition method proposed in Schwarz et al., 2019: a) model with 10 parameters $E_0 \ldots E_3$, $\tau_0 \ldots \tau_4$ and γ to estimate; surface and bottom are modelled by boxcars to account for beam spreading caused by off-nadir incidence; effects from the water volume are modelled by two exponential segments; the tail below the bottom takes care of late echoes, likely caused by multiple scattering; τ_{cog} is the centre of gravity of the bottom and tail segments with respect to τ_2 (Schwarz et al., 2019). b) Illustration of the application of this method in Richter et al., 2021.

Two main problems specific to bathymetric waveform decomposition must be kept in mind (Xing et al., 2019):

- Separating overlapping echoes in extremely shallow waters;
- Detecting weak seabed returns.

The methods presented above (Abady et al., 2014; Ding et al., 2018; Schwarz et al., 2017; Schwarz et al., 2019) show good results in shallow waters. However, in cases when the seabed return remains too weak, it can still be difficult not to falsely mistake noise for a weak backscatter. Several solutions have been proposed to tackle this issue.

The first is to **enhance the signal level**, for example by correcting the effects of signal attenuation in water (Richter et al., 2017), or by smoothing out the detector and ambient light noises (Launeau et al., 2018; Saylam et al., 2017; Zhao et al., 2022).

Another possibility is to use **waveform stacking** techniques, namely **orthowaveforms** processing methods. They consist of stacking spatially close waveforms to increase the signal-to-noise ratio. In the case of orthowaveforms, this stacking is performed vertically using voxels to take the beam trajectory into account, as illustrated in Figure 2.8 extracted from Richter et al., 2022. The mean backscattered intensity of each voxel is then kept to obtain a spatially averaged and ortho-rectified waveform. Mader et al., 2019; Mader et al., 2021; and Mader et al., 2023 illustrate how such techniques can make weak returns much easier to identify, and can reduce the computational complexity of seabed detection by reducing the range of the waveform in which to search for the water bottom return.

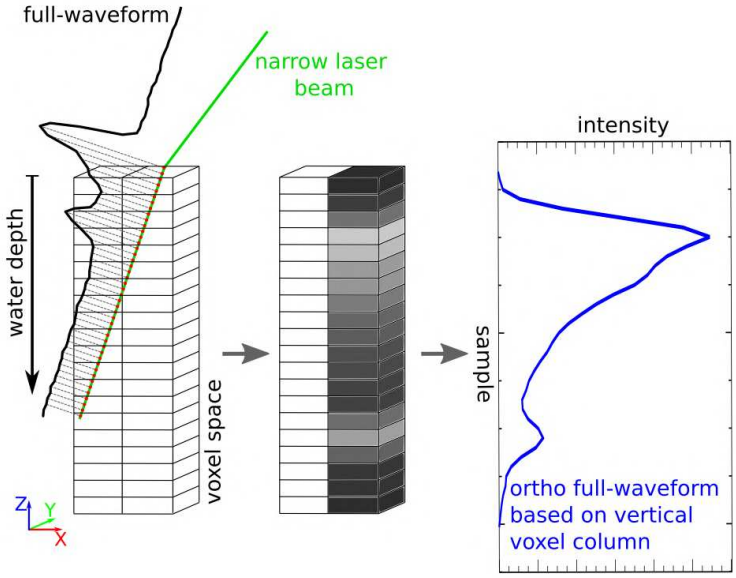


Figure 2.8: Principle of ortho full-waveforms generation, extracted from Richter et al., 2022.

For the separation of overlapping echoes, namely in extremely shallow water areas, a possibility is to work on the temporal sampling of the waveform. Using **temporal resolution augmentation** models as in (Yang et al., 2022) can help identify when one component ends and the other starts.

Since several studies (Parrish et al., 2011; Zhou et al., 2021; Xing et al., 2019) point out that there is no universal processing strategy, most of the existing bathymetric waveform decomposition methods are **designed for typical use cases**. Consequently, (Xing et al., 2019) suggests that the best alternative is to design methods adapted to different ranges of depths, rather than attempting to find a global method. The authors of Xing et al., 2019 indeed state that the high variability of depths and attenuation processes, in reality, is too difficult to model.

General fitting procedure Waveform decomposition is **iterative**, as peaks are fitted one after another. After the first fitting, the difference between the denoised waveform and the first fitting is analysed. If additional peaks are detected in it, a new optimization phase is launched. This allows the detection of additional echoes that could not be identified with peak detection initially. The process is repeated until the difference between the denoised waveform and the sum of the estimated components is low (Song et al., 2019). Examples obtained with different decomposition methods after optimization are presented in Figure 2.9, modified from Zhou et al., 2019.

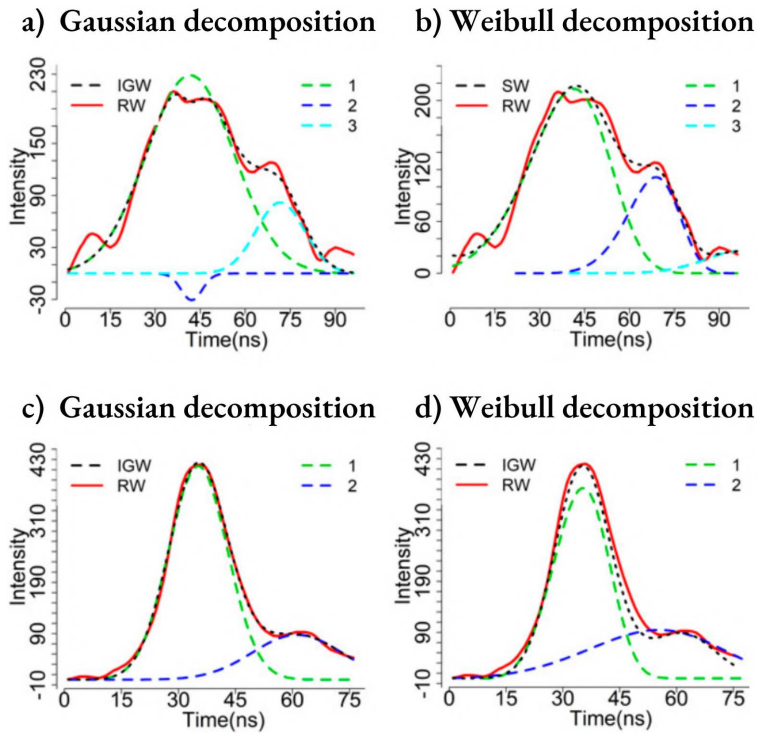


Figure 2.9: Modified from Zhou et al., 2019: comparisons of results using Gaussian (a),c)), and Weibull decompositions (b), d)) for two different waveforms. The black dashed line is the modelled waveform (IGW), the red solid line is the raw waveform (RW), and the dashed and numbered lines correspond to the detected components.

To avoid false detections, the results of waveform decomposition are sometimes filtered out, for example by comparing the full width at half maximum of the obtained component and that of the transmitted pulse, or by comparing its amplitude to the standard deviation of the waveform's background noise (Parrish et al., 2011; Song et al., 2019).

Waveform deconvolution

While decomposition analyses waveform components as they are received, another possibility is to reverse the effects of the transmitted pulse to access the intercepted surface characteristics independently from the influence of the sensor. As **the received signal is a convolution** between the transmitted pulse and the target response function (Chauve et al., 2007; Jutzi et al., 2003), performing **deconvolution theoretically gives access to the original intercepted surfaces' signatures**. For that, the system's impulse function must be known, which is often the case with full-waveform systems. If transmitted pulses are not recorded, the typical pulse function of a sensor can be estimated by analyzing waveforms resulting from shots on plane surfaces at a low incidence angle.

Deconvolution can be performed using the **Wiener filter** (Jutzi et al., 2006). However, this often requires a denoising step upstream, as this method is sensitive to noise. The other popular deconvolution method for waveform processing is the **Richardson-Lucy algorithm**, which is more robust to noise and allows to retrieve the main components of the waveforms successfully (Parrish et al., 2011; Wang et al., 2015; Wu et al., 2011; Zhou et al., 2021). The **Gold deconvolution** (Morháč et al., 2003) method is another alternative explored in Zhou et al., 2017, where the authors highlight its performances in dense vegetation and its low tendency to false detections.

Although these methods come at the expense of a higher computation cost, all comparative studies on waveform processing tend to agree on the superior performances of the Richardson-Lucy deconvolution, which often misses less backscatters (Parrish et al., 2011; Wang et al., 2015; Wu et al., 2011; Zhou et al., 2021).

However, it must be noted that **deconvolution techniques do not provide the same information obtained with waveform decomposition**. They consist more of a signal enhancement strategy, but output a processed signal and not discrete echo locations and parameters. They are consequently **used in combination with peak detection or decomposition** strategies (Xing et al., 2019; Zhou et al., 2017), which become easier to use because the signal is significantly denoised. An advantage is that, in theory, decomposing a deconvolved waveform gives access to the surface response function. On the contrary, when using decomposition alone, the fitted components are still the result of the convolution of the system impulse and the surface response. Comparisons of decomposition and deconvolution approaches for satellite lidar waveforms processing also highlight the superiority of deconvolution to identify several vertical elements, namely in vegetated areas, thus describing canopy structures better (Neuenschwander, 2008).

Deep learning for waveform return detection

Few studies use machine learning to perform lidar return detection in the waveform data. In (Aßmann et al., 2021), a **1D CNN encoder** is built to extract the **locations of peaks** in waveforms and thus detect

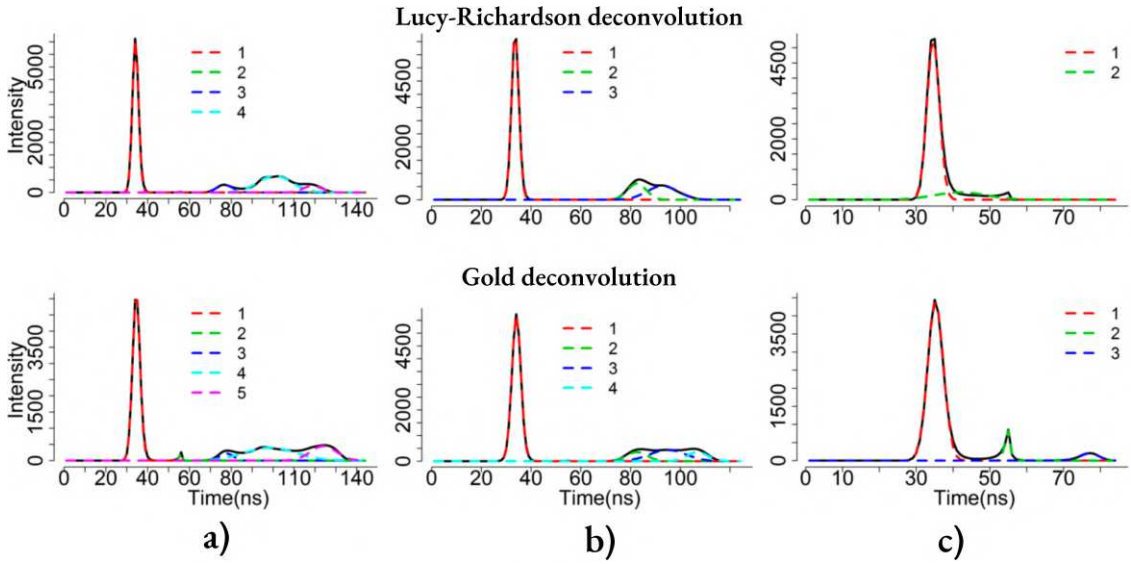


Figure 2.10: Illustration of waveform decomposition and deconvolution Zhou et al., 2019: Comparisons of the decomposition results with Lucy-Richardson approach and Gold approach for three sample pulses (a) to c)). The solid black line is the original waveform. The coloured dash lines are Gaussian components after Gaussian decomposition of the deconvolution result.

intercepted surfaces. A similar network was exploited in (Liu et al., 2021b) to **estimate the water depth** from bathymetric waveforms automatically. Finally, (Liu et al., 2022b) use **dense and residual networks to automatically estimate waveform peak location and amplitude**. Overall, these three investigations of the abilities of deep neural networks to perform waveform processing find high-quality results and show the possibility of using such approaches on large waveform datasets.

2.3.3 Identifying objects in the signal / extracting semantic information

Waveform classification may consist of attributing one unique label to each waveform or classifying each target in the waveform. In the latter case, processing can be closer to **semantic segmentation** or **object detection and classification** rather than usual classification. Similarly to point cloud classification, full-waveform classification can be made with pre-defined features, or with learned features.

Lidar waveforms feature extraction and classification

The most commonly used approach to identify the intercepted surfaces using waveforms is to **classify waveform features** with classical machine learning algorithms, namely Support Vector Machine (Launeau et al., 2018), Random Forests (Narayanan et al., 2009), or other algorithms taking spatial context into account such as CRF (Niemeyer et al., 2011). Often, the **features obtained through waveform decom-**

position and the optimization of finite mixture models on the components are used to describe the geometrical and optical parameters of the objects hit. Although the parameters derived depend on the model used, typically, they include the width of the peak or its full-width at half maximum, its amplitude, but also sometimes its skewness and kurtosis (Bruggisser et al., 2017).

Peak width informs mostly on the **geometry**, i.e. the slope and roughness of the intercepted surfaces (Song et al., 2019), but not about its material (Heinzel et al., 2011; Wagner, 2010). **Amplitude-related** statistics, on the other hand, are more linked to the **material** of the targets. However, these observations remain empirical, and some papers consider the geometric and radiometric properties of the targets too correlated in the waveform shapes to be accessible separately (Chauve et al., 2007).

Once peaks are identified, other parameters can be computed, even if they are not related to the function used for decomposition. Usual parameters computed to describe waveforms include peaks-wise parameters and global signal characteristics. On peaks, area-under-curve, standard deviation or variance, mean or median amplitude, skewness, or kurtosis are common parameters (Heinzel et al., 2011; Mallet et al., 2011; Narayanan et al., 2009; Neuenschwander, 2009; Niemeyer et al., 2011). The location at which the waveform amplitude overcomes the noise threshold, the location of the last component's centre, and the difference between both are other informing features about the global nature of the hit area (Pirotti et al., 2011; Drake et al., 2002). Over the whole waveform, the number of components hit, and spatial filters about the neighbouring waveforms are examples of classification predictors (Heinzel et al., 2011). However, among these examples, (Heinzel et al., 2011) identify peak intensity – i.e. amplitude – as the most important parameter for classification, along with the median peak width and the mean total number of targets within a waveform. Similarly, Neuenschwander et al., 2008 cite the total integrated waveform energy as an efficient feature to discriminate vegetation from built surfaces. It is interesting to note that Neuenschwander, 2009 also observe higher accuracy when performing the same landcover classification task using waveform features than classical satellite imagery.

Applications to land-water surface covers classification The specific task of **classifying land and water surface covers** with lidar waveform features has been explored multiple times (Teo et al., 2017; Collin et al., 2012; Eren et al., 2018; Tulldahl et al., 2012; Hansen et al., 2021; Schmidt et al., 2014; Wang et al., 2007). Up to now, existing approaches only include **classical machine learning combined with handcrafted feature extraction**, except for land-water separation, which has been explored with deep learning, as the next section will explain.

Several trends emerge from the literature on this subject:

- There are **many approaches classifying either terrestrial or marine classes, but very few study both simultaneously**. For example, there are many applications of waveform feature clas-

sification to identify vegetation species or trees (Kinzel et al., 2013; Bruggisser et al., 2017), but even when applied to the characterization of coastal vegetation, they solely tackle emerged vegetation (Nayegandhi et al., 2006). The same can be observed for marine covers. Wang et al., 2007 analyse waveform features in shallow coastal areas to derive seabed type variations, thus mapping seagrass meadows and underwater sand at depths between 0.8 and 4.3 m, while Hansen et al., 2021 use topo-bathymetric lidar data but only classify emerged boulders of a coastal environment. Similarly, Schmidt et al., 2014 use waveform features with a contextual random field and a random forest to separate mussel bed, land, and water of a coastal zone on emerged areas only. A last example is the use of topobathymetric lidar for coral reefs morphology assessments (Wilson et al., 2019), which shows again how topo-bathymetric lidar is often used for separate marine and terrestrial surveys. Only Collin et al., 2012 classify a large range of terrestrial and marine classes simultaneously in 2D rasters using a bathymetric lidar.

- **Some studies only aim at separating water and land**, for example Guo et al., 2023 and Liang et al., 2022.

- **Many methods include a rasterization step**: most of the published research on lidar waveform processing in land-water areas produces 2D outputs (Wedding et al., 2008; Teo et al., 2017; Collin et al., 2012; Eren et al., 2018), or simple analysis of correlation between measured characteristics and waveform features (Wilson et al., 2019; Nayegandhi et al., 2009). Few publications exploit the spatial wealth of lidar data and render classifications in 3D (Hansen et al., 2021; Tulldahl et al., 1999).

- **Only two approaches propose the use of two wavelengths**. Although several previously cited papers exploit surveys relying on dual-wavelength systems (Wilson et al., 2019; Hansen et al., 2021), only Collin et al., 2012 and Leigh et al., 2016 exploit both wavelengths. Among them, Leigh et al., 2016 proposes to use waveform features extracted from orthowaveforms from both wavelengths, while Collin et al., 2012 only extracts features from the green waveforms considered to be in the water.

Other studies such as (Butler et al., 2020; Grande et al., 2009; Chust et al., 2010) used 2D lidar-derived data and imagery along with machine learning classifiers to map coastal land- and sea covers. However, they did not exploit the full-waveforms and obtained low accuracy when classifying only lidar digital elevation models, and thus had to exploit additional imagery.

Feature learning on lidar waveforms

Feature learning directly on lidar waveforms without previous transformation into images or voxels **has not been investigated much**. Existing research is limited to classification – one label per waveform – of urban environments surveyed with topographic lidars (Zorzi et al., 2019; Hu et al., 2019), or worldwide regression of canopy heights with a 1D CNN (Lang et al., 2022).

Some approaches consider individual waveforms only, such as FWNet, using a PointNet encoder on waveforms and their first return's coordinates (Shinohara et al., 2020). In Hu et al., 2019, waveforms are also used individually to discriminate shallow water and land with accuracies between 98% and 99% when comparing a 1D fully connected network and a 1D CNN, respectively.

In Liang et al., 2022, a sense of spatial consistency is introduced by associating waveforms obtained using eight different lidar channels with varying footprints in an ensemble learning approach using one 1D CNN per channel and a voting system to perform ocean-land waveform classification with 99% of resulting accuracy.

Waveforms can also be associated with traditional lidar data rasters to improve the obtained results. 1D CNN predictions on waveforms are used in combination with DEMs and 2D CNNs in Zorzi et al., 2019 to combine the wealth of spatialized representations and waveform data for urban environment classification.

2.3.4 Estimating environmental parameters from lidar waveforms

We showed the use of waveform decomposition parameters as proxies for categorical variables extraction. **Waveform decomposition can also open opportunities for regression approaches with the objective of estimating environmental parameters**. With the advent of satellite lidar sensors, this becomes even more promising, as worldwide products of water optical properties and seabed or vegetation characteristics become feasible (Lu et al., 2022).

Regression can be performed on topographic waveforms, mainly to extract estimations of vegetation parameters, or on bathymetric waveforms, in which case the regression typically targets the optical properties of water.

Topographic waveforms

Vegetation parameters can be derived from topographic waveforms with two main approaches:

- **Radiative transfer models inversion**; for example, Koetz et al., 2006 inverse a 3D radiative transfer model (RTM) to derive forest biophysical parameters from the interactions between the infrared laser and the canopy structure. Their approach is tested on synthetic and real data, and they

manage to successfully retrieve parameters such as leaf area index (LAI) or vertical crown extension. Such methods are common in forestry, but can also be applied to other types of vegetation, namely crops. Hmida et al., 2017 show that RTM model inversion can be used to estimate the height, LAI and ground reflectance for different cultures - namely maize and wheat - and at different phenological states.

- **Regression based on waveform decomposition parameters**; Adams et al., 2012 show that correlations exist between waveform shape parameters and the LAI in forested areas. They mainly outline the role of the decays of energy in waveforms acquired over coniferous foliage, and the width of the decomposed peaks. They suggest that such parameters are more reliable for LAI estimation than intensity-based parameters that are strongly affected by shadowing. Similarly, Hancock et al., 2015 explore the use of waveform decomposition methods to estimate the reflectance of vegetation for different types of lidar systems.

Bathymetric waveforms

Over water extents, bathymetric lidar has been shown to be an efficient tool for the **estimation of optical parameters linked to the water body**. These parameters are mostly linked to **the way water attenuates or scatters light**. The **diffuse attenuation coefficient**, K_d, which is the factor of the exponential attenuation with depth of the downwelling irradiance, is one of the most studied parameters. It is also used to approximate **water turbidity** more globally, as it is linked to the presence of suspended sediments and other yellow substances. K_d is not strictly an inherent water parameter, as it is not invariant to the incident radiance, and thus depends on ambient illumination conditions. **Inherent optical properties** (IOP) include the **scattering and absorption coefficients**, which respectively describe how light is attenuated and scattered by water. Their sum is the **beam attenuation coefficient**, which is sometimes also studied with lidar waveforms. K_d, the scattering, the absorption, and the beam attenuation coefficient depend on the distance covered by the beam and on the wavelength of the incident radiation.

Other examples of IOPs are the more specific parameters describing the scattering events. The **volume scattering function** (VSF), describes how a volume scatters incident light across angles varying between π and 2π depending on its volume. The scattering coefficient is the integral of the VSF over a sphere. The **backscatter coefficient** describes the portion of radiation scattered in the backward hemisphere. The way light is scattered across angles - the VSF - depends strongly on the characteristics of the particles present in the water. They tend to be inhomogeneous. For example, in natural waters, they are peaked in the forward direction. VSF observations made in Petzold, 1972 show that about 75% of the scattering takes place within ten degrees of the incident direction. The **particulate backscattering co-**

efficient b_{bp} denotes the scattering due solely to particulates when removing scattering by freshwater or seawater.

One last example of IOP is the **single scattering albedo**, sometimes denoted ω_0. It corresponds to the ratio of total scattering to total absorption. It ranges from 0 to 1, 0 designating pure absorption, and 1 pure scattering. Physically, the value of ω_0 depends on the presence of organic and inorganic particulates.

Watkins et al., 2023 explain the significant potential of lidar-derived estimations of water properties. As the authors recall, K_d and b_{bp} are key descriptors of the water quality of inland water extents and are usually hard to measure. They are either obtained via expensive and time-consuming sampling campaigns or using passive imagery, which can result in errors of more than 50%. By comparing in-situ measurements of K_d and b_{bp} to estimations made with ICESat-2 data acquired the same day, Watkins et al., 2023 find high agreement, which the authors interpret as a confirmation of the potential of lidar sensors for freshwater lakes monitoring.

While other properties such as phytoplankton subsurface stratification can be approached too (Lu et al., 2021), the literature on this subject mostly features estimates of water optical parameters, namely the beam attenuation coefficient or turbidity.

Globally, 3 approaches can be distinguished:

- Using **exponential fitting on the waveform** to approach K_d through an empirical relationship
- **Inverting radiative transfer models** (RTM) based on the lidar equation and optimizing their parameters to best fit the observed waveforms
- **Using regression methods and empirical observations** of K_d to have training couples of waveforms and K_d values

Use of exponential fittings and waveform decomposition parameters Most papers dealing with the estimation of IOPs or other water parameters use **exponential fittings on bathymetric waveforms**. The **slope of the exponential is then used to estimate the turbidity** of the water column (Richter et al., 2017) through an approximation of K_d, K_{estim}, based on a relation between the waveform power p and the depth z (Lu et al., 2023):

$$K_{estim} = 0.5 \frac{d}{dz} \ln(pz)$$

This procedure may be applied to orthowaveforms to benefit from a better signal-to-noise ratio (Pan et al., 2016; Richter et al., 2021; Richter et al., 2022). For example, Richter et al., 2021 show that fitting

an exponential between the maximum and the minimum of the useful part of the waveform allows an efficient estimation of the value of K_d. Figure 2.11 illustrates this approach.

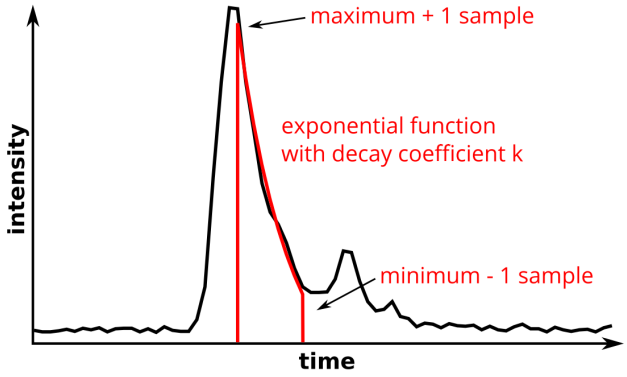


Figure 2.11: Extract of Richter et al., 2021: Schematic representation of volume backscatter extraction and exponential function approximation.

In another paper, the authors also explore the fitting of several exponential segments to also work in water with pronounced turbidity stratification (Richter et al., 2022).

Additionally to exponential functions, the parameters of fitted triangle functions or Weibull distributions are also used as proxies of suspended sediments concentration or of the diffuse attenuation coefficient (Zhao et al., 2018b; Montes-Hugo et al., 2016).

However, it should be noted that **depending on the sensor's FOV, the approximated value of K_d may not be the true diffuse attenuation coefficient**, but a parameter linked to the sensor, as all the diffused energy may not be captured by the receiver. Values estimated with narrow FOV lidars are thus closer to a K_{lidar} parameter than to the real K_d. Feygels et al., 2003 exploit this nuance to derive knowledge on water optical properties. The authors show that using sensors with three different FOVs - one narrow, one intermediate, and one large - results in three different estimations k. They suggest using the differences between these three estimations as proxies of different functions of seawater IOPs, which are established through physical simulation models.

Lu et al., 2023 use exponential fittings on ICESat-2 waveforms to map K_{lidar} around the globe during daytime and nighttime and across seasons, and observe a relation between its values and the sea ice concentrations. Their spatial rendering of K_{lidar} confirms the potential of lidar for water extents monitoring identified in Watkins et al., 2023. Figure 2.12 is an extract of Lu et al., 2023 illustrating typical results obtained with their approach.

Regression approaches **Several regression approaches are based on observations of MODIS**, NASA's Moderate-Resolution Imaging Spectroradiometer to train models able to predict water proper-

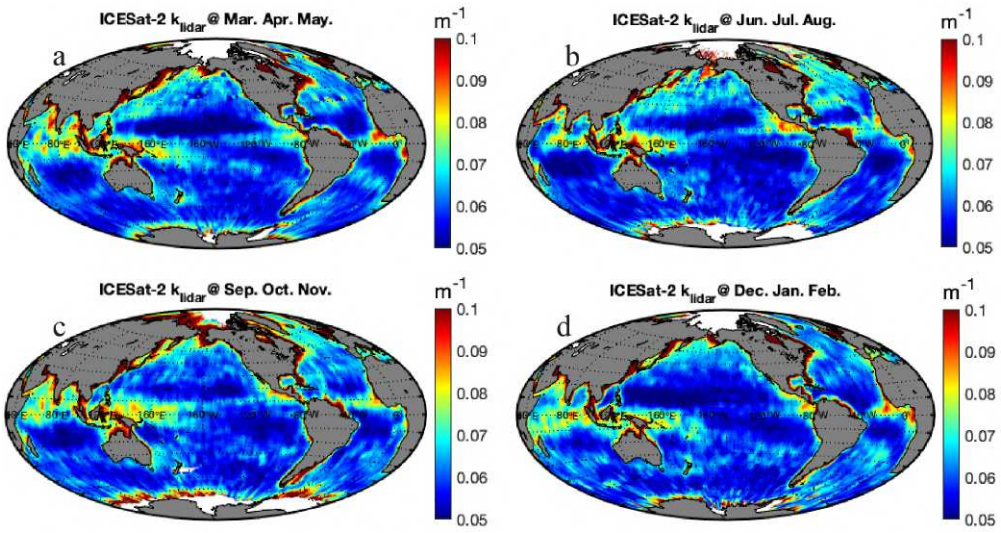


Figure 2.12: Extract of Lu et al., 2023: Seasonal distributions of ICESat-2 observed effective lidar attenuation coefficient (klidar) during nighttime (a) March-May; (b) June – August; (c) September – November; and (d) December – February. Data are seasonally averaged climatologies for the 2018 -2020 period binned to 0.5 latitude by 0.5 longitude pixels.

ties from satellite lidars.

For example, Zhang et al., 2023 build a MLP to retrieve particulate organic carbon and b_{bp} from measurements of the CALIOP polarized lidar. They use specific measurements available through this sensor - the ocean column-integrated depolarization ratio and the subsurface column-integrated perpendicular backscatter data - to predict b_{bp} and particulate organic carbon. The relation between these CALIOP features and b_{bp} and particulate organic carbon are learnt using MODIS products. The authors even highlight that their network has higher estimation performances than physical models-based estimations usually exploited to analyse CALIOP data.

Corcoran et al., 2021 develop an approach based on similar principles to estimate the diffuse attenuation coefficient K_d from ICESat-2 measurements. They also use a regression model - here, a random forest - and train it using MODIS measurements of K_{d490} transformed into K_{d532} with a physical relation.

Inversion of radiative transfer models Finally, **RTM inversion** can also be used on bathymetric waveforms, this time to retrieve water optical properties. However, caution is advised in Tuell et al., 2005, which states that a procedure of **normalization of the waveform power** should be applied to compensate for the effects of signal compression and of the frequency of the receiver. The authors thus develop a method to normalize and calibrate measurements of lidar waveforms to enable the estimation of IOPs from these signals. Kopilevich et al., 2005 then use these normalized waveforms to estimate seafloor re-

flectance and IOPs, using an analytical solution to the RTM.

2.3.5 Current state of full-waveform lidar data exploitation

The potential of full-wavefom lidar data analysis lies in two possibilities: improved backscattered pulse detection and ranging, and access to physical parameters of the intercepted surfaces. Indeed, compared to real-time pulse detection systems, a finer look at the signal can allow the retrieval of weaker echoes, sometimes surrounded by noise and difficult to separate with fixed amplitude thresholds.

In particular, full-waveform data processing can significantly improve the detected surface backscatters in submerged areas, and under vegetation, where they still are the most challenging. Additionally, the possibility of retrieving information about the geometry and the material of the objects surveyed opens opportunities for various environmental applications relying on 3D structural assessments or on the possibility of accessing covered surfaces.

Examples of applications enabled by full-waveform analysis include point clouds densification in rugged, vegetated or submerged areas, coastal habitats 3D mapping, vegetation health, species, biomass, and structure evaluation.

However, the presented approaches include rasterized products of lidar waveforms and do not always compensate for effects affecting full-waveform component shapes, making them site-limited.

Additionally, there is a lack of methods allowing to classify the intercepted surfaces directly from lidar waveforms, without intermediary steps, and without summarizing them into one class. Namely, semantic segmentation of full waveform or simultaneous target detection and classification could be further investigated, in light of recent advances in machine learning applied to 2D images.

Waveforms also have underexploited potential to provide additional environmental parameters, namely in topo-bathymetric environments, where they could be used to derive:

- Aquatic and terrestrial vegetation health, structural complexity, volumes, types, and biomass
- Water quality
- Marine, aquatic and terrestrial ecosystem functions or services
- Land-water ecosystems 3D connectivity
- Blue and green carbon stocks
- Blue and green corridors

All of these directly or indirectly influence waveform shapes and could thus potentially be estimated from full-waveform topo-bathymetric lidar surveys.

Lidar waveform processing is however still limited by two main factors: data availability, and data complexity. The first problem, **accessibility**, is due to the **rare delivery of full waveforms** by sensor constructors along 3D point clouds. This is partially due to the confidentiality of proprietary formats and associated pre-processing that play a major role in today's sensor evolutions. Another explanation is that waveforms are not commonly requested data.

This is linked to the second limitation: the **complexity of full-waveform data**. Full-waveforms are **noisy data, occupying a large volume of memory, in unusual formats**. As a result, waveforms are difficult to visualize and manipulate. Although basic processing can be achieved with limited knowledge, retrieving additional information from these signals requires an advanced understanding of how lidars work and how green and/or infrared lasers interact with their environment.

The **noise** in the data is a largely limiting factor for bathymetric waveform analysis, in which weaker seafloor returns can be surrounded by noise, and thus considered false detections, as the estimated level of noise is often used to filter the results of waveform decomposition or deconvolution (Parrish et al., 2011).

Another difficulty is the **difficulty of interpreting backscattered intensities**, as explained in Chapter 1 (Kashani et al., 2015). Although emitted pulse information can be used to calibrate or correct the waveform intensities, their information is still not trivial to interpret, as it is linked to various optical and geometrical processes. **Overall, there is a lack of accessible methods to develop and popularize the use of full waveforms to improve environmental assessments.**

The development of such methods is also hindered by the **low availability of benchmark labelled dataset**s on which to base new approaches and the complexity of producing labelled waveform data. Currently, a possibility is to use **waveform simulators** to investigate further how to work with these data. In a world where data is the key, the main obstacle remains the access to representative enough samples to provide the lidar community with powerful foundation tools to make more of their surveys.

Globally, stimulating full-waveform analysis and encouraging the development of processing methods would also allow for improvement in the methods used by lidar sensor constructors to produce and deliver the data, and thus to increase the quality of lidar-derived products for the whole community. Exploiting lidar surveys more deeply would also justify their use in environmental remote sensing despite their significant environmental cost.

2.4 Conclusion

Lidar data has the potential to provide knowledge on many environmental phenomena, in particular over areas at the interface between land and water. This potential is still largely underexploited due to the lack of methods both adapted to the complexity of lidar data processing and accessible – in terms of ease of use and of predictions interpretability – to people who could make the most of lidar data: thematic specialists, and governmental agencies.

There are two main types of approaches for data processing: classical machine learning which relies on the optimization of functions on handcrafted features of data, and deep learning, which learns the function directly on the data. Handcrafted descriptive features of 3D point clouds are numerous and efficient in describing scene geometry and spectral characteristics. On the other side, feature learning on 3D point clouds is recent and still difficult to apply easily to large datasets for thematic specialists due to its complexity and computation cost. Overall, there is a need for accessible point cloud information extraction methods, that would allow thematic specialists with no specific machine learning knowledge to exploit their lidar data in 3D and perform advanced analyses not allowed by 2D data.

On the other hand, waveform processing methods are more scarce; they are limited by noise and do not fully exploit the possibilities offered by machine learning, which could be useful considering the modelling complexity necessary to exploit them. Full-waveform lidar data exploitation is also held back by the low accessibility to data.

Globally, the current state of machine learning is highly favourable to the development of methods for topo-bathymetric lidar point clouds and waveforms processing. Numerous methods to solve complex problems now exist, for 3D data and for 1D time series. In the following chapters, we propose learning-based methods to contribute to the development of knowledge extraction from lidar surveys. In Chapter 3, we explore bi-spectral point clouds classification; in Chapter 4, we experiment with machine learning and deep learning to extract semantic information from waveforms; lastly, in Chapter 5 we exploit physical models to go further in waveform processing.

CHAPTER

3

BI-SPECTRAL 3D POINT CLOUDS CLASSIFICATION

Contents

When it comes to extracting knowledge from lidar data acquired over complex natural environments, a first possibility is to process the most common lidar product: the 3D point clouds. As explained in Chapter 2, there are currently no methods specifically designed to exploit the wealth of bi-spectral lidar point clouds, and thus few detailed classification approaches adapted to topo-bathymetric environments. In this chapter, we introduce a workflow specifically designed to exploit the three sources of information contained in dual-cloud surveys: the information contained in the first point cloud, in the second point cloud, and the knowledge represented by the sampling differences between them. The result of these developments is a CloudCompare plugin called 3DMASC, for 3D classification using Multiple Attributes, Scales and Clouds. The global concept is explained below, in figure 3.1. This work was submitted to the ISPRS Journal of Photogrammetry and Remote Sensing on February 6th, 2023. It has been through two rounds of review, the first answered on August 4th, 2023 and the second on September 25th, 2023, and is currently under review.

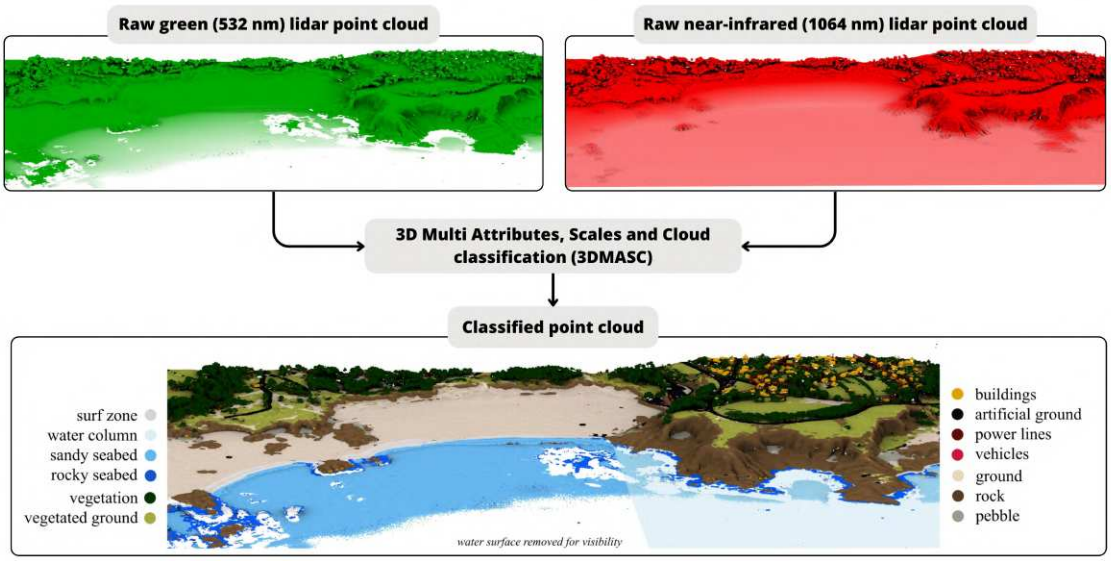


Figure 3.1: Graphical abstract of the method presented in chapter 3.

This work was developed over several years: the initial idea and the first developments were part of another research project led by Arthur Le Guennec, Dimitri Lague, Sebastien Lefèvre, and Thomas Corpetti. Part of this PhD was dedicated to deepening the initial developments by:

- investing more features - namely experimenting with previously unexplored combinations of statistical operators and lidar point cloud descriptors such as mean values of echo features -;
- designing and implementing the feature selection, classification evaluation, and model explainability procedures;
- manually labelling data and performing an extensive set of experiments necessary to better assess the performances of the plugin and its application to multiple use cases;
- contributing to the publication of the plugin, of open-source Python scripts and labelled datasets, and ensuring its diffusion and accessibility to the community through scientific communications, online documentation, and a dedicated workshop at a summer school.

A careful reader may wonder why no neural networks are involved in this chapter, despite their undeniable usefulness and growing popularity for lidar data processing. This is a deliberate decision that resulted from a thorough reflection on the current state of 3D point cloud classification in geosciences and of 3D deep neural networks dedicated to classification or semantic segmentation. We indeed realized that although significant advances had occurred in computer vision that benefitted 3D data classification, there were still no tools accessible to thematic experts relying on 3D point clouds. As a matter of consequence, geoscientists who find advantages in using 3D data still lack accessible tools to automatically process them and often end up under-exploiting their costly datasets or converting them to 2D rasters at the expense of information loss. With this work, our principal aim is to open 3D point cloud classification to users with other areas of expertise than 3D data processing or machine learning. Our belief is that at the moment, deep neural networks are not the optimal solution to fulfill this need. Indeed, efficient 3D neural networks are still recent and, although highly performing, they remain - for now - experimental and less accessible to non-specialist users than more classical machine learning algorithms. They require dedicated and well-parameterized graphics processing units (GPUs), which limits their accessibility to a wide range of environmental researchers. Their increased complexity to handle, optimize, and parameterize without extended knowledge about machine learning, scientific programming and mathematics is another factor limiting their deployments for 3D analyses by thematic users. The recent advances towards lower supervision and smaller amounts of labelled data still show lower performances than fully-supervised approaches. Finally, there are, to our knowledge, no methods capable of considering two point clouds derived from two different wavelengths. As a consequence, because of the lack of foundation models for

3D data processing in natural environments, the low availability of labelled data in such scenes, and the more complex and debated explainability of deep learning models, we prefer to turn towards machine learning on handcrafted features, as our goal is to provide a generalizable and accessible tool.

3.1 3DMASC: classification of bi-spectral point clouds

3D data are becoming increasingly popular among geoscientists, as they constitute major opportunities for enhanced observation of natural processes and more precise risk assessments. In particular, in complex, natural environments combining vegetated terrains, artificialized portions, and submerged areas, specific 3D point clouds obtained through topo-bathymetric (TB) lidar sensors are an opportunity to gather knowledge inaccessible to other surveying methods. Indeed, TB lidar sensors were introduced specifically to enable the documentation of shallow waters at high resolution (Fernandez-Diaz et al., 2014; Fernandez-Diaz et al., 2016; Mandlburger et al., 2015; McKean et al., 2009; Quadros et al., 2008; Wang et al., 2007). These sensors combine the strengths of topographic lidar sensors, equipped with small footprints near infrared (NIR) lasers and high shot densities particularly useful to image vegetated areas, that cannot penetrate water, and large footprint bathymetric lidars, able to image seafloors deeper than 30 m in clear waters (Guenther et al., 2000; Philpot, 2019), but with reduced point density and spatial resolution, and high mobilization costs. TB lidar sensors practically consist in combining both types of sensors; a NIR laser ($\lambda = 1064$ nm) and a green laser ($\lambda = 532$ nm), and their respective benefits. Associated TB lidar datasets are bi-spectral, consisting of one point cloud (PC) per wavelength, with submerged topographies as detailed as emerged parts (see Figure 3.2). Due to the different specificities of each laser, namely their footprint size, scanning angle range, and wavelength, the obtained PCs are systematically different, and provide distinct samplings of the same scene, in particular over vegetation and submerged surfaces. TB lidar sensors are useful in the study of varying subjects. Combining high-resolution data about the submerged and emerged surfaces offers new opportunities to map habitats in fluvial (Fernandez-Diaz et al., 2014; Mandlburger et al., 2015; McKean et al., 2009; Pan et al., 2015) or coastal (Chust et al., 2010; Hansen et al., 2021; Launeau et al., 2018; Parrish et al., 2016; Smeeckaert et al., 2013; Wilson et al., 2019) environments, improve high-resolution modelling of flood inundation (Lague et al., 2020; Mandlburger et al., 2015) or track sediment transport at the land-water interface. These bispectral sensors also showed useful for vegetated or urbanized terrain assessment (Dai et al., 2018; Ekhtari et al., 2018; Laslier et al., 2019; Morsy et al., 2017b; Wichmann et al., 2015). However, to fully use them and leverage the scientific potential of such extensive datasets over complex natural scenes, adapted processing methods are necessary. In particular, automatic classification of the green lidar data directly at the 3D PC level is essential.

In such areas, the exploitation of the PCs obtained with both wavelengths is beneficial, mainly (i)

because the rich spectral information provided by the combined surveys can be leveraged to distinguish vegetated, submerged, and urban objects that can be mixed in the same scenes, and (ii) because the increased geometric information provided by a simultaneous sampling with two lasers provides additional information and chances to image a greater portion of the terrain.

Several methodological challenges complexify the development of adapted classification workflows. First, the refraction of the green laser in water makes it critical to detect all green points below water during data production to subsequently perform accurate refraction correction on the position of the received echoes. This correction requires accurately knowing the spatial extent of water in the scene and the local water elevation, which the NIR channel gives when data are available in the area. While relatively straightforward in coastal environments or large lakes as water has a constant elevation, detecting water surfaces is far more challenging in fluvial environments for four reasons: (i) water elevation decreases downstream, sometimes abruptly at the vicinity of dams; (ii) rivers can have several active braids or complex hydrological connection with abandoned channels or lakes in adjacent floodplains; (iii) full mirror-like NIR reflection may occur on flat water such that the NIR PC may lack water surface echoes over large areas (Figure 3.2); (iv) vegetation frequently grows on the floodplain such that river banks and small lakes may be completely below vegetation making things even more complex as canopy interception reduces the backscattered intensity and the likelihood of having a water surface NIR echo and bottom green echo (Figure 3.2).

Second, the backscattered green laser energy generates two prominent echoes in an ideal clear water column. The first is a volume echo located just below the water surface. Its position can deviate from the actual water surface by several dozens of centimetres depending on water characteristics, leaving no other possibility than to use the corresponding NIR survey to derive the real water surface.(Guenther et al., 2000;Lague et al., 2020; Philpot, 2019). Though this volume echo is of no use, it is systematic for any shot. The second echo corresponds to the bathymetry. However, in turbid or deep waters, it sometimes has such a weak amplitude that its signal-to-noise ratio hinders its detection. For a given sensor and flight elevation, the maximum measurable water depth thus highly depends on water clarity and bottom reflectance (Guenther et al., 2000; Lague et al., 2020; Philpot, 2019). For instance, in clear coastal waters, the Teledyne Optech Titan sensor can reach depth down to 10-15 m over bright sand but can be limited to 0.5 m over dark rocks and will typically only reach depths of 1-4 m in rivers owing to the reduced water clarity (Lague et al., 2020). Thus, it is commonplace in inland water surveys that deeper parts of rivers or lakes are locally not detected due to green laser extinction.

Consequently, as for ground detection below dense vegetation, one cannot assume that simple operations such as picking the lowest green point over a specific area or extracting the last recorded echo in the green PC will systematically isolate the seabed or riverbed. Similarly, because of the green water surface

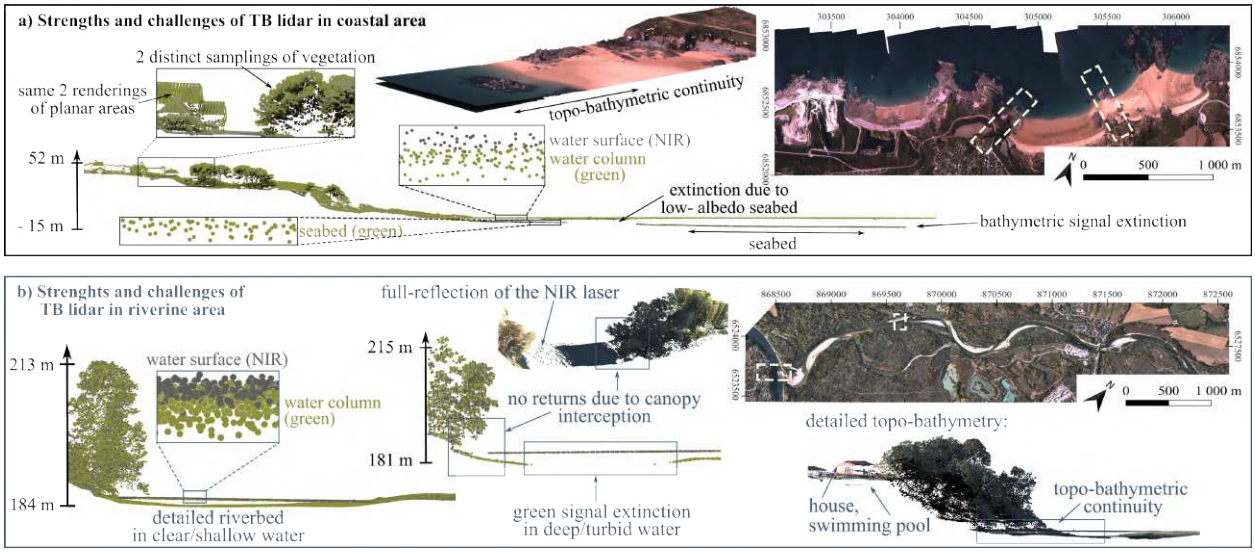


Figure 3.2: Strengths and challenges of topo-bathymetric lidar data. Examples of (a) the coastal setting of the surroundings of Fréhel (France) and (b) along the Ain River (France). Datasets are presented in the RGF93 coordinates system.

uncertainty, and the incomplete sampling of the NIR water surface, removing all green points below a given depth is impossible as a large part of the very shallow seabed will be discarded, and depth may be mis-estimated.

While the use of both NIR and green PCs appears essential to derive an accurate classification of submerged parts, the exact method is not straightforward, and there is currently no available solution to automatically separate bathymetric echoes from volume echoes over large PC datasets in complex inland water environments.

Additionally, beyond the detection and separation of bathymetric and volume echoes of the green laser, classifying the nature of the land-water continuum – seabed or riverbed covers and above-ground features – on 3D PCs is a significant challenge. Most of the existing approaches rely on 2D rasters classified with traditional algorithms like maximum likelihood, support vector machine, or decision trees (Letard et al., 2021b; Sun et al., 2017; Tulldahl et al., 2012; Wedding et al., 2008; Zavalas et al., 2014). Although these methods exploit geometrical features, they analyse averaged features due to the rasterization step, which may produce mixed pixels (Pi-Fuei Hsieh et al., 2001) and smooth out the geometry of the scene, as the spatial point pattern information is lost when the data is condensed into regularly spaced observations. Few studies provide 3D classifications of underwater environments using bathymetric lidar (Hansen et al., 2021; Letard et al., 2022b; Letard et al., 2022a). Additionally, some approaches require full-waveform data (Letard et al., 2022b; Letard et al., 2022a), which is complex to process and often unavailable or unpublished. Over land, the bi-spectral backscattered intensity of TB lidar offers new clas-

sification opportunities, as explored in urban environments by (Morsy et al., 2017b; Morsy et al., 2022; Teo et al., 2017). The two distinct samplings of each laser, provided by their different footprint sizes can also provide useful information. However, there have been, to date, no applications attempting to classify both clouds of a topobathymetric survey directly, and fusing them into a single cloud to apply workflows existing for forested or urban environments is impossible. Indeed, mixing the data obtained with the two different sensors would be incorrect due to their different precisions, densities, spatial resolutions, and optical characteristics. Spectral information would become unusable, and, due to the reasons explained above, the water surface sampling obtained would be unexploitable. Additionally, the capacity of one or the other sensor to image specific parts of natural scenes is information in itself, that would be lost if both clouds were to be fused for processing, but that can be exploited by directly operating on the differences between the point clouds. A processing method adapted to multiple clouds and applicable to configurations in which vegetated, urban, and submerged settings are combined is thus expected.

This work presents an original framework called 3DMASC for 3D point classification with Multiple Attributes, Multiple Scales, and Multiple Clouds, and its application to coastal and fluvial TB airborne lidar datasets. 3DMASC operates directly at the 3D PC level to produce outputs in 3D, thus preserving the rich information of spatial point patterns. This classification process relies on multiple 3D features that make it generalisable to a variety of 3D data types and to different point classes. By assessing point cloud characteristics at different scales simultaneously, it can distinguish classes characterized by different sizes while balancing salt and pepper-like noise, or errors at the borders between classes, which often come with small and large scales, respectively. Finally, our workflow operates directly on the differences between distinct samplings offered by multiple point clouds, thus leveraging the underexploited knowledge of multi-cloud surveys. 3DMASC combines proven classical elements of single PC semantic classification, such as geometric feature extraction from multi-scale spherical neighbourhoods (Brodu et al., 2012) or k-nearest neighbours (Thomas et al., 2018) and a random forest model (Breiman, 2001). In addition, it adds new features specifically engineered to leverage the NIR and green PCs. Our contributions consist in:

- Designing new joint-cloud features calculated on two PCs using their local geometry and backscattered intensity. 3DMASC uses a flexible method to compute features from two PCs, resulting in more than 80 different features;
- Screening over 80 features, both classical and new, to select the essential features and scales contributing to 3D point classification to optimize classifiers in terms of computational efficiency, generalization ability, and interpretability;
- Demonstrating that with limited training data (< 2000 points per class) and less than ten features

and five scales, the classification accuracy of TB lidar datasets can be excellent (>0.95);

- Providing a plugin in the open source software Cloudcompare (CC) that can be used easily by non-specialists to classify any 3D PC and by experts for fast 3D feature computation and visualization;
- Sharing two manually labelled state-of-the-art lidar datasets with two different levels of detail (up to 13 classes).

Sections 3.2 and 3.3 introduce our methodology. Experimental results are then shown in section 3.4, associated with a discussion in section 3.5, and a conclusion in section 3.6.

3.2 Description of the 3DMASC workflow

In this section, we describe the 3DMASC method, included in a CloudCompare plugin (Girardeau-Montaut, 2022). The method consists first of computing descriptive features at multiple neighbourhoods of different sizes, involving one or two different PCs covering the area considered for classification. An optional features and scales selection method is then applied to reduce feature redundancy and ensure the relevance of the kept attributes. Finally, a random forest model is trained and used for classification, and its results are analyzed using prediction probability, Shapley explanations and feature importance values. Section 5.5 provides details of the implementation and operation.

3.2.1 3D features extraction

3DMASC operates directly on unordered sets of points, producing a 3D classification without requiring an intermediate rasterization step. A PC is a set of n 3D points Pk | k=1,…,n in which each element Pk is a vector of coordinates (x,y,z) with associated point-based features: intensity, multi-echo characteristics, and potentially red – green – blue (RGB) colour (see Section 5.5 for a detailed list of features). On top of point-based features, 3DMASC uses neighbourhood-based features defined using a spherical neighbourhood or a k-nearest neighbour search (kNN). A maximum of four different 3D entities are involved in the process of neighbourhood feature extraction:

- 1-2) two point clouds. The originality of 3DMASC lies in using up to two PCs to characterize the scene of interest. For topo-bathymetric applications, they originate from different wavelengths, typically 532 nm and 1064 nm. We refer to them as PC1 and PC2, respectively.
- 3) A set of core points (Brodu et al., 2012), denoted PCX, that 3DMASC classifies at the end of the process. They may be a subset of points from PC1 or PC2 with a regular subsampling or other positions spread within the extent of PC1 and PC2.

- 4) An optional context PC, denoted CTX, containing any relevant context information in its Classification attribute at a potentially much lower resolution than PC1 or PC2. A typical CTX would be previously classified ground points at 2 m spatial resolution.

Neighbourhood selection and scales

3DMASC mainly uses a spherical neighbourhood search in the relevant PC – PC1 or PC2, depending on the feature to compute – to capture the surroundings of each core point (Figure 3.3a). The neighbourhood scale is defined as the sphere diameter. 3DMASC uses a multi-scale classifier computing multiple neighbourhoods for each core point (Figure 3.3c). The user typically provides minimum and maximum scales and a step (e.g., 1 m) between successive scales. The minimum scale must be consistent with the PC's density to compute features for most core points. The largest scale is typically set by the size of the objects of interest. Defining the optimal set of scales for various types of TB airborne lidar (e.g., coastal, fluvial…) is a challenge not yet resolved that we address in this work. Beyond ensuring classification success, it is also crucial for operational efficiency, as the feature computation time increases strongly with the scale and number of different scales.

3DMASC also supports kNN to measure the vertical or horizontal distance between PC1 and PC2, or CTX (Figure 3.3b). This supplements relative position measurements between PCs where diameter-based features are impossible to compute due to a lack of neighbours.

Single cloud neighbourhood-based features

Single cloud features describe PC1 or PC2 once at a time. Since many criteria characterize a 3D object and can help identify its nature, the plugin natively encompasses 15 different features (see Section 5.5 for the complete list of features). The broad set of features available is presented in the following paragraphs.

Six **dimensionality-based features** aim to describe the local PC's general aspect and identify if the object has a linear, planar or spherical outlook (e.g., Brodu et al., 2012; Gross et al., 2006; Vandapel et al., 2004). They rely on the eigenvalues of the covariance matrix of the points. 3DMASC can directly use the 3 normalized eigenvalues or classical combinations resulting in sphericity, linearity and planarity metrics.

Six **geometry-based features** inform on the shape of the PC. 3DMASC computes and uses the slope angle, the detrended roughness, the curvature, the anisotropy, the number of points at a given scale and the first-order moment, introduced for contour detection in Hackel et al., 2017.

Three **height-based features** characterize the vertical structure of the local neighbourhood with respect to the minimum elevation Z_{min} and maximum elevation Z_{max}. For a core point with elevation Z, 3DMASC computes $Z_{max} - Z$, $Z - Z_{min}$ and the local thickness of the point cloud $Z_{max} - Z_{min}$ as explained in Chehata et al., 2009.

Optional **contextual features** are used to place each core point in its spatial context and get its position relative to the ground, the water surface, or any specific pre-existing class, labelled in the CTX point cloud. They are computed with a kNN neighbourhood. They generalize the distance to ground feature used in Chehata et al., 2009 and Niemeyer et al., 2012.On top of these classical features, we propose novel 3D descriptors based on the application of **statistical operators on point-based features within spherical neighbourhoods**. Six statistical descriptors can be used: mean, mode, median, standard deviation, range and skewness. They are designed to inform on the variations of backscattered **intensity** and **multi-echo LiDAR features**: return number, number of returns and their ratio called echo ratio. These 4 point-based features combined with the 6 statistical descriptors result in 24 neighbourhood-based features at a given scale. To our knowledge, these types of rich multi-scale statistics were never used before for raw 3D PC classification.

Similar features can be built from the 3 components of the **RGB** colour information, and we evaluate at a later stage the benefits of this information for classification.

Dual cloud features

Dual cloud features describe the geometrical, spectral, height statistics or multi-echo characteristics differences between the neighbourhood of the core point in PC1 and PC2. Spectral ratios have been introduced in the context of multi-spectral lidar classification (Chen et al., 2017; Morsy et al., 2017b; Wichmann et al., 2015), but geometrical, height statistics and multi-echo characteristics are new contributions. We designed them to leverage the bi-spectral information and improve the descriptions of scenes characterized by a different 3D aspect in PC1 and PC2. In TB lidar datasets, the NIR and green PCs are most significantly distinct above water and vegetation (Figure 3.2), but they can also be slightly different over other surfaces. This is due to the different surface optical characteristics and the NIR and green laser emitters that can have different angles of incidence or aperture. These may cause differences in the returned signal intensity, and the 3D position of the points. The definition of these features assumes that both PCs are correctly registered and that the alignment error is as low as possible for geometric differences to be related to objects' characteristics and not registration errors.

Dual-cloud features result from **mathematical operations** between single cloud features of the same core point's neighbourhood in PC1 and PC2. They can be feature *differences*, *additions*, *multiplications*, or *divisions*. Here we have used *differences* to measure dissimilarity in particular for elevation, geometry and multi-echo features, and *divisions* to normalize one feature by another, typically for intensity. Figure 3.3a illustrates two examples of dual cloud features: the *mode of the difference of elevation* that is expected to be close to zero on the ground, but different over water; and the *median intensity ratio* between the green and NIR channel that is expected to be distinct over different grounds. A selection of dominant

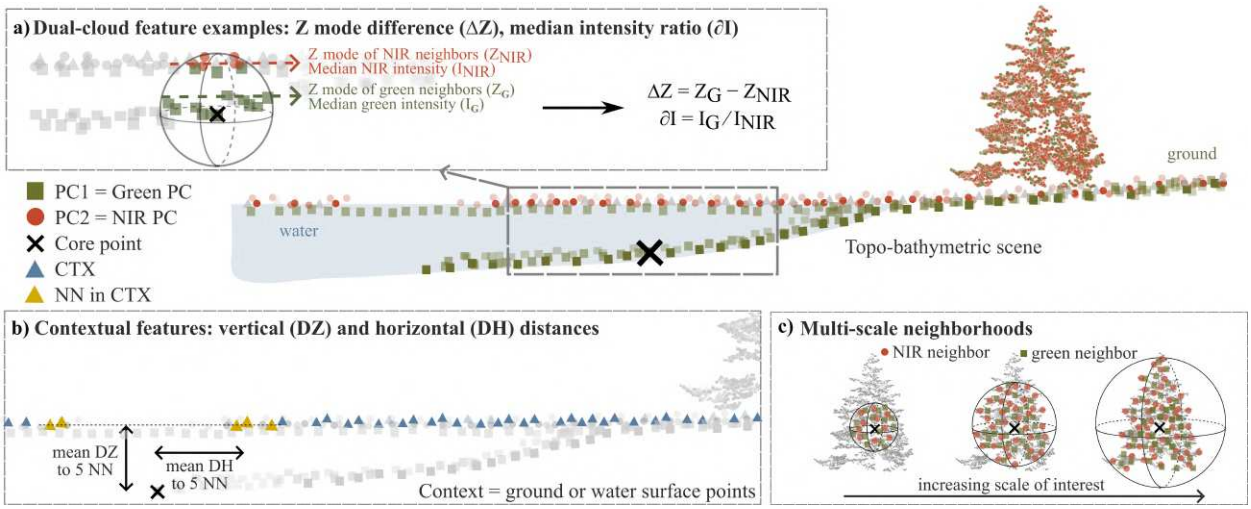


Figure 3.3: Illustration of the main characteristics of 3DMASC: a) examples of new dual-cloud features providing a better description of the differences between clouds, b) the generalized contextual attributes placing each point in its spatial setting and c) the multi-scale spherical neighbourhoods used to describe the many aspects of 3D objects.

features is presented, illustrated and explained in Section 3.4.

Dual cloud features also encompass a distance computation (vertical or horizontal) between the core points PC and another PC (PC1, PC2, or CTX), using kNN.

3.2.2 Random forest classification

3DMASC uses a Random Forest (RF) algorithm (Breiman, 2001) to perform PC classification, i.e., predict a label $y \in \{1, 2, \ldots, c\}$ for each point P_k of the input PC P, using the predictor vector F $\{F_{kij} | k = 1, \ldots, n; i = 1, \ldots, f; j = 1, \ldots, s\}$, where f = number of features, s = number of scales and n = number of core points and each $F_{kij} \in \Re$. For instance, the label can represent the type of object sampled by P.

Here, the feature importance is the product between the probability of reaching a node (i.e., the proportion of samples that get that node) and the Gini impurity decrease of that node. Feature importance is normalized to sum up to 1. A higher value symbolizes a more significant influence of the feature on the prediction.

RF does not handle Not-a-Number (NaN) values which may be present with our features (depending on the scale, NaN values can occur if the spherical neighborhoods does not contain the minimal number of points required to compute the covariance matrix for example). This requires specific pre-processing of the predictor vector. Indeed, replacing NaNs with a fixed value may imply irrelevant representations of the local sub-clouds and thus incorporate bias in the classifier training. To tackle this issue, 3DMASC

relies on the RF implementation of the cross-plateform library OpenCV (Bradski, 2000), which incorporates surrogate splits to handle missing measurements. We use base settings and forests populated with 150 decision trees, having a maximum node depth of 25, as advised in (Oshiro et al., 2012). We also compared it with the RF implementation in the python library Scikit-learn (Pedregosa et al., 2011), and found similar results.

To further improve the robustness of the classifiers, we also exploit the prediction probability output by RF and use it as a classification confidence indicator, as seen in Brodu et al., 2012 and Letard et al., 2022b. The prediction probability corresponds to the proportion of forest trees that voted for the class assigned to the point. It ranges between 0 and 1.

3.2.3 Features and scale pre-selection to control the size of the predictor vector

We propose a feature selection routine (Dash et al., 1997) to improve the explainability and efficiency – through the number of predictors – of the trained algorithm, as there can be almost 90 features per scale in TB environments.

Although information redundancy supposedly does not impact RF performances, it disrupts the explainability of the feature importance values, since if two features bring similar information, their relative importance will be underrepresented. Thus, we keep only a set of uncorrelated features, by using a bivariate feature selection (Dash et al., 1997; Guyon et al., 2003), incorporating an assessment of the features' Information Gain (IG) (Dash et al., 1997) and the Pearson linear correlation coefficient of attribute pairs. The correlation threshold and the scale at which each feature is evaluated are user-defined, and determined after an empirical investigation.

The same bivariate procedure allows the selection of scales. However, we also decided to promote small scales to limit the computation cost of the classifier. The selection process relies on a majority voting procedure. Since it is impossible to consider a scale independently from its application to a feature, we retain the scales that are the most often selected when they are evaluated for each feature independently.

Considering the variety of features included in 3DMASC, removing correlated features and scales does not provide a significantly smaller set of features. Typically, in our experiments on two datasets developed in the next sections, around 40 features per scale of interest remain after correlation-based pruning. The classifier obtained may thus not be easier to explain, and the application steps may be unnecessarily computationally heavy.

To further reduce the dimension of the predictor vector, we considered a feature ranking depending on the IG. However, defining a fixed number of features and scales is highly task- and site-dependent, and filter-based selection would not consider internal synergies between features. Consequently, we use an embedded backward feature selection, relying on the RF feature importance, as detailed in (Dash et al.,

1997; Aggarwal, 2014). This selection is performed on the uncorrelated set previously obtained. The optimized predictor vector is then identified through automatic out-of-bag score (OOB) monitoring. OOB corresponds to the performance score obtained by the RF algorithm on a sample that is not part of the subsample of data used to build the trees (Breiman, 2001). It is used to estimate the generalization ability of the model, i.e. the performance of the classifier on new, unseen data. In practice, we tested the performance of each model build during optimization on our test dataset and indeed observed that prediction accuracies on the test data varied similarly to OOB (see Section 5.5, Figure 5.20). We thus use a sliding window to monitor the variations of OOB over a given number of iterations (in our experiments, ten) and keep the last best iteration before OOB starts to drop (i.e. varies more than a user-defined threshold).In the rest of the paper, we will refer to this step as classifier optimization, which, since the OOB score is obtained using the training data, is performed completely independently from the test data.

3.2.4 Framework implementation

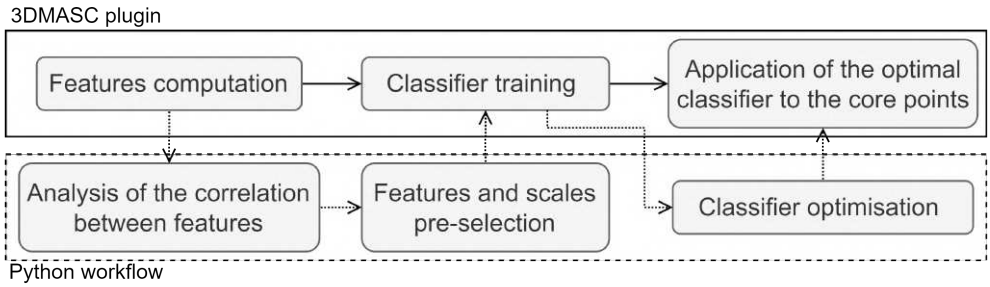


Figure 3.4: The 3DMASC workflow.

Figure 3.4 sums up the global framework introduced in this work and illustrates how the different steps explained follow each other when processing a PC. As detailed in Appendix A, the Cloudcompare 3DMASC plugin can be used at two levels of complexity: for beginners, a complete graphical user interface (GUI) exists from feature computation to classifier training and class inference; for expert users, 3DMASC can be called through command line solely for fast feature computation with its parallelized C++ implementation, and the results subsequently used in any other environment such as python. Feature and scales preselection, as well as classifier optimization described in section 3.2.3 follows this latter approach and operates through a complementary Python script. To avoid feature preselection and classifier optimization for non-specialist users, a key objective of this work is to identify a minimal set of features and scales that can systematically be used for TB lidar classification.

3.3 Published datasets and experiment protocol

3.3.1 Experimental datasets and classes

To illustrate the use of 3DMASC for bispectral lidar data classification, we selected two topo-bathymetric lidar datasets, representing one coastal and one fluvial environment, respectively (Figure 3.2). These two datasets only differ in the type of environment they model. They were both surveyed with a Teledyne-Optech Titan airborne lidar with two wavelengths, 532 nm and 1064 nm (Lague et al., 2020). The green laser points with a forward pitch of 7°, which is necessary to avoid strong surface reflection on water and has a beam divergence of 0.7 mrad. The NIR laser has no forward pitch and a beam divergence of 0.3 mrad. Consequently, the incidence angle, surface sampling and laser spot size are never the same at a given location of the scene for the two lasers. The sensor produces high-density PCs, typically 36 pts/m^2 on land – when combining both wavelengths – and 18 pts/m^2 under water in a single pass (Lague et al., 2020). More details about the sensor and the acquisition conditions – typical aircraft altitude, speed, overlap between flight lines and preprocessing – are available in (Lague et al., 2020). The mean vertical offset between the two channels measured on flat horizontal surfaces is typically less than 1 cm. The precision evaluated as the standard deviation of point elevation measured on flat horizontal surfaces is around 5 cm on topography and 10 cm on submerged surfaces.

The first site lies on the French coast of the Channel, in Britanny, near the town of Fréhel; the second is a portion of the Ain River in South-Eastern France near its confluence with the Rhône River. The surveys were conducted in May 2021 and September 2016, respectively. Figure 3.2 features the two scenes. They both contain natural and anthropic land covers and include a part of the bathymetric environment: in the first case, shallow sea water with green laser extinction at 10.5 m; in the second case, a river with green laser extinction at 3.5 m. The flights combined lidar surveys and simultaneous RGB imagery acquisitions with the control camera, which produced orthoimages with ground sampling distances of 25 cm and a registration error of about 20 cm. As RGB imagery acquisition was not the main objective of the surveys, pronounced shadows exist in particular on the Ain survey as it happened late in the afternoon.

3.3.2 Classes definition and 3D annotation

We evaluate the performances of 3DMASC on two levels of detail: a primary classification of 5 land covers – strictly identical for both areas – and advanced labelling of 11 and 13 types of objects on the Ain and Fréhel datasets, respectively. We chose the classes depending on the diversity of land and sea covers we could observe in each area. Table 3.1 contains all the categories that we use for the primary and advanced classifications.

Basic classes (primary classification)	Detailed classes (advanced classification)	
Both	Ain (river)	Fréhel (coast)
Ground	Bare ground	Sand
		Pebble/Cobble
		Rock
	Artificial ground	Artificial ground
	Vegetated ground	Vegetated ground
Vegetation	Intermediate	Vegetation
	High	
Artificial elements	Buildings	Buildings
	Power lines	Power lines
	Vehicles	Vehicles
Seabed/Riverbed	Riverbed	Sandy seabed
		Rocky seabed
	Swimming pools	/
Water	Water column	Water column
		Surf zone

Table 3.1: List of classes defined for the experiments on two levels of detail.

Artificial ground includes roads and surfaces covered with concrete or tar (parking lots, dykes). Vegetated ground is grass or other low vegetation, such as low-growing heather in moors. In the Ain survey, intermediate vegetation is defined as bushes or shrubs with a different aspect than high trees and a smaller growing height. We did not use a classical classification based on a strict height threshold, as usually made in vegetation mapping applications (Letard et al., 2022a). Our objective was to avoid the traditional misclassification of low branches attached to high trees as shrubs while they are points belonging to high vegetation. The definition of intermediate vegetation and high vegetation therefore balances 3D aspect and height above ground. Compared to other classes that can be objectively defined, our separation between intermediate and high vegetation is rather subjective. The lack of various types of vegetation in the Frehel datasets prevented us from refining the vegetation class.

We annotated portions of data manually using visual interpretation of the PCs and the RGB imagery acquired simultaneously using Cloudcompare (Girardeau-Montaut, 2022) including new specific developments for quick labelling of 3D point clouds. Four training and test datasets – one for each level of detail of each scene – were created, all labelled and balanced, for the classification experiments. They all contain 2000 points of each label. To eliminate potential spatial bias due to the use of multi-scale spherical neighbourhoods, we forced each training and test point of the same label to be at least 20 m away, considering we used spheres with diameters up to 15 m. Figure 3.5 illustrates the resulting sets of points labelled for training and testing. The annotated datasets are available along with the source codes of the plugin and of the scripts used to perform further analysis at the following link `https://github.com/p-leroy/lidar_platform`. These datasets contain the NIR and green high-density surveys of both areas, a context PC representing a raw ground detection of each site at 2 m spacing, and the annotated training and test points, with two levels of detail, one with 5 broad classes, another with 11 or 13 detailed classes depending on the area.

3.3.3 Evaluation metrics

We use the **Overall Accuracy** (OA) to quantify the correct proportion of global predictions. The **precision** estimates, for each label, the actual correct proportion of positive predictions. The **recall** value evaluates the part of true positives identified correctly (Sokolova et al., 2006). Precision thus tends to outline over-estimation of some classes, while recall highlights under-estimation. Often, the goal is to balance precision and recall so there is no clear tendency of over- or under-estimation of the classes considered. In a context in which under- or over-estimation of a given class is not targeted, the smaller the difference between both metrics, the better the result. The **F-score** combines the information provided by both precision and recall through their harmonic mean (Sokolova et al., 2006).

In addition to performance quantification, we wish to address the explainability of the method. Ac-

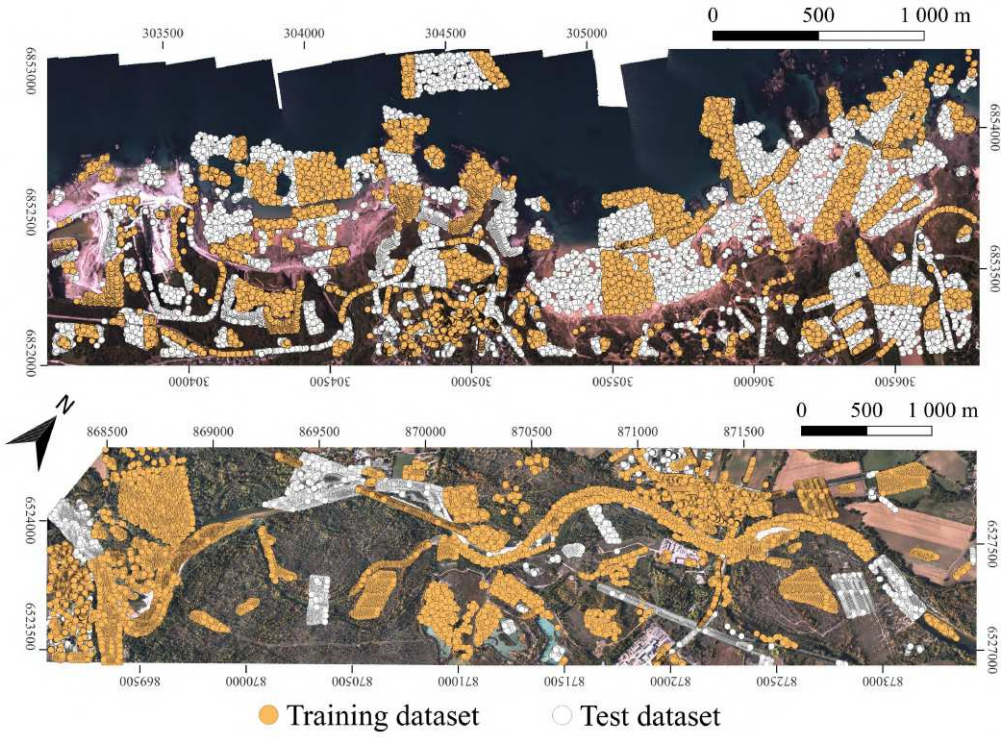


Figure 3.5: Location of the training and test datasets in both experimental areas; the coast around Fréhel (up) and the surroundings of the Ain river (down) (RGF93).

cording to Roscher et al., 2020, explainable machine learning in the natural sciences should incorporate three elements: transparency, interpretability, and explainability. We address each of these elements by, respectively, detailing the method, exploring feature and scale importance and selection mechanisms, analysing explicative elements of the decision rules behind the results obtained and their reproducibility among different environments. The class-wise performances are explainable with the approach of Lundberg et al., 2017 by computing the **Shapley values** (Shapley, 1953). These range between 0 and 1 and quantify, for each point, the influence of each feature on the label prediction, based on game theory concepts. We performed this analysis using the SHAP Python library (Lundberg et al., 2017). Using these values as a complement to the variable importance measurement and a low number of predictors in the optimized models allows us to have a more robust explanation, less dependent on the randomness of descriptors and sample selection at each node of the decision trees.

3.4 Results obtained with 3DMASC

This section first presents the overall classification results obtained in the fluvial and coastal environments and the impact of feature preselection and optimization. We then present the class-wise results and the dominant scales and features that emerge from the experiment. Finally, we explore the benefits of using RGB information, contextual data and classification capabilities when using only green lidar data. All results presented are obtained on a test dataset strictly different from the training dataset.

3.4.1 Overall classification results depending on the number of predictors

We use three different terminologies: Full predictor Set (FS) classification implies the computation of all features at all selected scales; Optimized Classification (OC) does not consider all features at all scales; Single Scale Classification uses an identical single scale for all feature calculations. The starting set of features contains 88 features, which include all possible features of 3DMASC computed on PC1, PC2 and their difference or ratio between both PCs (see Appendix B). They are calculated at 29 different scales from 1 m to 15 m with a 0.5 m increment and for kNNs with k in $\{1; 2; 3; 4; 5; 10\}$. The complete predictor vector has 2011 columns (4 point-based features, 23 features computed at 29 scales for 3 spherical neighbourhoods – green, NIR, both – and 6 kNN-based features).

To determine the scale to use for feature evaluation – i.e. IG assessment – we analyzed the OA obtained when selecting features based on their IG at scales varying from 1 m to 12 m. This first analysis shows that features computed at 2 m allow the best selection for OA (see Supplementary Material). Similarly, testing for the optimal correlation threshold results in a value of 0.85 to obtain the highest OA (see Supplementary Material). A scale of 4 m would have been valid too for the Ain dataset, but since there

was no difference between using 2 m and 4 m for this area, we picked 2 m in order to ensure comparability of the results between the two zones.

Impact of correlation and feature pre-selection

It is expected that the same feature computed at two scales separated with a small gap will produce redundant information. Here we explore if all types of features exhibit similar levels of correlation with scale. Figure 3.6 presents the mean absolute Pearson coefficient between features computed at scales separated by 1 m or 3 m for different types of single-cloud or dual-cloud attributes: geometrical, echo-based, and intensity-based.

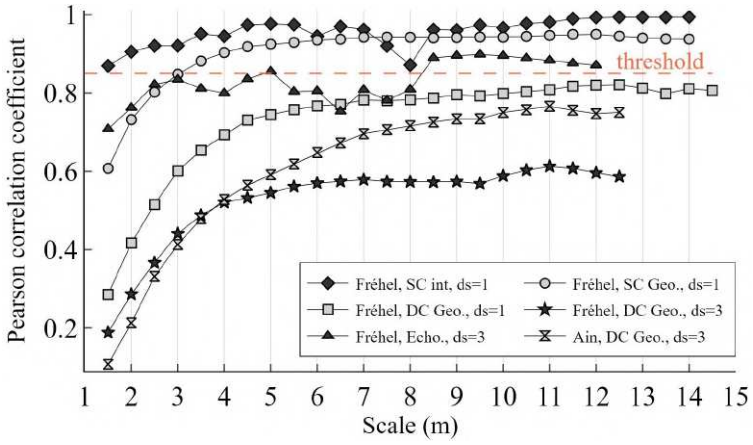


Figure 3.6: Linear correlation between features computed at scales separated by ds=1 m or ds=3 m for different examples of features. SC = Single Cloud; DC = Dual Cloud. The threshold at 0.85 emerged from an empirical analysis.

The general tendency is for absolute linear correlation to increase with scale and to saturate or increase only slightly above a threshold scale of about 4 to 6 m. The maximum linear correlation level depends on the type of feature and environment. Dual-cloud geometric features are less correlated than single-cloud features. Intensity-based features exhibit high linear correlation levels suggesting a potentially strong redundancy across scales. The comparison between the dual cloud geometric features at Fréhel for steps of 1 m and 3 m shows that the larger is the step between scales, the lower the linear correlation. As expected intuitively, the step between scales should thus tend to increase with scale, in particular above 6 m, to limit information redundancy.

These results indicate that given the high linear correlation of certain features, especially above 4-6 m, there is hardly a need for a large number of individual scales above this scale, in particular for single cloud intensity and geometric features. Consequently, we enforce the maximum number of scales kept in the preselection phase to be 10, compared to the initial 29. Finally, Figure 3.6 demonstrates that intra-

feature scale correlation is site-dependent and that there are no clear principles ruling correlation dynamics. Consequently, selecting scales for features based on their linear correlation is impossible without first computing them.

After feature preselection accounting for absolute linear correlation, different features are eliminated depending on the site. Overall, there were fewer correlated features on the Ain site and more correlation when using a higher number of classes (and therefore feature samples). The number of features passing the selection step ranges between 36 (Fréhel, primary classification) and 44 (Ain, primary classifier). Height-derived and dimensionality-based attributes were the most pruned types of features during correlation filtering. NIR and green roughness, and return numbers are strongly correlated in both areas. Measures of echo ratio were too much correlated in the Ain, whereas they were not in Fréhel, which reflects the differences between riverine and coastal TB surveys.

Impact of predictors number and optimization: from full predictor set to optimized classifiers

We explore the influence of the number of features and the number of scales used on the OA and present the results in Figure 3.7.

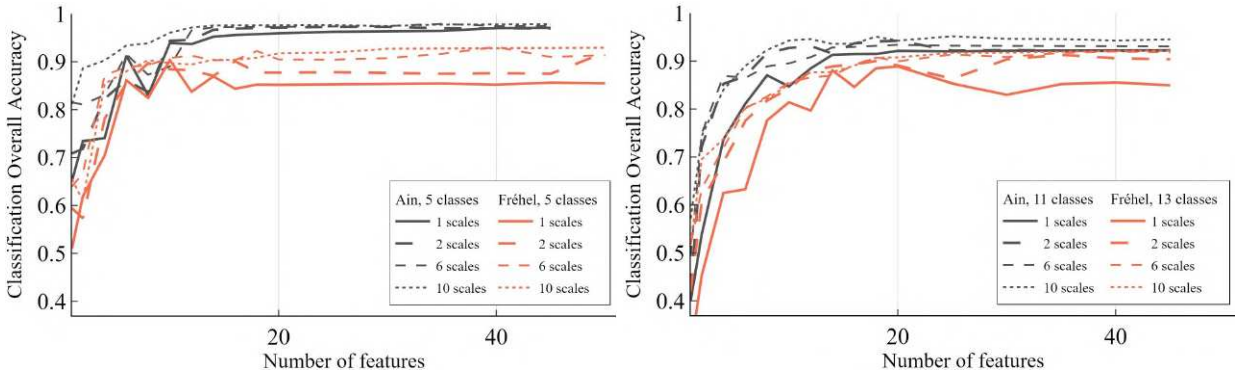


Figure 3.7: Classification accuracy depending on the number of features computed at different numbers of scales.

The results confirm the conclusions of (Brodu et al., 2012; Thomas et al., 2018) on the superiority of multi-scale algorithms compared to single-scale classifiers. This analysis also illustrates the decreasing benefit of increasing the number of features and scales past 20 features and 6 scales, even using uncorrelated entities only. Figure 3.7 highlights that adding features increases OA more than adding scales. For instance, adding a second scale to a single feature classifier systematically results in an OA surge, while harvesting 10 features at the same two scales produces more accurate results than relying on two features computed at the same 10 scales. Due to the majority voting used for scale selection, the scale used for

single-scale classification varies, explaining the variations of accuracy (see Figure 3.7, single-scale curve), and showing the dependence between the features' relevance and their computation scales.

Since the accuracies presented are the results of applying the trained algorithms to data unseen during training, these results also showcase the stability of RF relative to overfitting and generalization. Even when training the model with hundreds of predictors, OAs remain stable (between 92% and 98% depending on the use case) when classifying the distinct test points (see Figure 3.7). Furthermore, 3DMASC's features succeed at characterizing the nature of the objects lying behind the points, as accuracies converge towards values ranging between 92% (Fréhel, advanced) and 98% (Ain, primary). It is, however, delicate to determine the ideal number of features and scales to retain. The optimization procedure provides more information on the required number of predictors to achieve high-accuracy identification of the different classes. Figure 3.8 presents the OOB evolution when reducing the predictor set iteratively. The small-range oscillations observable in the OOB score can be explained by the random drawing used to get the out-of-bag samples on which to compute the score, which results in the non-respect of the samples' spatial independence, as the 20 m distance criterion only applies between samples of the training and test sets and not within each of them.

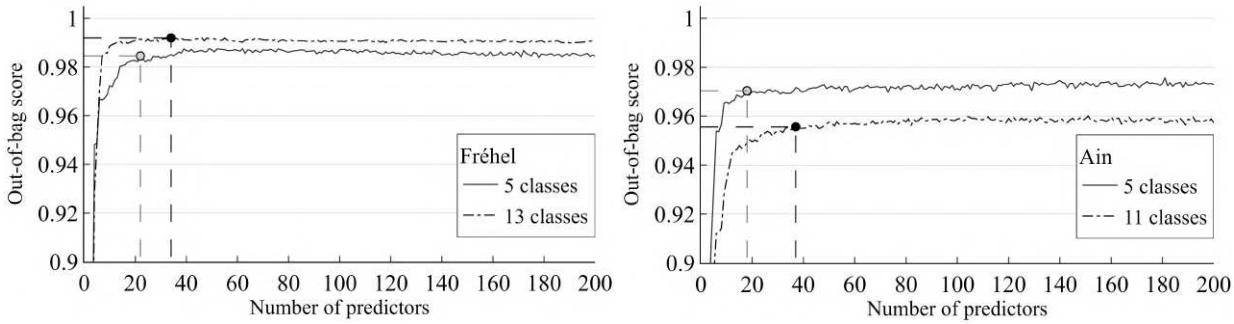


Figure 3.8: Out-of-bag score depending on the number of predictors used during classifier optimization.

With the automatic monitoring of the OOB's significant variations, a set of parameters is chosen, providing optimized classifiers for the four experiments and corresponding training datasets. Table 3.2 gathers the main characteristics of the optimized classifiers.

The results in Table 3.2 confirm what we observed in Figure 3.7: a small number of features and scales produces highly accurate classifications. The most complex classifier obtained incorporates 37 predictors, including 16 features and nine different scales. Table 3.2 also outlines that more predictors are needed to correctly identify a larger number of labels: advanced classifications require 19 and 12 more predictors than primary for the Ain and Fréhel areas, respectively. The optimized models obtain accuracies ranging between 90.7% and 97.6% and harvest more features than scales, confirming the superior efficiency of

Classifier	Ain (5cl)	Ain (11 cl)	Fréhel (5 cl)	Fréhel (13 cl)
FS OA	97.9%	94.6%	92.8%	91.9%
FS nb of pred.	371	352	315	330
OC OA	97.6%	94%	91.6%	90.7%
OC nb of pred.	18	37	22	34
Nb of features	12	16	12	19
Nb of scales	5	9	6	6
Mean confidence	0.93	0.9	0.89	0.83

Table 3.2: Characteristics of the four models. FS = full predictor set, OC = optimized classification. Nb of pred. refers to the number of predictors used.

feature diversity over scale abundance. Overall, the maximal difference in OA between full set and optimized classifiers is 1.2%. Models are highly simplified: on average, the optimization reduces the predictor vector's dimension by 93%. However, the fully iterative procedure is necessary to determine the number of predictors to use and limit the loss of OA. For example, when using the 18 highest-ranked predictors at the first RF classification of the Ain, the OA is 94%, which is almost a 4% difference.

3.4.2 Class-wise results with optimized classifier

Class-wise metrics

Figure 3.9 illustrates the application of the optimized classifiers for advanced classification.

The land-water transition is well identified and the main elements such as ground and above-ground features are separated. Figure 3.10 sums up the class-wise results obtained for each experiment. The main classes of the Ain site obtain F-scores higher than 85%. In the coastal area, they are distinguished with F-scores over 88%.

The difficulty imposed by the distinction of objects with similar geometries does not impact the performances severely. All F1-scores are higher than 83%, and average confidences are over 70% and 80% in primary and advanced cases, respectively, except for *vehicles* in the advanced Fréhel experiment. These observations suggest efficient construction of the classifiers, as correct predictions obtain the vote of most of the decision trees.

The identification of *water column* echoes is highly accurate (99%) in both advanced classifications, but there is more confusion in the primary experiments, where the broader classes may be harder to define.

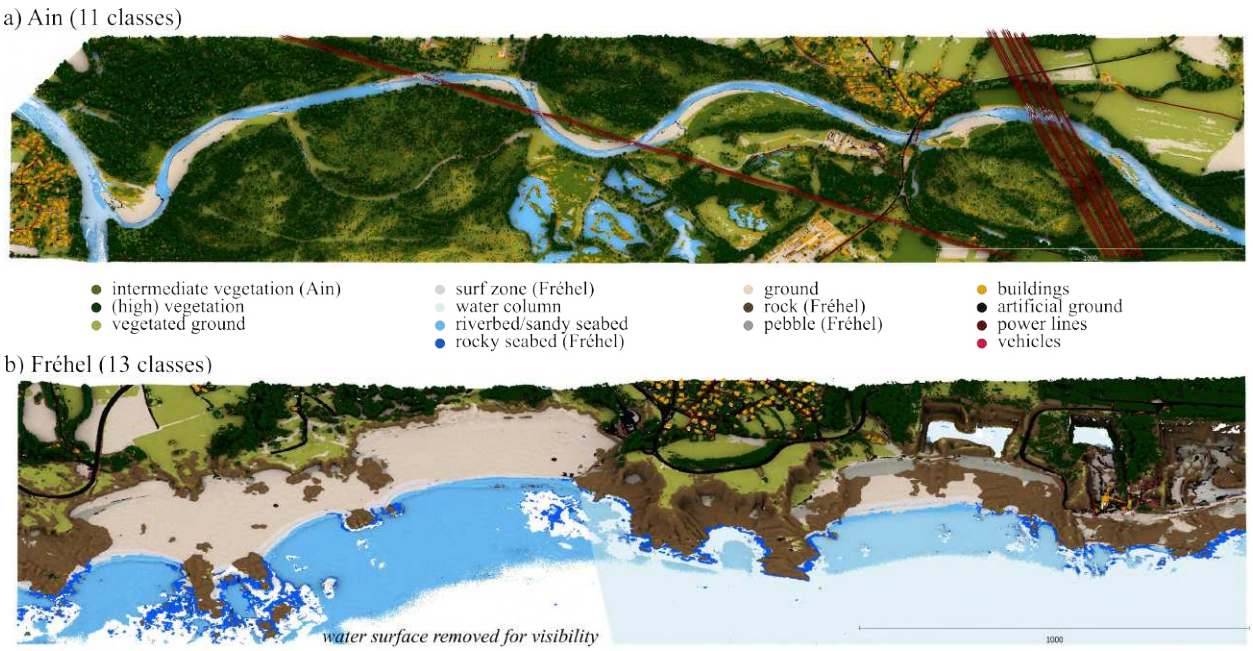


Figure 3.9: Classified point clouds of both areas using the optimized classifiers, with Fréhel on top and the Ain under.

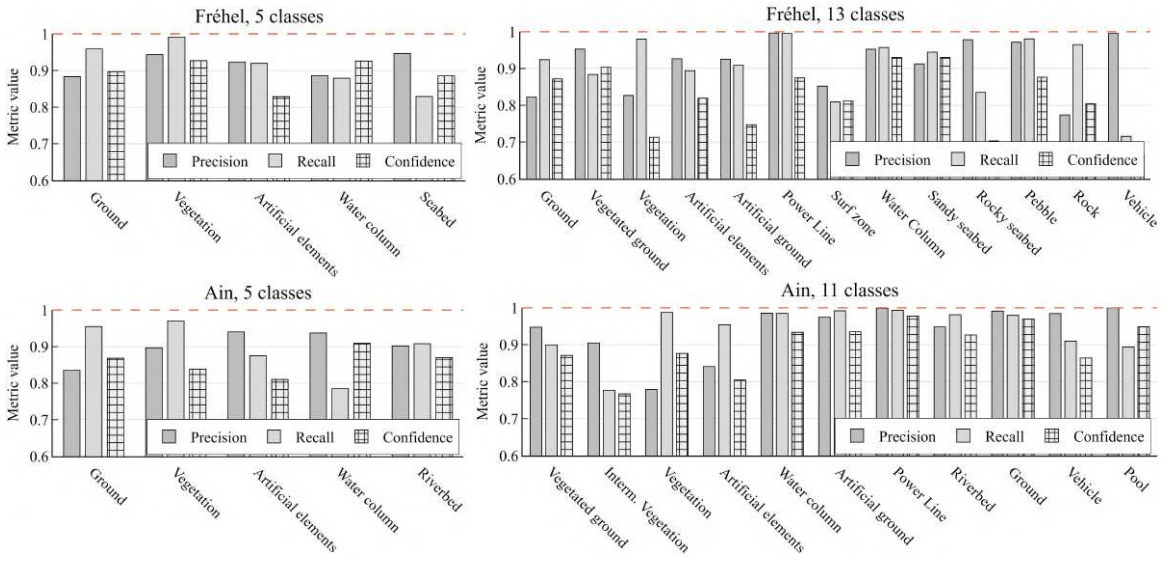


Figure 3.10: Precision, recall, and prediction confidence per class for the four classifiers after optimization.

Classifier	Selected scales
Ain, primary	1.5 m, 4 m, 5.5 m, 14 m, 10NN
Ain, advanced	1.5m, 2 m, 2.5 m, 3 m, 3.5m, 4 m, 5.5 m, 7.5 m, 5NN
Fréhel, primary	3 m, 3.5 m, 5.5 m, 6 m, 14.5 m, 1NN
Fréhel, advanced	3.5 m, 4.5 m, 6.5 m, 7 m, 14.5 m, 1NN

Table 3.3: Remaining scales in the four optimized multi-scales classifiers. kNN indicates k nearest neighbours.

The *surf zone* is also challenging to distinguish from *ground* or *rocky seabed* in some areas of Fréhel.

Some classes show gaps between precision and recall, reflecting the over-detection of *building* in rocky areas or of *intermediate vegetation* in the Ain (see Figure 3.9).

Dominant scales analysis

The optimized predictor vectors indicate that some features are particularly informative at specific scales, and conversely, some scales are essential for given features only. The optimization phase alters the systematic multi-scale character of the classification since the number of predictors in the optimized models is smaller than the product between the number of scales and the number of features. For example, the advanced classification of the Ain has an optimized predictor vector exploiting 16 features at nine scales, yet, its total size is 37. In contrast, if the optimized classification systematically used all available scales for a feature, it would be 144.

Table 3.3 sums up the specific scales retained for each experiment. It shows that finer scales are necessary to describe the Ain site: the minimal scale selected is 1.5 m, whereas it is double for Fréhel. All classifiers follow a similar pattern: they exploit small to medium scales up to about 6 m, and a much larger scale of about 14 m without transitioning via a medium value. The advanced models both reuse similar scales to their primary equivalents but incorporate new ones in between, reducing the typical sampling step of object sizes. However, the 11-label classifier of the Ain is the only one discarding the 14 m scale, thus exploiting only small to medium diameters.

To better identify the contribution of specific scales to various classes in the two environments, Figure 3.11 shows the Shapley analysis for the standard classification.

Dominant scales are different between the Ain area and Fréhel. Water column and seabed/riverbed are dominated by features computed with around 3 m, 6 m and 14.5 m diameter in Fréhel, whereas 6 m and 10NN features are more useful in the Ain. Similarly, artificial elements and trees do not exploit the

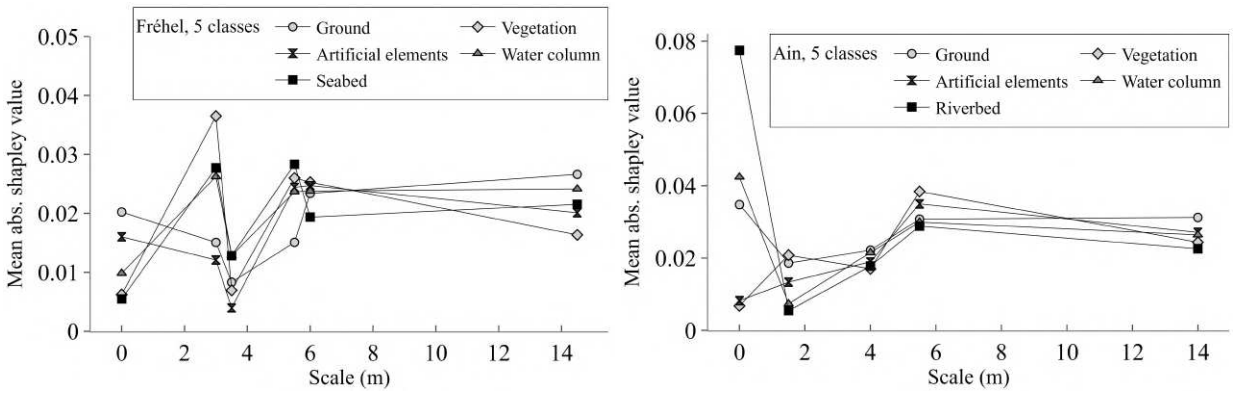


Figure 3.11: Mean absolute Shapley value obtained by each scale of the optimized predictor vector depending on the class considered (a scale of 0 m represents features computed with a kNN search).

same sphere sizes over the two sites. The scales also adapt to each label. For example, *artificial elements* – containing buildings, vehicles, and power lines – rely less on kNN features than *riverbed* in the Ain or *ground* in Frehel.

We can also identify two groups of classes having similar scale contribution patterns. The first includes *water column* and *seabed/riverbed*, and the second includes *ground* and *artificial elements* in Fréhel, while it is composed of *ground* and *vegetation* in the other area. In Fréhel, *vegetation* follows similar trends as the bathymetric classes.

Dominant features analysis

To simplify, we only review the dominant features of the primary classifications in this section. Several features stood out from the rest and passed both the selection and optimization phases. They theoretically contain the essential information to distinguish the defined classes. Table 3.4 introduces and illustrates each of them.

The optimized sets of predictors obtained, presented in Figure 3.12 and Table 3.4, seem to be tailored to each site. The Shapley analysis, in Figure 3.12, corroborates this observation. **Only five features common to both sites are identifiable**: *vertical distance of green points to their NIR neighbours (kNN)*, *mean green intensity*, *mean echo ratio in the green neighbourhood*, *mean number of returns in the green neighbourhood*, and *mean number of returns in the NIR neighbourhood*.

Two groups of labels have similar feature contribution patterns. *Ground*, *seabedriverbed*, and *water column* on one side and *vegetation* and *artificial elements* on the other. The first group is mainly identified by multi-echo features, NIR intensity, and dual-cloud features. The second relies primarily on dual cloud features – median intensity differences namely – and NIR multi-echo attributes.

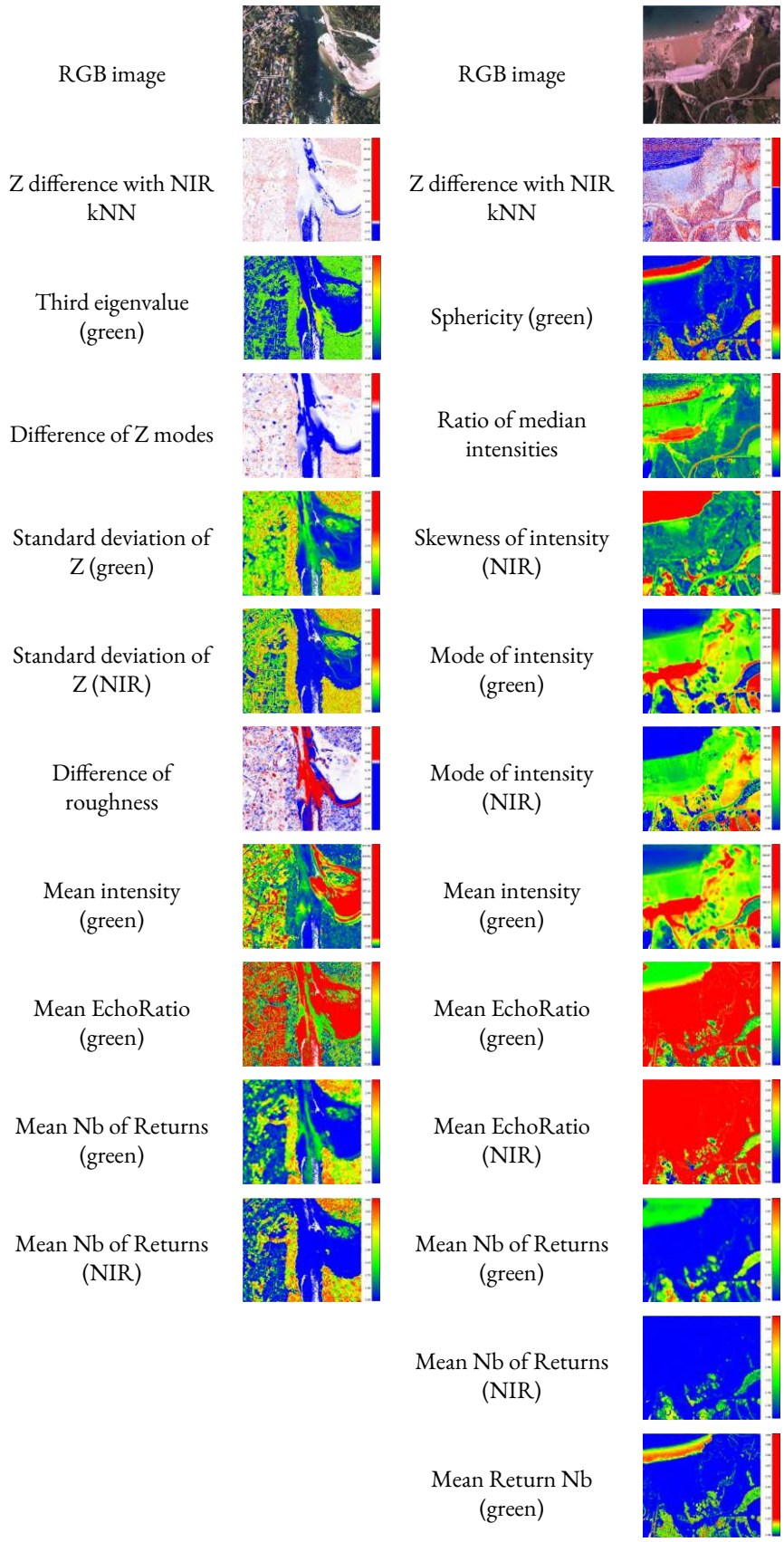


Table 3.4: Optimized set of features for both sites. For dual-cloud features corresponding to a difference, blue values are negative, white is zero, and red is positive. For other features which are strictly positive, blue indicates lower values, green intermediate values, and red higher values.

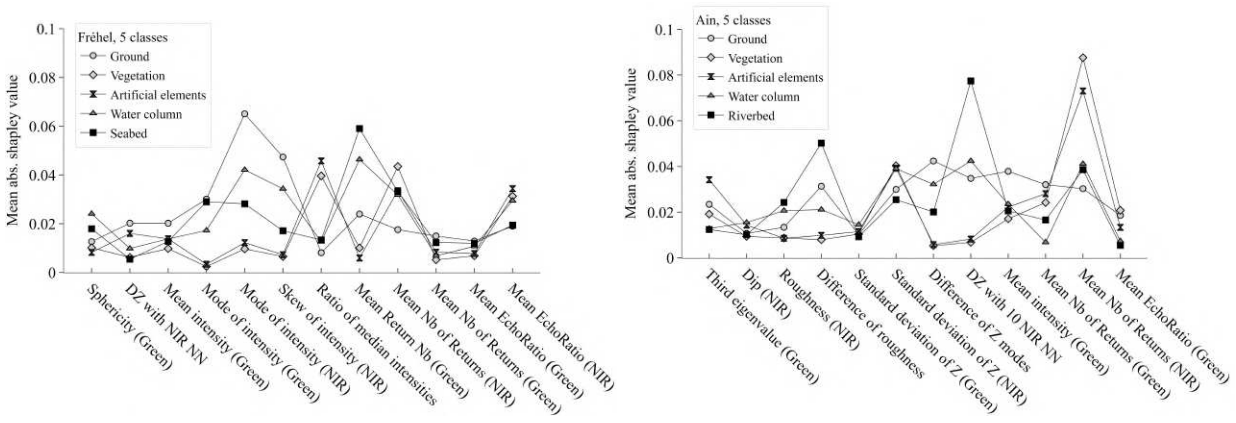


Figure 3.12: Mean absolute Shapley value obtained by each feature of the optimized predictor vector depending on the class considered.

In both cases, **the TB aspect of the datasets is fully exploited**: in both areas, there are four NIR PC features and five to six green PC attributes involved in the optimized sets. NIR PC-derived features are more contributive to topographic objects, while both PCs are equally crucial for *ground/seabed/water column* distinction. The experiments on Fréhel also draw more on NIR intensity-derived parameters than the models to process the Ain, in which only one green spectral parameter is involved with low relative importance (Figure 3.12). The class-wise feature importance analysis also shows that **features do not have the same descriptive power in both NIR and green domains**. The number of returns of the NIR echoes is more informative on the nature of the surface than its green equivalent.

Both results show a **predominance of newly introduced 3DMASC features over classical features** used in other studies (Chehata et al., 2009; Hackel et al., 2016; Thomas et al., 2018; Weinmann et al., 2015). 8 out of 12 for the Ain site and 10 out of 12 for Fréhel are attributes we propose with 3DMASC: means, modes, or skewness values of PC characteristics. Geometrical and dimensionality-based features are scarce: only NIR PC roughness, NIR PC dip, green PC sphericity, and green PC third eigenvalue pass the optimization phase. The mean green intensity is the only other example of a classical feature observable (see Table 3.4). Intensity-based features constitute nearly half of the predictors of the Fréhel optimized classification but are few in the Ain model. The other half of the Fréhel predictor set is dominated by multi-echo features of both wavelengths, that evict height-based features and geometrical features. In the Ain, they appear through the differences in elevation modes. Other dual cloud features stand out: *vertical distances between green and NIR points*, *elevation mode differences*, and *median intensity differences* between the two PCs (Table 3.4).

Figure 3.12 also reveals that the **new 3DMASC features outperform usually dominant characteristics** like intensity. The difference in elevation modes between NIR and green PCs, is more relevant

to identify vegetation than intensity, in particular in the Ain. Similarly, the roughness difference between PCs dominates single cloud NIR roughness, even for ground and over-ground object separation. The ratio of median NIR and green intensities is particularly useful for outlining vegetation and artificial elements. Dual cloud features are present in both OC classifiers, illustrating how they complement separate single cloud attributes. Multi-echo features also contribute significantly to the predictions. The mean number of returns is particularly useful to characterize *vegetation* and *artificial elements*.

3.4.3 Results using other predictors

In this section, we test 3DMASC in different settings: using a context PC, RGB information, and simulating the unavailability of the NIR wavelength. All results are summed up in Figure 3.13. They are obtained by running the complete framework on initial predictor vectors including contextual features, RGB-derived features, or green features only. The presence or absence in the optimized predictor set of each tested attribute is thus already an indication of their informative character.

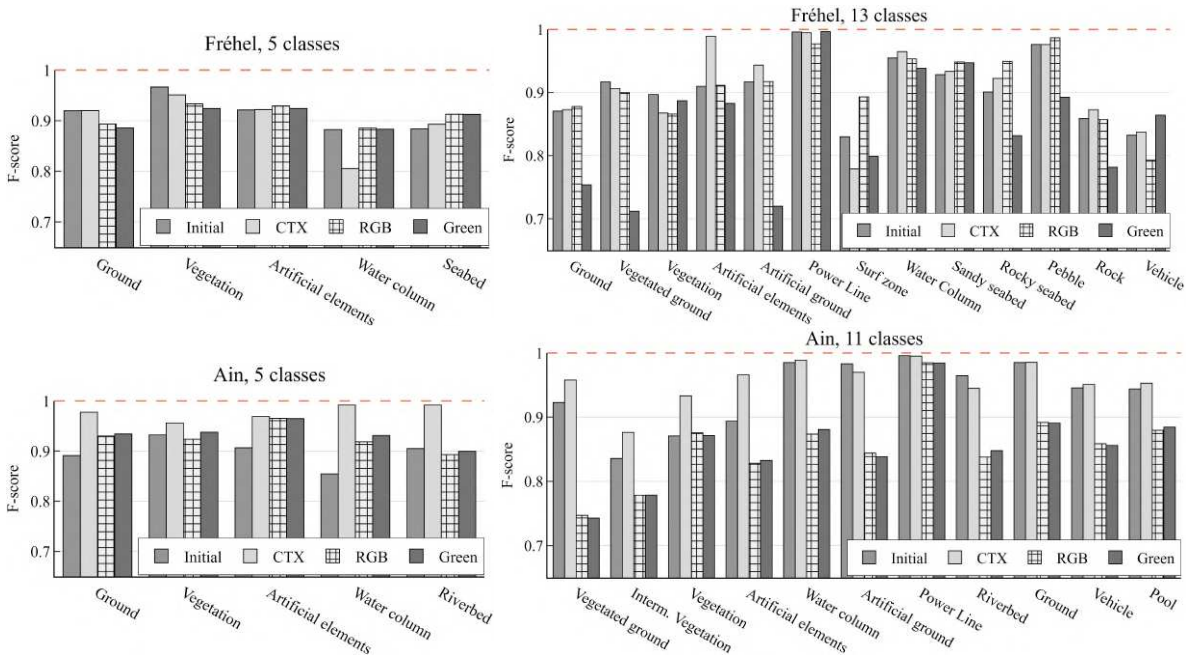


Figure 3.13: F-score obtained for each class depending on the experiment. Initial = optimized classifier obtained with the initial set of predictors. CTX = optimized classifier obtained when adding contextual features. RGB = optimized classifier obtained with RGB features added. Green = optimized classifier obtained using only green features.

The contextual features used were vertical distances to a PC containing only ground or water surface points, for different scales (1 m, 3 m, 5 m, and 10 NN). These predictors allowed to use smaller scales (see

Supplementary Materials) and improved the prediction confidence and quality of all classes in the Ain area. In Fréhel, they improved the accuracy of *seabed*, but tend to penalize *water column* and *vegetation*.

The reflectance in the blue domain is the only RGB-derived attribute that passed optimization. Its mode is used in two models: Fréhel primary and Fréhel advanced. This shows that RGB features are not crucial to detect the classes of the Ain but may serve to differentiate coastal land and sea covers. They also seem to penalize our classifier optimization framework when they are used but do not appear in the best models, as the losses in F-scores on the advanced classification of the Ain reveal. This shows that RGB parameters may evict some more useful features and reduce the classifier's abilities on unseen data. However, in the primary case in this area, F1 scores are higher when the optimization is performed on this extended feature set: when including the RGB attributes, the OOB significantly drop when the number of predictors is below 40, causing the presence of additional intensity-based 3DMASC attributes in the optimized classifier, which seems to increase F1 scores of several classes.

When using green laser data only, OAs range between 85% (Fréhel, advanced) and 94% (Ain, primary). Predictor vectors are dominated by multi-echo features and intensity-derived attributes. In Fréhel, dip and standard deviation of intensity are the only new features selected. In the Ain, point-based echo ratio, mean return number and mode and standard deviation of intensity appear. Overall, more scales are used per feature and seven and eleven features are selected for Fréhel and the Ain, respectively. In the 5 class experiment of the Ain, a performance decline is only observed for *riverbed*. Although its F-score drops by 2%, it remains at 90%, showing that a single bathymetric PC already provides highly accurate detections of the *water column* and the *riverbed*. In Fréhel, the classification of *seabed* is even improved when excluding NIR data. In both settings, the distinction of topographic classes is less accurate when discarding NIR information.

3.5 Discussion

Starting with a set of 88 features computed at 29 scales, we obtained optimized, compact classifiers ranging from 18 to 37 predictors – scales and attributes – and resulting in good to excellent classification for up to 13 classes. In this section, we discuss these results with respect to existing work on PC classification.

3.5.1 Classifier optimization and number of predictors used

Through 3DMASC, we obtain classifications of TB scenes with OAs over 90%, using light classifiers that harvest a maximum of 37 predictors (Table 3.2), some of them using as little as 18 predictors (Ain, 5 classes). Average prediction confidence is high and accompanied by high accuracy, a synonym of effective classifier training. Low values of confidence can be linked with classification errors and used to filter out

Classifierr	Confidence threshold	OA	Remaining points
Ain, advanced	0.5	95%	96%
	0.6	97%	92%
	0.7	98%	87%
	0.8	98%	80%
Fréhel, advanced	0.5	94%	92%
	0.6	96%	84%
	0.7	97%	76%
	0.8	97%	67%

Table 3.5: Overall accuracy depending on the confidence threshold applied to filter the predictions.

misclassified points. Table 3.5 shows the results of applying a confidence threshold below which points are removed. It illustrates that there is necessarily a balance to find between result quality and spatial resolution of the classified PC, as aiming at fewer classification errors means accepting to reduce the local density of the data.

The optimization step seems to efficiently balance computational efficiency and high-quality classifications. The low number of predictors makes the models applicable to large datasets, easily explainable with Shapley values and thus accessible to non-specialist users. These characteristics allow 3DMASC to be an interesting alternative to current state-of-the-art methods that are 3D deep neural networks. Discussing the performances of deep neural networks is out of the scope of this study, but they have proven to output significantly good results on 3D semantic segmentation applications (Guo et al., 2021b). However, to our knowledge, no available and accessible deep learning framework for multiple 3D PC classification or for airborne lidar data in similar natural settings exists that could be easily reused by environmental scientists without a significant background in deep learning. Indeed, the developments still mostly focus on urban areas (Huang et al., 2021; Lin et al., 2021; Mao et al., 2022a; Schmohl et al., 2019; Wen et al., 2021; Yang et al., 2018; Zeng et al., 2023; Zhang et al., 2022; Zhao et al., 2018a), where the variety of scales and geometry is quite dissimilar to what we observe in natural, TB settings. Neural network hyperparameters are also harder to optimize without expert knowledge and require more complex and intensive training, and thus, computing power and are thus less easy to master for thematic users.

In this context, we addressed the need for an approach allowing for high-accuracy multi-class clas-

Samples per class	Overall accuracy			
	Ain		Fréhel	
	5 cl.	11 cl.	5 cl.	13 cl.
1600	98%	95%	91%	91%
1200	98%	95%	91%	91%
800	97%	95%	91%	90%
400	96%	95%	90%	90%
100	94%	93%	89%	90%

Table 3.6: Classification accuracy depending on the number of training samples used. Tests are performed using the complete set of 3DMASC features

sification while relying on less complex computation, and staying accessible to environmental scientists through open-source software. We chose to experiment on datasets containing 2000 labelled points per class, but when randomly subsampling the labelled data, we observe that high accuracies are already possible with a few hundred ground truth points per class, as featured in Table 3.6.

Neural networks are also more abstract and thus harder to decipher, contrary to 3DMASC thanks to feature importance and Shapley values that contribute to an explainable machine learning approach. Indeed, we refer to the definition of Roscher et al., 2020, identifying transparency, interpretability, and explainability as major traits of explainable machine learning. In this work, transparency is addressed through the detailed description of the method construction choices and the exploration of some key parameters such as feature selection choices, scales choices, etc. Interpretability is addressed through extensive analysis of the features' importances, the iterative pruning of the descriptor vector, the exploration of meaningful scales and the consideration of two different environments and datasets. Finally, explainability is considered through the analysis of the decision process in the two environments depending on the class – typical feature and scales for different classes – but also through the experiments on two datasets and the application to much larger sets of points, giving insights on the robustness and reproducibility of the method.

3.5.2 Dominant scales

Taking advantage of the explainability of the method, we identify typical characteristics of OC classifications. First, **a typical set of scales emerges from the experiments**, including small and medium sphere diameters ranging between 1.5 m and 7.5 m and one larger scale around 14 m (Table 3.3). The global range of scales selected does not vary between primary and advanced classifiers, except for the Ain where we can expect that the introduction of smaller-scale objects - *vehicles*, *swimming pools*, *intermediate vegetation* – penalizes very large scales. Advanced classifiers rather add scales within the core range, reducing the step between two options.

Second, **the exact optimized sets of scales that arise are specific to each environment**, which questions the possibility of identifying optimal neighbourhoods without analyzing their application context. For example, out of four experiments, three different optimized NN neighbourhoods stand out: 10, five, and one (see Table 3.3), contrasting with the conclusions of Niemeyer et al., 2011 that select one single optimal scale of seven NN for their different experiments, and with the results of Dong et al., 2017 who also find a single scale of five NN as the most often selected neighbourhood. Furthermore, the fact that each selected scale is not used for each feature tends to be consistent with the work of Dong et al., 2017, choosing to optimize each feature's neighbourhood rather than identifying a global optimal scale.

Third, **scale selection results are consistent with the intra-feature correlations** we observed in Figure 3.6. Although these estimates could be complemented with other measurements of non-linear correlations as in (Weinmann et al., 2015), this first consideration for correlation already provides insightful information on the relevance of the different scales for our study areas. These suggested that less scales were needed above 6 m than below, which is in line with the fact that we only obtain one large scale. This large scale also outlines the necessary **trade-off between classification accuracy and classification resolution**. If we investigate the role of this much larger scale, we find that, though it helps to mitigate some errors linked to larger scale roughness in the PCs – for example confusion of rocks with buildings – it also smoothes out the results, blurring classes borders and even missing smaller objects like cars. In Figure 3.14, cars can be identified in the PC, but many of them are missed and labelled as ground when large scales are used. Limiting the range of scales to 7 m produces a result in which these cars are correctly detected, but the ground incorporates false *building* labels.

Our observations thus **question the relevance of large scales**, which appear to be selected for certain point types as they pass the score filtering selection, but end up penalizing the global classifier application through several aspects. Table 3.7 illustrates the confidence filtering analysis obtained on the Fréhel advanced classifier optimized on scales within 1 to 7 m only. It shows that, without the possibility to select larger scales, the classification reaches similar accuracies and confidences. However, they clearly affect the computation efficiency. The advanced Fréhel OC classifier obtained on scales up to 7 m incorporates ten

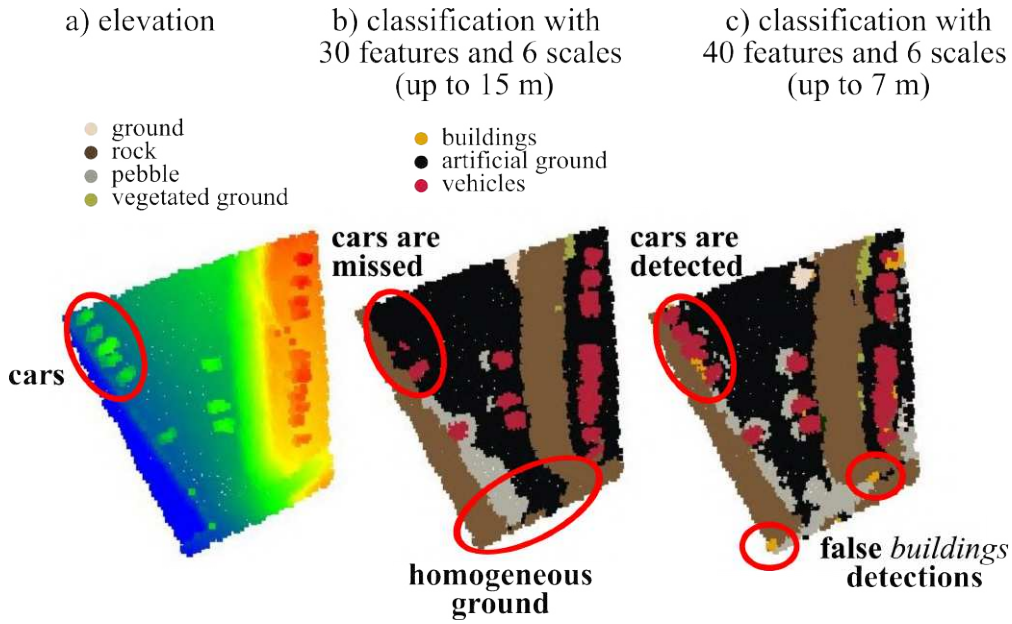


Figure 3.14: Extracts of classification results obtained depending on the maximal scale included.

more features, but the computation time is divided by three (3450 points per second versus 1102 points per second). Suppressing large scales may thus improve classification speed, while maintaining high OAs.

3.5.3 Computation time

Computational efficiency is an important aspect of 3DMASC. The computation of the spherical neighbourhoods is the main bottleneck of the workflow, similar to what is observed in other studies (Hackel et al., 2016; Weinmann et al., 2015) and sometimes even drives the choice of the neighbourhood type. Table 3.8 illustrates the time necessary to compute all implemented features at different single or combined

Classifier	Confidence threshold	OA	Remaining points (%)
Fréhel, advanced (Max scale = 7 m)	0.5	94%	91%
	0.6	96%	84%
	0.7	97%	76%
	0.8	98%	66%

Table 3.7: Overall accuracy depending on the confidence threshold for a reduced set of possible scales.

	1 m	4 m	7 m	10 m
Single scale (pts/s)	28082	7337	2109	928
		1-4 m	1-7 m	1-10 m
Multi-scale (pts/s)		4069	854	275
			Up to 7 m	Up to 15 m
Optimised multi-scale (pts/s)			3450	1102

Table 3.8: Feature computation time depending on the scale set.

scales.

These results were obtained using a computer equipped with 128 Go of memory and a 12 core AMD Ryzen ™ 9 5900X CPU. The test file was the data of the Ain area; it contained 106 410 018 green points, 61 043 388 NIR points, and 5 700 844 core points having a 1 m spacing. Table 3.8 shows how crucial scale selection and optimization are: without optimization, computing scales from 1 m to 10 m lasts 5 hours and 45 minutes (275 pts/s). After optimisation using scales up to 7 m, the computing time drops to $\simeq$ 28 minutes (3450 pts/s). This computation speed could be increased by implementing pyramidal computation into the 3DMASC plugin, which consists of subsampling the data when increasing the neighbourhood size as made by Thomas et al., 2018.

Additionally to the selection of a scale range, the number of different diameters within the interval, and the number of features to compute for each neighbourhood also has an impact – though less significant – on the processing time. Table 3.8 shows that the optimised descriptor set relying on scales up to 7 m is three times faster to compute than the complete set of features on scales between 1 and 7 m. Consequently, although predictor selection is not crucial for classification performance (see Table 3.3), it is essential to the practical applicability of the method.

3.5.4 Class-wise results: dominant features

Five of the maximal 12 features needed to perform basic classification are common to both experiments. These are *multi-echo features computed on both PCs*, *vertical distance of green points to their NIR neighbours (kNN)*, and *mean green intensity*. They are then combined with site-specific attributes. In Fréhel, the optimized predictor set retains mainly multi-echo attributes and intensity-derived information. In the Ain, multi-echo features and height-derived parameters dominate. However, **classical features of 3D data interpretation** such as dimensionality-based features delineating the shape of local PCs from the combination of eigenvalues (Brodu et al., 2012; Gross et al., 2006; Vandapel et al., 2004; Weinmann et al., 2013) **are almost unused**. They only become more prominent when complexifying the number and

types of classes to detect. This is also certainly linked to the fact we analyze airborne lidar data, while these features were designed in priority to describe terrestrial and mobile laser scanning, that include a greater diversity of surface orientations.

Point-based attributes are also absent from the optimized classifications. Newly introduced features based on statistical operators applied to multi-echo features or intensity values systematically outperform them in terms of contribution. Such operators had been tested on height-derived values (Antonarakis et al., 2008; Dong et al., 2017) but never applied to other types of features. **The use of statistical operators is particularly informative and able to drastically improve the informative power of point-based characteristics**, namely multi-echo attributes, that never particularly stood out in existing PC classification literature but appear essential to the success of our experiments. We interpret this result as a way for decision trees to compensate for their inability to consider spatial relationships between points and include a level of spatial consistency of the considered attributes. These **operators also limit bias linked to intensity values, that are unavoidable** to classify diverse environments (Song et al., 2002; Yan et al., 2015). Intensity median, mode, skewness or ratio values constitute half of the primary predictors in Fréhel, and are prominent in both advanced models. By being less sensitive to outliers, standard deviation and skewness mitigate the limitations of this measure, which varies with the acquisition conditions and does not constitute an absolute estimation of surface reflectance (Kashani et al., 2015). Overall, **the features we present seem to describe natural environments better**.

We compared classifications of the Ain obtained with 3DMASC features and with features used in Thomas et al., 2018; Hackel et al., 2016; Chehata et al., 2009; and Rusu et al., 2009. In order to ensure a fair comparison, all features were computed at the same scales and on the green PC only, and then classified with a RF model. Consequently, the 3DMASC version compared to these existing methods only includes single cloud features. In practice, only the features differ, with 3DMASC incorporating spatial statistics of multi-echo attributes and all other features, but no information about the NIR cloud – and thus no dual cloud feature, or contextual feature.

Additionally to the point feature histograms developed in Rusu et al., 2009, the compared approaches rely mainly on features derived from the covariance matrix of the core point's neighbourhoods, on height-based parameters, and, less frequently on echo-based parameters. Due to the unavailability of waveform data on our test areas, we did not include waveform-derived attributes originally exploited in Chehata et al., 2009.

We computed each feature set on the green PC only and at multi-scale spherical neighbourhoods with diameters of 2, 3, 4, 5, 6, and 7 m. The feature point histograms were computed only on the green PC too, with a normal computation scale of 0.5 m and a feature computation scale of 2 m. Histogram-based features also involve the spatial repartition of features (Osada et al., 2002; Rusu et al., 2009; Tombari et al.,

Dual cloud	Single cloud				
Dual-cloud 3DMASC	Single-cloud 3DMASC	Thomas et al., 2018 (covariance-based)	Hackel et al., 2016 (covariance- and height-based)	Chehata et al., 2009 (covariance-, height-, echo-, plane-based)	Rusu et al., 2009 (fast point feature histogram)
97.6%	93.7%	74.3%	82.6%	84.9%	71.3%
With single-cloud 3DMASC		93.7%	93.7%	93.7%	93.7%

Table 3.9: Classification overall accuracies obtained with different types of single-cloud features on the 5 classes of the Ain dataset.

2010). Although they have been used for classification before with satisfactory results (Arbeiter et al., 2012; Blomley et al., 2016; Blomley et al., 2017; Garstka et al., 2016; Himmelsbach et al., 2009; Wohlkinger et al., 2011), we made the choice not to include them in 3DMASC to avoid increasing the predictor selection difficulty, as they require the choice of two scales – one to compute the normal, the other to compute the features – and selecting the number of bins to use. Their higher computation time also played a role in this decision (Garstka et al., 2016). Details about the features used in each experiment are provided in Supplementary Materials. Overall Accuracies obtained on the test set for 5 classes in the riverine area by each approach are summed up in Table 3.9.

These results show that in natural environments, using our features produces systematically higher results than other existing features and that none of these features, in particular fast point feature histogram improved the single cloud 3DMASC classification results. They also show that using solely covariance-based features produces among the lowest OA in our riverine environment, while it generated more precise classifications of urban environments (Thomas et al., 2018), highlighting the need for methods adapted to the different types of 3D data currently in use.

We also **introduce new measures of the optical behaviour of the surfaces present** in the PCs, which were mostly estimated through mean intensity, and propose new inter-channel ratios to complement existing multispectral attributes (Morsy et al., 2017b; Wichmann et al., 2015). Previous studies analyzing multispectral lidar faced the difficulty of linking points to their equivalents in PCs of other wavelengths since they are never in strictly identic positions due to the sensor configuration (Lague et al., 2020). These new ratios, along with our dual-cloud features compensate the limits of point matching, used in existing multispectral lidar analysis work (Morsy et al., 2017b) when they are used on datasets with correct geometrical and radiometric calibration (Kashani et al., 2015; Yan et al., 2012).

Dual-cloud features systematically stand out among highly contributive features. Their lower

inter-scale correlation likely contributes to their more informative character, along with their ability to compensate for the limits of shallow learning classifiers, that are unable to learn features, and thus to bring out and use connections between features. For example, a difference in roughness between NIR and green PC is particularly high for points belonging to the water column, and much lower for the riverbed or the bottom of swimming pools, due to the full reflection of NIR laser on the water surface and scattering of the green light in the water column. The same optical phenomenon explains the higher difference in elevation modes between PCs in *swimming pools* and *riverbed*. The inherent points position differences of TB sensors, illustrated in Figure 1.7, explain the varying vertical distances between green and NIR PCs in vegetated areas, and their systematically negative value in bathymetric zones. Similarly, **the use of a previously classified ground PC as a contextual feature** allows for improvement in the labelling of points at the limit between ground and above-ground features, namely building walls and lower tree branches, explaining the improvement observed when they are included, and the smaller scales needed to capture the signature of such variations.

Using these observations, we **recommend the following set of features to use on topo-bathymetric environments**: the NIR and green *number of returns* and *echo ratios*, the green *return number*, the *vertical distance* to the 1 and 10 *nearest neighbours of the core points in the NIR PC*, the *mode* of the green and NIR *intensities*, the *skewness* of the NIR *intensity*, the *ratio of median intensities*, the NIR and green *elevations' standard deviation*, the *difference of elevation modes*, the *NIR roughness* and the *difference of roughness*, the NIR *dip*, and the green PC *sphericity*. With these 19 features computed at scales between 1.5 and 14 m, we observed OAs of 98% and 91% for 5 classes on the riverine and coastal datasets, respectively, and 94% and 90% on their 11 and 13-class versions.

3.6 Conclusion

In this chapter, we have introduced 3DMASC, a method for explainable machine learning multispectral point cloud classification. 3DMASC operates directly on sets of unordered, unstructured points and predicts a label for each, with a confidence index and information on the origin of the decision, through feature importance. It differs from previous point cloud classification methods in its capacity to handle multiple clouds simultaneously and describe the spatial and statistical repartition of point cloud attributes, introducing indirect context consideration in the model and new multispectral feature ratios. 3DMASC also stands out from state-of-the-art 3D classification methods with its accessibility: it is explainable using Shapley values, usable without dedicated GPUs, and easy to handle for thematic specialists such as geomorphologists, ecologists, or cartographers. We focus on providing compromise in terms of computation cost, processing time, complexity, and resulting metrics, with respect to the current state-of-the-art

methods. We demonstrate the performance of the approach on two different airborne lidar use cases: the detection of land and sea covers in (1) a fluvial environment and (2) a coastal area. Results show that the method produces highly accurate classifications of basic or detailed categories of points. Furthermore, models excel in TB environments thanks to the newly introduced features and require a limited number of training points ($\leq$ 2000 per class), scales, and attributes. We also implemented a feature selection framework that allows us to draw three main conclusions about the definition of the predictor's vector: (1) statistics of point-based attributes are more informative than classical dimensionality or geometrical features on this type of data, (2) multi-echo features, vertical distances between the two PCs and mean intensities appear to constitute an essential base of features to use and (3) dual cloud features are highly contributive to separate ground, artificial elements and vegetation. Our results also stress the superiority of multi-cloud classification compared to single-cloud, especially for bi-spectral lidar. We release our source code through an open-source plugin in CloudCompare (Girardeau-Montaut, 2022), hoping it will help applications of 3D remote sensing for earth observation and conservation. Although this chapter illustrates specific use cases of the workflow on topo-bathymetric lidar datasets, 3DMASC can be extended to PC time series analysis, and 3D data interpretation in general. It may be applied to terrestrial laser scanning data, to structure from motion PCs, or even to drone lidar sensors, which are still too compact to incorporate dual-wavelength lidar sensors but already enlarge the access to lidar surveys.

CHAPTER

4

EXTRACTING SEMANTIC INFORMATION FROM LIDAR WAVEFORMS

Contents

In Chapter 3, we exploited the geometric information contained in 3D PCs to perform semantic segmentation. However, as explained in Chapter 1, lidar surveys initially produce much more detailed data about the terrain surveyed: the full waveforms. Although the availability of this information is a major issue and constitutes a limitation to their systematic use for lidar data classification, waveforms provide additional knowledge on the structure and nature of the surfaces surveyed. Each object of the surveyed environment illuminated by the sensor's laser reflects light in a specific way, generating a characteristic signature in these signals. For specific applications, waveforms can thus be particularly useful.

In this chapter, we investigate how these raw spectral measurements can be used to derive semantic information about topo-bathymetric (TB) environments and we propose methods to perform 3D mapping of land- and sea-covers based on lidar waveforms.

As introduced in Chapter 2, when it comes to learning-based supervised classification or semantic segmentation, there are two main possibilities: learning separation rules of handcrafted features computed upstream or learning data representations and their separations into classes simultaneously. In this chapter, we investigate both approaches, as there is little information available in the scientific literature about land-water interface classification with TB lidar waveforms. We propose two workflows based on bispectral lidar waveform features to map the land-water continuum seamlessly. With these methods, we aim to bridge the gap between marine and terrestrial surveys, to demonstrate that efficient methods can be developed to automatically map the land-water interface surface covers, and to show that an integrated vision of coastal zones is feasible and advised. The results and methods presented in the following Sections resulted in three conference proceedings: one for the International Geosciences and Remote Sensing Symposium (IGARSS) of 2021 (Letard et al., 2021b), another at the occasion of IEE OCEANS conference of 2021 (Letard et al., 2021a), and the last during the International Society of Photogrammetry and Remote Sensing Congress of 2022 (Letard et al., 2022b). The methodology presented in Section 4.2 was published as a paper for a special issue on coastal ecosystems observation in MDPI's Remote Sensing in 2022 (Letard et al., 2022a).

4.1 Context

4.1.1 Methodological context

Though many methods have been proposed to process airborne topographic lidar waveforms (Mallet et al., 2009; Mallet et al., 2011; Reitberger et al., 2009; Zorzi et al., 2019), bathymetric lidar waveforms are, to the best of our knowledge, much less explored. They are often only analyzed to retrieve bathymetry, and are even less employed for classification tasks. Yet, these data include more detailed information on the

physical properties of the environment surveyed than the elevation contained in rasters or PCs, by registering also the way the Earth's covers interact with light. This information is particularly useful to classify covers that have distinct spectral signatures (e.g., Letard et al., 2022a) or subtle geometric features at the decimeter scale that discrete echoes cannot capture (e.g. Launeau et al., 2018). Exploiting full-waveform lidar data requires adapted processing methods, as this sensor produces rich but complex information on the environment, with sometimes dozens of point records and several waveforms of up to a thousand samples per square metre. Although efficient tools exist to process lidar-derived rasters or PCs, methods to efficiently exploit the knowledge enclosed by the waveforms are still expected, as they remain mainly experimental. Lastly, we pointed out the absence of solutions for bi-spectral PC processing; the same issue persists for bi-spectral waveform data exploitation, which is still largely under-explored (the only experiment on bi-spectral waveform features being that of Leigh et al., 2016 using orthowaveforms).

Classification of land or water covers using lidar data has been scarcely explored recently. Even when using waveform data, most of the published research is based on 2D data classification (Wedding et al., 2008; Teo et al., 2017; Collin et al., 2012; Letard et al., 2021b; Eren et al., 2018) while fewer articles exploit them directly in 3D (Hansen et al., 2021; Tulldahl et al., 2012; Chehata et al., 2009). Many studies researching ways to classify lidar data used machine learning algorithms such as support vector machine (SVM), maximum likelihood, or random forests. Currently, there is a consensus on the efficiency of random forests thanks to their ease of application to large datasets and their low tendency to overfit (Yan et al., 2015). Furthermore, random forests offer the possibility to retrieve information about the predictors' contributions easily.

In Section 4.2, we thus tackle the problem of **bi-spectral waveform classification** for the identification of multiple surface covers **in TB environments**. To this end, we exploit the strengths of random forests and implement a feature extraction process to derive 3D maps of marine and terrestrial habitats without interruption between land and water.

However, the exploitation of full-waveform data often relies on backscatter detection (Klonowski et al., 2007; Sandidge et al., 1998; Kutser et al., 2006; Wedding et al., 2008) followed by feature extraction and classification (Klonowski et al., 2007; Adler-Golden et al., 2005; Collin et al., 2012; Eren et al., 2018; Tulldahl et al., 2012). Whether based on simple peak detection (Collin et al., 2012) or on decomposition approaches (Song et al., 2019; Shen et al., 2017), the necessary identification of waveform components is often application-dependent and requires advanced settings (Sandidge et al., 1998; Kutser et al., 2006). Both traditional peak detection and feature extraction methods rely on computations that are very sensitive to noise (Kutser et al., 2006) which may hinder the obtention of reliable and generalizable results.

In Section 4.3, we thus explore a **semantic segmentation** approach to **classify each waveform component in one shot**, without having to pre-detect them with decomposition methods.

Before diving into the specifics of both methods, we present the study area and the datasets used in the following paragraphs.

4.1.2 Datasets used for methodological developments

Study area

The study area was chosen along the northern coasts of Brittany, France, near the town of **Fréhel**, for its ecological diversity and because of the availability of full-waveform lidar data acquired by the French Hydrographic Office (Shom) as part of the Litto3D® project (Pastol, 2011). This seaside region is set in an ecologically rich environment encompassing: fine sand and pebble beaches, a sandy dune, rocky areas provided with seaweeds, seagrass meadows, wooded areas, crop fields and salt marshes through which a river flows towards the sea. These ecosystems host a great variety of species: shellfishes, endemic dune plant vegetation, green, red or brown seaweeds, *Zostera Marina* plants, evergreen and deciduous trees, crops and endemic salt marshes plants such as *Halimione portulacoides*, sea poa or purslane. There are also typical urban land covers, due to the presence of a small resort town, **Sables d'Or les Pins** (48°38'27"N: 2°24'24"W). Buildings, tar or concrete-covered paths, boats in mooring, and vehicles in parking lots are thus also present in the selected zone.

Data acquisition

The lidar data[1] used for this research were acquired over the coast of **Sables d'Or les Pins** in **September 2019** by the Shom as part of the Litto3D® project (Pastol, 2011), using a Leica HawkEye III 4X sensor. This sensor produces laser pulses at **wavelengths of 513 nm and 1064 nm** on three different channels. To reach dozens of meters deep seabeds, the system is set to emit more powerful beams in the green wavelength, compromising on point density to meet both accuracy and safety requirements.

Depths under 10 m have a dedicated shallow green laser, while a more powerful laser, the deep channel, is used to detect deeper seabed. These two channels provide PCs with a density of at least five points per m^2 and one point per m^2 and they have a laser spot size diameter of 1.8 m and 3.4 m, respectively. The NIR laser has a laser spot size of 0.2 m and a point density of at least 10 points per m^2. Each green waveform and every 32 NIR waveforms were recorded with a time-frequency of 1.8 GHz, resulting in a backscattered intensity value every 556 picoseconds. The survey was conducted with constant laser amplification. Due to the power needed to penetrate through several meters of water, the shallow laser' backscattered intensities tend to be saturated over highly reflective land surfaces, but they are still usable for surface

1. Available online at https://diffusion.shom.fr/donnees/altimetrie-littorale.html

characterization. The deep channel's returned intensities, however, are systematically saturated and do not provide usable information for land cover classification.

Over the studied area, there are on average 6.7 green waveforms and 0.5 IR waveforms per square metre. The reanalyzed echo PC of the shallow green wavelength was also used to accurately position the raw waveforms since this PC underwent refraction correction before delivery. The effects of refraction were not corrected in the raw waveform files.

A **ground-truth data** acquisition campaign took place **in 2021** to gather knowledge on the land and sea covers in this area, through **uninhabited airborne vehicle** (UAV) and **uninhabited surface vehicle** (USV) **RGB imagery**, and geolocalized **photoquadrats**. They helped label the lidar data to perform semantic information extraction.

The UAV imagery was acquired over five smaller areas of interest, each representing typical coastal habitats, in March and April 2021 using an RGB DJI Phantom 4 Pro V2, and a Parrot Sequoia+ including a NIR nadiral sensor (770 nm to 810 nm) with a zenithal irradiance sensor. These flights were calibrated with a total of 55 ground control points. 150 photoquadrats were captured with RGB cameras and georeferenced, to seize the ecological diversity of the study area. Over a seagrass meadow of the study site, a PowerVision unmanned surface vehicle (USV) was used to acquire underwater images in September 2021.

Finally, an **RGB orthoimage**[2] acquired **in 2014** over the whole area was also used to give extra information on the habitats present on site four years prior to the data acquisition.

Five orthoimages and digital surface models were derived from the UAV imagery using the photogrammetric reconstruction procedure on the Pix4D software. Two separate processes were performed for each acquisition: an RGB and a Green-Red-NIR reconstruction. The positions of the ground control points were integrated to improve the quality of the derived products. The photoquadrats used as ground truth were simply linked to their position after GNSS data post-processing.

2. The "Ortho-littorale v2": https://www.geoportail.gouv.fr/donnees/ortho-littorale-v2

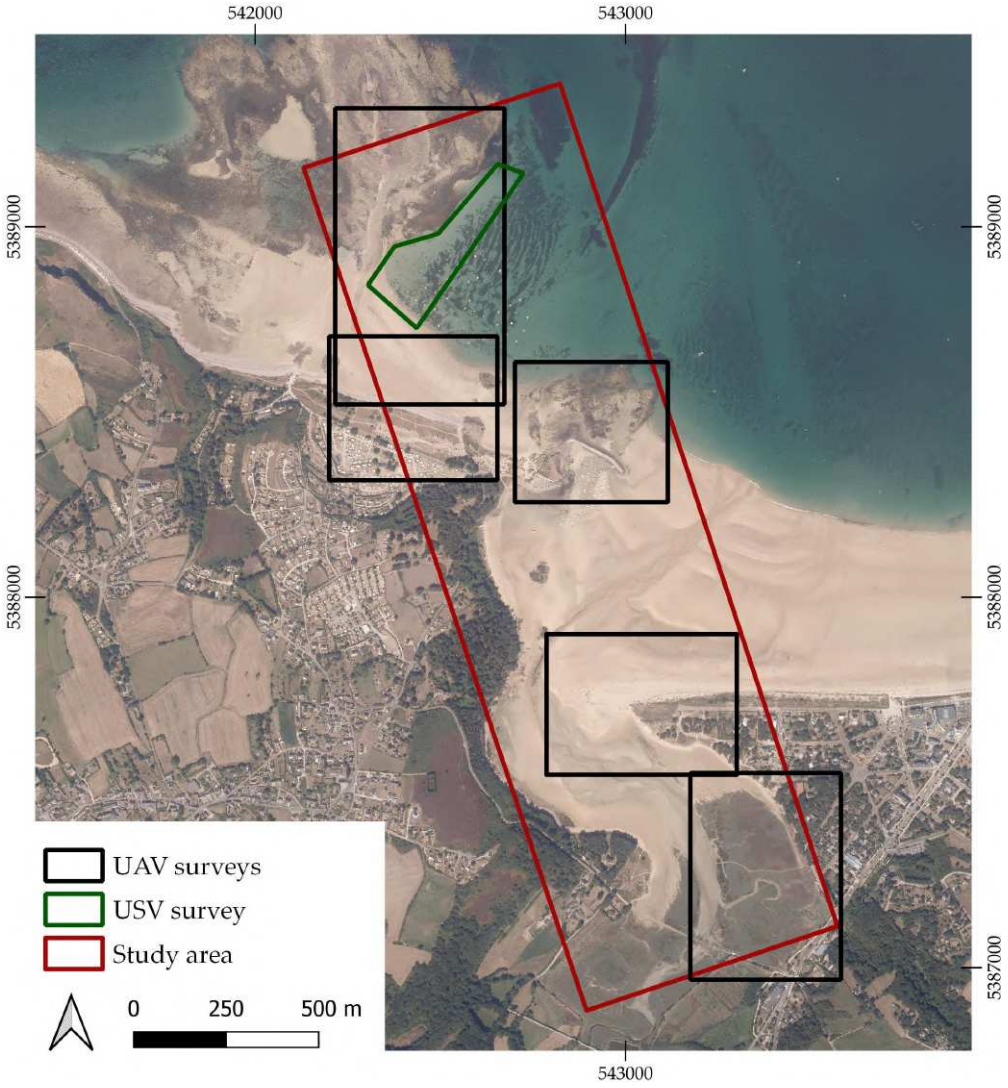


Figure 4.1: Study area and ground truth data spatial coverage (datum: WGS 84; projection: UTM30N).

Figure 4.2: Orthoimages resulting from the ground-truth data acquisition campaigns in Sables d'Or les Pins.

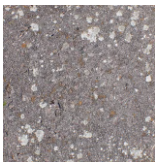
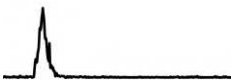
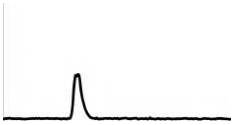
Class name	Illustration	Waveform	Class name	Illustration	Waveform
Algae			Seagrass		
Submerged rock			Submerged sand		
Rock			Pebble		
Wet sand			Dry sand/bare ground		
Artificial ground			Roof		
Boat			Car		
Salt marsh			Low vegetation *(lawn)*		
Intermediate vegetation *(shrub)*			Tree		

Table 4.1: Classes of land and sea covers identified in the study area and studied in the next sections. Waveforms were normalized between 0 and 1 over the scene for this illustration.

4.2 Classification of bispectral waveform features with random forests

To enhance land-water areas surface cover mapping, we first developed a 3D classification approach relying on **handcrafted bispectral** - green and NIR - **waveform features**. The resulting methodology contains three main steps:

- Green and NIR waveforms are first processed individually to derive handcrafted features;
- The resulting features are matched using a spatial criterion to produce a vector of bispectral information usable for classification;
- A random forest model is used to produce a classification of the feature vector, which is then rendered in 3D.

In the following sections, we detail each step and provide insights into the technical choices that resulted in this method. We then illustrate the performances obtained with it for the classification of diverse land- and sea-covers of the study area.

We evaluated our new workflow in different experiments with incremental complexity. In this chapter, we only provide a concise summary of the main observations we made, and present the most advanced results we got.

4.2.1 Methodology

With this methodology, our main contributions consist in:

- introducing bi-spectral waveform features-based coastal surface covers mapping;
- enhancing existing approaches using waveform features classification in coastal settings by avoiding data rasterization,
- illustrating the potential of TB waveform data to distinguish between terrestrial and benthic objects in complex natural environments.

Waveform features extraction

Extracting **tailored features** from the relevant part of the waveforms is the base of the approach. Since more than half of the registered lidar signal is noise, extracting waveform features first implies outlining the actual portion of information in these signals. Here, we wish to describe the surfaces and their

covers. In bathymetric waveforms, this corresponds to the **benthic return**, i.e. any return detected after the water surface component, that we discard from the information considered for feature extraction. In topographic cases, this consists of the complete waveform portion containing signal and not solely noise. Consequently, peaks corresponding to surface returns were isolated from noise in emerged domains, while peaks corresponding to benthic surfaces were isolated from noise and water surface or column returns in submerged domains. Figure 4.3 illustrates the portion of signal considered for feature extraction in bathymetric and topographic cases.

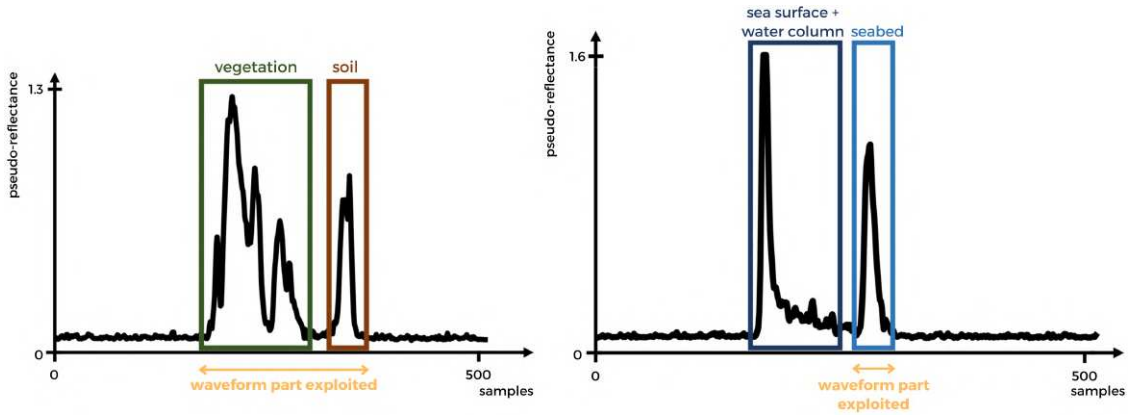


Figure 4.3: Illustration of the topographic (left) and bathymetric (right) waveform portions used for coastal surface cover classification.

Waveforms peak detection Since our approach aims at **identifying surface covers** in the waveforms, but not at separating them when they are multiple, isolating noise or irrelevant information - water surface and water column in our case - from surface signatures amounts to detecting the **start of the first relevant peak** in the waveform, and the **end of the last return**. We thus perform **peak detection** to identify the portions of the signals to describe with features and to classify.

To this end, we start by **smoothing** all waveforms with a Savitzky Golay filter, whose principle is to estimate piecewise polynomial functions. This step attenuates the noise and eases the identification of waveform peaks while limiting the number of false detections linked to abrupt sign changes that occur due to high-frequency noise.

Green and IR waveform peaks are not processed the same way since both wavelengths interact differently with water. In both cases, the **first derivative of the smoothed waveform** is computed and thresholded to make the increases in received energy stand out from the parts containing only noise.

Bathymetric waveforms segmentation Of both NIR and green wavelengths, only green penetrates water. Consequently, **bathymetric waveforms** can by definition only be green waveforms. However, green waveforms can also be acquired over land, and thus correspond to topographic domains. To make the distinction, we relied on a pre-classification identifying submerged areas made by LSS - Leica's Survey Software - before the data was delivered to us. In submerged areas, we monitored the number of peaks detected to isolate the seabed from the water surface and the water column.

In bathymetric cases, one of the main difficulties is to handle variations of several orders of magnitude of the waveform energy (Guenther et al., 2000). We thus implemented a **two-step thresholding** to balance over- and under-detections of seabed peaks. The first, higher threshold T_1 is well suited to the detection of the most significant peaks; however, depending on local conditions affecting the reflection of light, some bottom returns may be less intense and hard to expose. Thus, a second, lower threshold T_2 is used when only one peak is identified with T_1, to try to detect peaks after the water surface (i.e., the peak already detected with T_1). This lower threshold would exacerbate noise if it were used on the whole waveform, but it is adapted to the detection of weaker returns when used on the underwater part of the waveform (i.e. only the portion following a first, major peak detected with T_1). If no additional return is identified with this thresholding, we conclude that no seabed was recorded/detected and discard the waveform, since there is no seabed return to compute features on.

Topographic waveforms segmentation In the case of **topographic waveforms** - i.e. NIR waveforms or green waveforms **acquired on land** -, the separation of noise and useful information is more straightforward, as we wish to analyse all the backscattered energy. The sole processing is thus to isolate signal from noise. To this end, we evaluate the mean level of noise observable at the beginning and at the end of the waveform - i.e. before and after any scene-originating reflection - and isolate the part of the waveform having an intensity above this noise level. In this context, the number of peaks present has no effect on the procedure we adopt: if there are several peaks, features are computed on the group of peaks. If there is only one peak - which occurs with bare ground or low vegetation -, features are computed on this peak.

Correction of light attenuation in water As explained in Chapter 1, green light is attenuated by water exponentially with depth depending on its turbidity. This could induce a bias in our classification. For example, benthic surfaces with high reflectances located in deeper waters could be falsely mistaken for seabeds characterised by low reflectances, only because their backscatter has travelled a longer path in water and has thus been more severely attenuated. To limit such classification biases, we estimate the attenuation of the signal and reverse it to **correct the exponential decay** of bathymetric lidar waveforms.

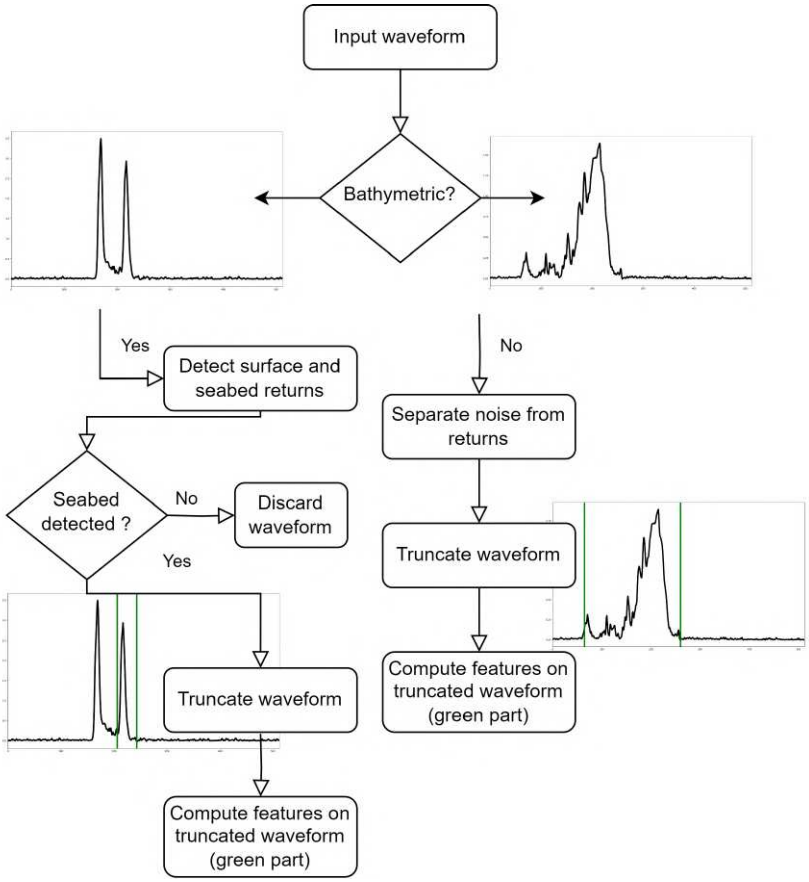


Figure 4.4: Waveform processing method flowchart and illustration on two different waveforms: one acquired over the sea, the other over land.

This attenuation coefficient, which is close to K_d because we use data acquired with large FOV sensors, is estimated by **fitting a decreasing exponential** function to the water column component of the signal, which allows the evaluation of the intensity gradient with depth. This component is delimitated using the peak detection made to locate the seabed, which necessarily involves the detection of a surface peak.

However, there are mathematical limitations to this approach: in very shallow water areas, no correction is applied since the exponential fitting is impossible on less than two waveform samples, which corresponds to depths under 0.125 m. In places where depths are smaller than 0.125 m - and over land -, the attenuation coefficient is thus fixed at 0.

Figure 4.5 illustrates the effects of attenuation correction on two types of bathymetric waveforms.

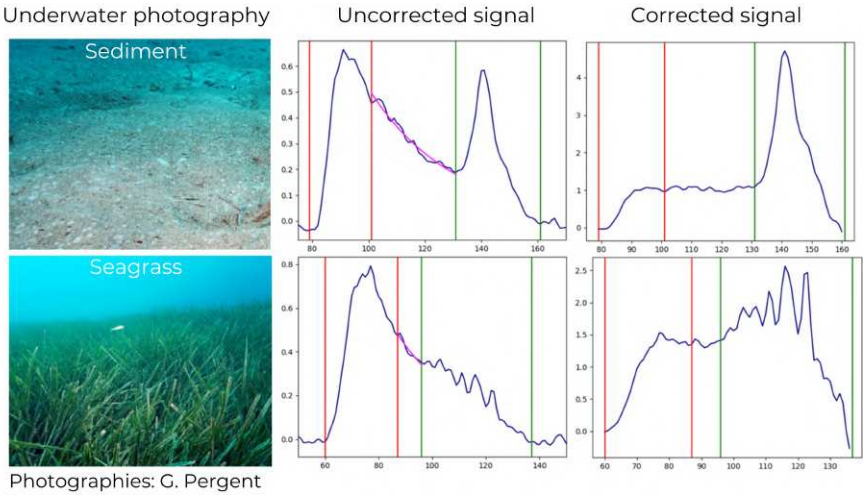


Figure 4.5: Illustration of the peak detection and signal attenuation correction obtained on two different bathymetric waveforms. Red and green vertical lines illustrate the resulting delineation of the air/water interface and water bottom returns, respectively.

Parameters extraction Before feature extraction, all truncated waveforms are divided by the peak energy of their incident pulse - i.e. the laser pulse they reflected. This allows us to obtain **pseudo-reflectances** and limit potential bias induced by variations of emitted laser intensity. Table 4.2 presents the **features** computed for each waveform after delineation of the useful signal and conversion to pseudo-reflectances.

Precisions about the NIR intensities To simulate a case in which only one wavelength is equipped with full-waveform recording - which is frequently the case - we made experiments with **green waveforms only**, in which NIR data were incorporated through the PC's intensity attribute. Intensities of

Name	Definition
Diffuse attenuation coefficient estimated value	Value of the coefficient of attenuation of light in water (0 for depths below 0.125 m and on land)
Complexity of the peak	Number of sign changes of the peak's first derivative
Mean	Mean pseudo-reflectance of the peak (after attenuation correction)
Median	Median pseudo-reflectance of the peak (after attenuation correction)
Maximum	Maximum pseudo-reflectance of the peak (after attenuation correction)
Standard deviation	Standard deviation of the pseudo-reflectance of the peak (after attenuation correction)
Variance	Variance of the pseudo-reflectance of the peak (after attenuation correction)
Skewness	Skewness of the peak (after attenuation correction)
Kurtosis	Kurtosis of the peak (after attenuation correction)
Area under curve	Area under the curve formed by the peak (after attenuation correction)
Amplitude	Amplitude of the pseudo-reflectance of the peak (after attenuation correction)
Time range	Time duration of the peak (in number of samples)
Total	Sum of pseudo-reflectance values forming the peak (after attenuation correction)
Height	Difference of altitude between the peak of the first layer of cover and the last peak
Maximum before correction	Maximum pseudo-reflectance of the peak (without attenuation correction)
Position of the maximum in the peak	Position of the maximum in the peak (in sample indices)
DZ	Difference between the elevation of the IR return and the green return

Table 4.2: Name and definition of the features extracted from the green waveforms during processing and used as input variables to the random forest model.

the NIR PC were matched with each point of the green waveforms PC using the **median NIR intensity** of the 10 nearest neighbours of each green waveform. To this end, we used the PC processing software CloudCompare (Girardeau-Montaut, 2022), in which the neighbourhood search is made with 3D Euclidean distances. The NIR PC was cleaned manually beforehand, to ensure all noise points, significantly above the surface, were removed from the data. The median intensity of the 10 closest neighbours in the NIR PC was chosen for two reasons. First, the number of 10 neighbours was relevant considering the difference between the two lasers' spot sizes and the resulting density of the PCs. Second, the use of the median intensity was more suited to the task than the mean intensity to avoid outliers' artefacts in spheres located at the interface of two habitats.

Precisions about the considered elevations For topographic waveforms, the elevation used corresponds to the last return's altitude (extracted from the PC). For bathymetric waveforms, it was computed using the depth of the last return identified by our algorithm and the altitude of our detected surface return, positioned with the PC. The vertical reference used is the French IGN 1969 system.

Classification

The features are directly classified to produce a **3D habitat map** so as to avoid information loss linked to rasterization.

We chose to use **random forest** models for their performance on multi-class problems implying dozens of features and their robustness to overfitting. The possibility to retrieve feature importance and prediction probability made them particularly suited to our needs. This type of algorithm has also been tested multiple times in 3D PCs classification research, with consistent observations of high accuracy in land cover identification (Yan et al., 2015; Chehata et al., 2009).

Our random forest classification model contains 150 trees and relies on classical parameters: nodes expand until the tree leaves are pure, and splits are made using Gini impurity, calculated using the following formula:

$$GiniIndex = 1 - \sum_j p_j^2,$$

where p_j is the proportion of samples of class j in the node. This criterion is close to 0 when the split is optimal, here when the nodes are pure.

Considering that we wish to apply the obtained model to a dataset containing millions of items, we choose a high number of trees, knowing that more trees theoretically equal better classification accuracy and that the number of trees needs to be adapted to the complexity of the dataset. We also based our choice on the observation made in Oshiro et al., 2012 on several different datasets that state that past 128

trees in the forest, classification accuracy gains become negligible for each additional tree, compared to computational demands.

We control the model's overfitting by monitoring the generalization score obtained on out-of-bag samples at each fitting step. The random forest implementation of the Python library scikit-learn was used to derive the results presented in Section 4.2.2 below.

Feature selection

To avoid potential negative feedback on the classification accuracy due to information redundancy among the N predictors, we conduct a **column drop-based feature importance** analysis and use it to select the most relevant classification attributes. To this end, we compute, for each predictor, the difference of accuracy obtained when removing it, compared to the performance of the complete feature set. The predictors that contribute negatively to the classification accuracy are not used to derive the final classification.

Production of a 3D map of land and sea covers

To visualize our classification results as PCs, we associate the predictions obtained with the planar coordinates of each waveform's last return and the elevation computed above. These coordinates are derived using the original green PC, which is corrected from the effects of laser refraction in water. This allows us to obtain the result under the form of a PC, and has the advantage of preserving the spatial density resolution, while also avoiding the critical issue of mixed pixels (Pi-Fuei Hsieh et al., 2001), often associated with rasterization.

Evaluation metrics

We assess the classifications obtained using the following metrics:

- **overall accuracy** (OA, ratio of correct predictions, best when its value is 1);
- **precision** (fraction of correct predictions among each ground truth class, best when its value is 1);
- **recall** (fraction of correct estimation for each predicted class, best when its value is 1);
- **F1-score** (harmonic mean of precision and recall, best when its value is 1).

As precision, recall, and F1-score as class-wise metrics, we use their unweighted mean to asses the results globally. A class-wise analysis is also performed, using confusion matrixes.

4.2.2 Results

The resulting workflow was successfully used for three classification tasks relying on lidar waveforms:

- Classification of *Posidonia Oceanica* seagrass meadows in Corsica, using bathymetric lidar waveform features and a 2D Maximum Likelihood classifier;
- Classification of 21 different land- and sea-covers in Sables d'Or les Pins using green waveform features and NIR intensities and a random forest model;
- Classification of 17 different land- and sea-covers in Sables d'Or les Pins using both NIR and green waveforms and a random forest model.

These three experiments were published in conference proceedings (Letard et al., 2021b; Letard et al., 2022b) and in a journal article (Letard et al., 2022a). Each experiment was the occasion to refine the method and explore its strengths and weaknesses. Since they were used to improve the design and implementation of the feature extraction method mostly, the different tests enumerated above were made with distinct settings - different training and test datasets, and different predictors and classifiers. In this section, we present a summary of the main observations we made, but all details and results about the three analyses can be found in the corresponding publications (Letard et al., 2021b; Letard et al., 2022b; Letard et al., 2022a).

Globally, all experiments show that full-waveform lidar is a powerful tool to classify TB areas. The obtained classification accuracies were consistently higher than 80% for the different tasks, and the visual results showed high precision and detail in the classified products.

In this section, we present more in-depth the results of the most advanced experiment, in which we used NIR and green waveform features to identify the set of 17 classes defined in Table 4.1 in the study area. A set of 1000 samples of each class was used to train the model. Another set of 500 distinct samples of each class was then used to assess the quality of the model's predictions. These samples form the training and test datasets of 17000 and 8500 feature sets respectively, that are shown in Figure 4.6.

Performances obtained depending on the predictors used

Classification experiments were led on five different sets of predictors in order to evaluate their relevance and added value. These five sets are the following:

- all green waveform features (11 features)
- all NIR waveform features (11 features)

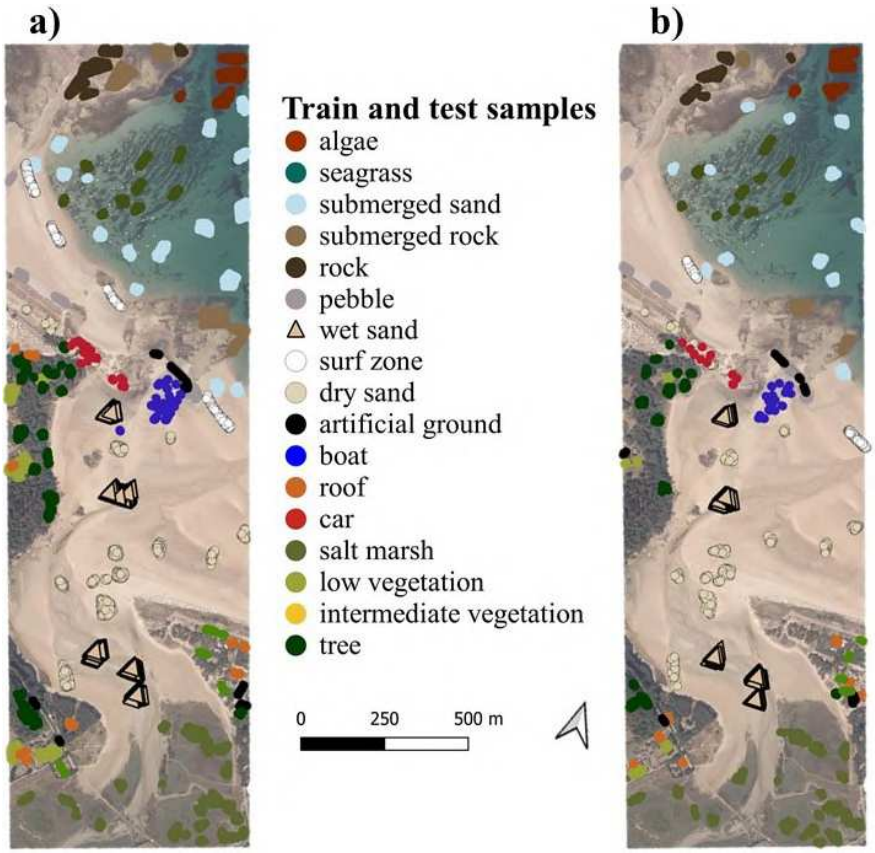


Figure 4.6: Distribution of the distinct a) train and b) test samples across the studied area (the natural colored imagery was acquired four years prior to the lidar survey).

Feature Set	Overall Accuracy	F-Score
Green waveform features	0.823	0.821
NIR waveform features	0.315	0.292
DZ	0.216	0.216
Green + NIR waveform features	0.846	0.842
Green + NIR waveform features + DZ	0.848	0.843

Table 4.3: Performance metrics obtained by different feature sets for the classification of 17 coastal surface covers.

- DZ (one feature only)
- green and NIR waveform features (22 features)
- all waveform features, plus DZ (23 features)

The performances of the classifications obtained for these sets are presented in Table 4.3.

Overall, the most relevant predictors for the classification of coastal land and sea covers are descriptors of the green lidar waveforms. IR data or differential elevation values appear to misclassify more than two-thirds of the points. The combination of green and NIR waveform features with elevation predictors produces the best classification performances.

Similar trends were observed in other experiments made with green waveform features, NIR intensities and elevations. Although the exact figures, detailed in the corresponding publication (Letard et al., 2022b) cannot be compared directly because the experimental settings are different, this shows that overall, there is a performance trend for coastal environments, in which NIR features < Green features < NIR + Green features < NIR + Green features + Z features.

Performance differences between NIR and green waveform features

Since each wavelength of the TB lidar was designed to survey a specific type of environment (NIR laser for topography and green laser for bathymetry), we performed an in-depth analysis of the classification results obtained using successively green data only and green + NIR data. This analysis features the mean precision and recall obtained for each class. It allowed us to observe which type of habitat was best described by each wavelength. The mean prediction confidence obtained for each point is also taken into account, in order to better understand the potential classification errors and what they imply. This **prediction confidence** corresponds to the probability that each point really belongs to the class it was assigned. The

complete class-wise analysis conducted for the green model and the Green + IR + DZ model is presented in Figure 4.7.

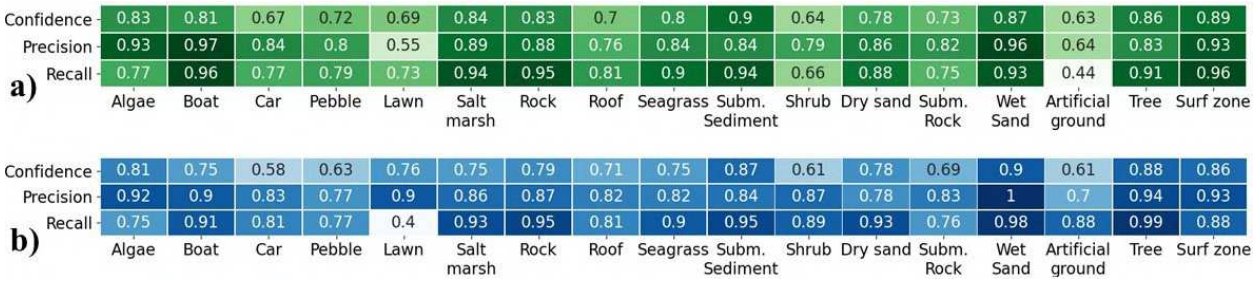


Figure 4.7: Class-wise classification metrics obtained when using a) green and b) bispectral waveform features as predictors.

The use of green waveform parameters produced accurate labelling, though the algorithm showed weaker performances on topographic classes such as lawn and artificial ground. The combination of both wavelengths and DZ produces a more accurate result and improves the recall of every class except low vegetation (*lawn*). Class-wise recall and precision values reveal that some classes were overestimated at the expense of others. This is the case of *submerged rock*, *algae* and *lawn*, which have lower recall values than precision.

Permutation-based feature selection

To generate the final coastal habitat classification, we analysed the contribution of each feature to the overall accuracy, and excluded from the predictors the features that impacted it negatively:

- green waveform skewness,
- IR waveform AUC,
- IR waveform skewness,
- IR waveform maximum,
- IR waveforms' mean.

Bispectral dataset classification

The metrics and the map obtained with the final set of features are presented in Table 4.4 and Figure 4.8.

Selecting attributes based on their importance makes the overall accuracy reach 86%. Globally, the classifier's tendency to overestimate *submerged rock* or *algae* identified with the low recall in Figure 4.7

OA	Precision	Recall	F-score
0.856	0.862	0.856	0.852

Table 4.4: Performance metrics obtained after bi-spectral waveform feature selection.

is still observable in the final result, illustrated in Figure 4.8, where these classes respectively invade the seagrass meadow and the surf zone.

Other obvious confusions exist between *rock* and *dry sand*, or *submerged rock* and *surf zone* or *submerged sediment*. They are also revealed by the precision and recall values of these classes in Table 4.7. One of the main confusions is between *pebble* and *sand*, yet it is not as clearly quantified by the metrics.

Prediction confidence analysis

To further assess the abilities of our method to classify land and sea covers, we analysed the prediction confidence obtained across the studied area for the bispectral waveform features classification. Figure 4.9 shows the maps obtained when setting a **confidence threshold**, below which the points are labelled as unclassified.

Most points are kept with a threshold set at 70% (which means the probability that the point belongs to the class it was given is at least 70%). This is in line with the mean confidence of 77% obtained on the test dataset (see Figure 4.7). However, when the threshold is increased at 90%, more complex areas, mainly at the interface between different classes, disappear, as they are classified with a lower confidence.

A closer look at the **misclassified samples** shows that the confidence level is globally lower for them. Indeed, on the test dataset, samples that were wrongfully classified have a median confidence of 47% with a standard deviation of 17%. The overall accuracy values obtained when filtering the points based on their confidence predictions confirm that misclassified points can be discarded using this criterion: Table 4.5 presents the accuracies obtained for different confidence thresholds.

Additional observations on the green waveform features made during side experiments

Though the exact OA figures cannot be compared directly because the exact training and test points were not identical, global performance trends depending on the type of predictors can be analysed.

Similarly to what we observed in Chapter 3 with PC features, the features failing at the selection step vary depending on the application case. Indeed, in the two other experiments made with this method, different **feature importance dynamics** were observed.

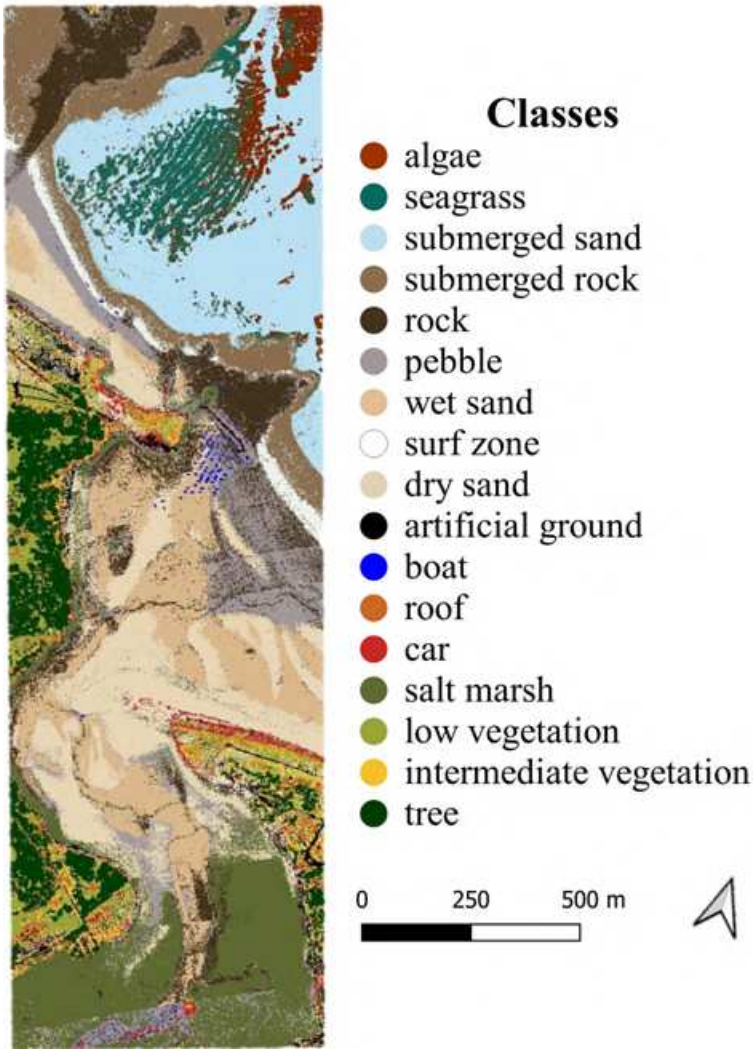


Figure 4.8: Coastal surface covers map obtained when classifying a selected set of bispectral waveform features.

Threshold	0.6	0.7	0.8	0.9
OA	0.95	0.97	0.98	0.99
Remaining points (%)	65%	55%	44%	32%

Table 4.5: Overall accuracy of the resulting classification depending on the prediction confidence threshold.

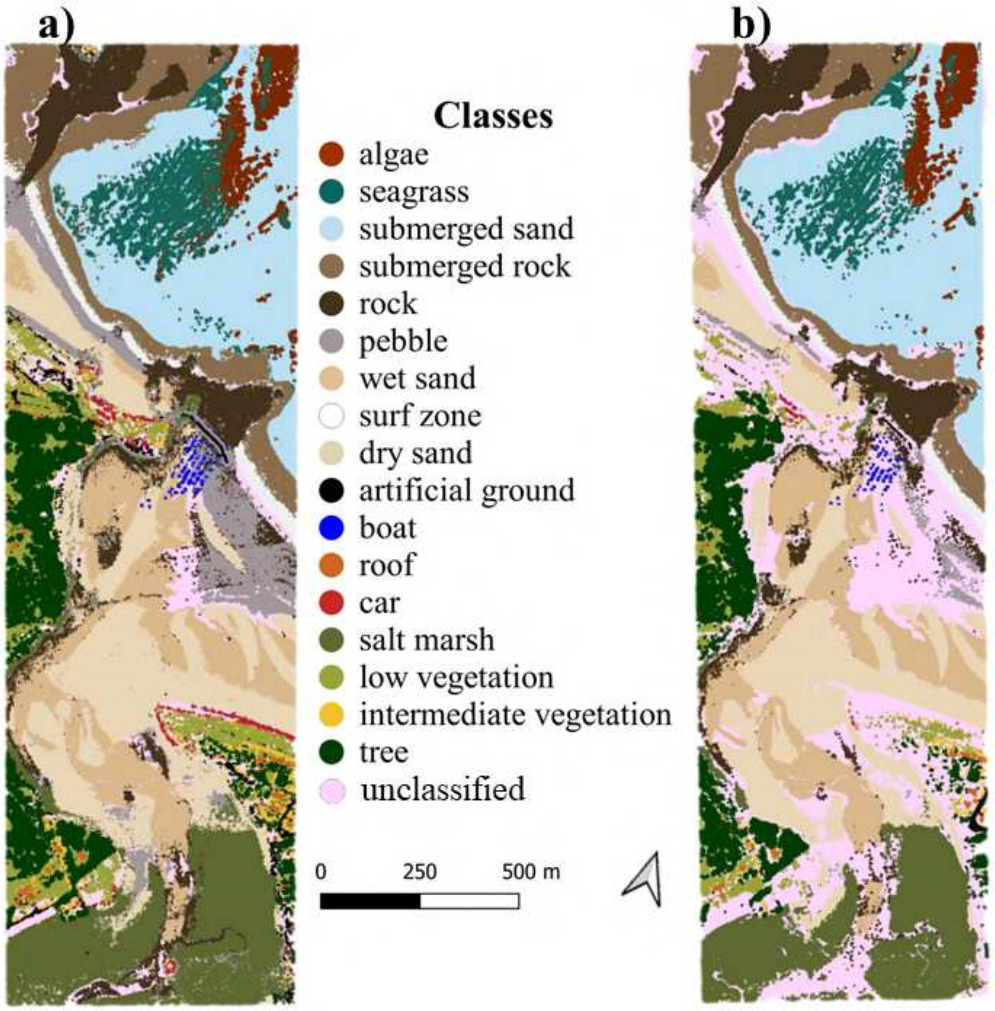


Figure 4.9: Land and sea covers map obtained at a) a 70% confidence level and b) a 90% confidence level.

When experimenting with a completely different dataset acquired in Corsica and containing seagrass green waveforms and sandy seabed green waveforms, the following features lowered the classification accuracy:

- Mean of the green peak after attenuation correction
- Maximum of the green peak after attenuation correction
- Standard deviation of the green peak after attenuation correction
- Green peak kurtosis
- Green peak AUC after attenuation correction
- Amplitude of the green peak after attenuation correction
- Total energy of the green peak after attenuation correction

Similarly, experiments on the Sables d'Or les Pins dataset including only green waveform features resulted in the following features eviction:

- Mean of the green peak after attenuation correction
- Maximum of the green peak after attenuation correction
- Standard deviation of the green peak after attenuation correction
- Variance of the green peak after attenuation correction
- Skewness of the green peak after attenuation correction
- Green peak AUC after attenuation correction
- Amplitude of the green peak after attenuation correction

A more detailed analysis of the contributions of the different types of green waveform features in this same experiment revealed that the **most informative green waveform features** were those describing **peak shape** - complexity, skewness, kurtosis, area under curve, time range, and height of the green waveform - and those linked to the **lidar acquisition** - diffuse attenuation coefficient estimated value, maximum, maximum before attenuation correction and position of the maximum in the peak of the green waveform.

4.2.3 Discussion

In the following paragraphs, we discuss the results obtained with our bi-spectral waveform classification method based on handcrafted features extraction.

Usability of full-waveform lidar for coastal habitat mapping

The final result obtained confirms the observations of (Mallet et al., 2011) and the observations we made in our two other experiments: they illustrate the potential of lidar waveforms for classification tasks. Here, a single dataset made the classification of 17 different land and sea covers possible with high accuracy (86%, see Table 4.3). The resulting 3D map is presented in Figure 4.10. It is dense and has a high spatial resolution, suited to the realisation of ecological assessments such as ecosystem services evaluation, as performed in Martínez et al., 2007 and Costanza et al., 1997.

Considering the thematic objectives of this research, the results are promising, since all ecosystems that provide goods and services are described with relatively high precision (on average, 87%, see Figure 4.7) and in 3D, contrary to other methods developed in existing papers (Collin et al., 2012).

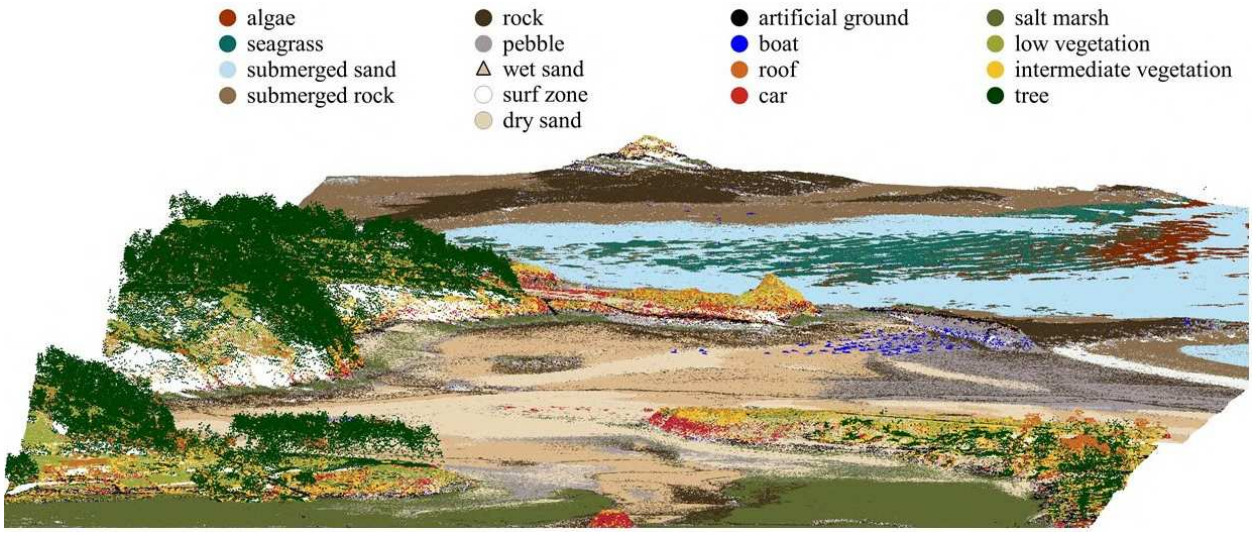


Figure 4.10: Resulting 3D PC of a coastal area and its habitats classified with full-waveform topo-bathymetric lidar.

One of the main objectives of this study was to develop a method for seamless spatial modelling of marine and terrestrial habitats. Here, the land-water continuum is classified without interruption, which is a key methodological aspect of coastal habitat monitoring (Collin et al., 2012). Figure 4.10 shows this uninterrupted restitution of the land-water interface. However, our approach has limitations that necessitate further investigation: the classifier has a tendency to overestimate algae at the expense of seagrass,

and pebble at the expense of sand, for example, as observed in our other experiment using only green waveforms (Letard et al., 2022b). Besides the quality of the training and test datasets established, a source of explanation for these classification errors could be found in the technical specifications of the sensor. The diameter of the HawkEye III's green laser's footprint is 1.80 m, which means that the returned waveform condensates information in a 2.5 m2 area. This parameter may have an influence on the ability of a given array of features to describe pebbles or sand, mostly at interfaces between classes.

Even though these classes are close semantically and can have similar waveform signatures due to their size, physical properties, and texture relative to the laser spot size, a better distinction is needed for ecological applications. Boats and cars are also often falsely detected. Figures 4.8 and 4.10 illustrate well this issue, as a great number of points are classified as cars on land, and part of the boulders of the dyke are labelled as boats. Figure 4.11 focuses on the sandy dune, which features many false detections of cars.

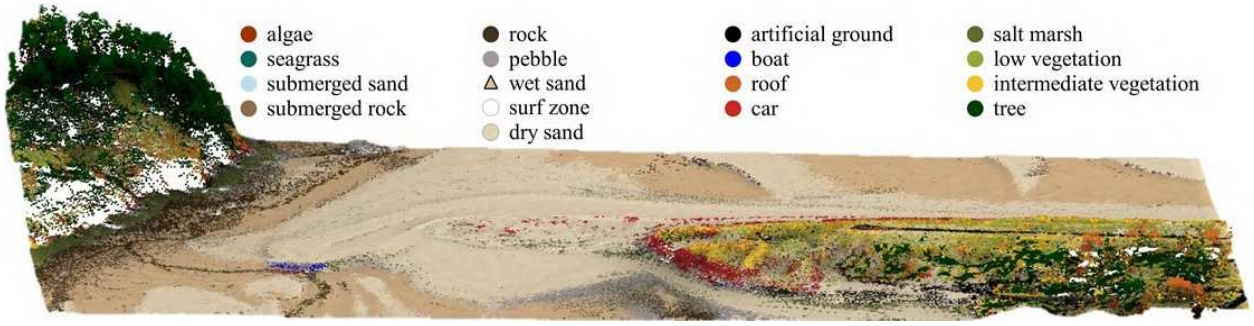


Figure 4.11: Extract of the resulting 3D classification: the sandy dune of Sables d'Or les Pins and its surroundings.

These errors may be explained by the albedo of cars and boats, which can vary considerably between two different types of vehicles. The wide colour spectrum they can have is difficult to model in a training dataset which can lead the classifier to learn confused information. The procedure to adopt for these types of classes in order not to compromise the wider objective - ecological monitoring - should be further discussed. These classes could be merged in a more global vehicles class, or an unclassified class could be added to handle unusual feature vectors and avoid their detection at the expense of natural habitats mapping, as in Figure 4.11 where the sandy dune - a key ecosystem - is mapped as car. Despite these mistakes, the possibility of having 3D assessments of the spatial repartition of different ecosystems already bears encouraging perspectives for coastal ecology and could enhance results outlined by previous studies using rasterized TB lidar data (Wedding et al., 2008; Collin et al., 2012).

The highly informative content of lidar waveforms, already stated in existing research on the topic (Mallet et al., 2011; Collin et al., 2012) is illustrated by their constant contribution to the resulting classification accuracy (Table 4.3). Lidar waveforms contain enough information to describe surface covers,

despite their lack of information on their neighbourhood's geometry and spatial repartition (Table 4.3). This loss of spatial context information - compared to PC classification techniques involving neighbourhoods (Brodu et al., 2012) - is also one of the strengths of waveform-based processing. It avoids spatial averaging of information that can result in classification artefacts depending on the neighbourhood radius defined (Brodu et al., 2012). Classifications based on waveforms and not on spatial context may consequently gain in horizontal resolution, keeping in mind the influence of the laser's footprint diameter. However, both wavelengths do not perform equally, as expected since they were each specifically designed for different environments (Philpot, 2019). Infrared waveforms cannot be used alone to study both dry and wet environments: the metrics obtained for the IR model (Table 4.3) quantify the limitations of topographic lidar for the survey of highly diverse environments. Green lidar waveform features perform better: their classification reaches 82% of overall accuracy, and similar values of precision and recall. Their ability to label some classes of ground is limited, which is why dual-wavelength datasets are relevant.

Comparison to existing waveform-based methods

Our results showed that TB lidar is tailored for the classification of coastal habitats. Elevations and green waveform features were complimentary and achieved high-precision results when combined. To the best of our knowledge, no similar papers proposing point-based land and water cover mapping from bispectral lidar data were published, so no direct comparisons of results are possible. However, our observations corroborate those made in Chehata et al., 2009, which successfully used random forest algorithms to classify full-waveform lidar data over urban areas and obtained an overall accuracy of over 94% when identifying four types of land covers. Chehata et al., 2009 only focuses on terrestrial areas but confirms the high accuracy we observe when using waveform features without rasterization for mapping purposes. Class-wise, our results seem more homogeneous for the land covers we have in common, although this means that our approach performs less accurately than theirs on some urban classes. Indeed, Chehata et al., 2009 presents a recall of 94.8% for buildings, which is higher than what we obtain on our roof class (81%), but our vegetation classes (trees and shrubs) have an average recall of 94%, while theirs is 68.9%, and our natural ground classes (soil, lawn, salt marsh) reach an average recall of 82.6%, higher than the 32.7% presented in Chehata et al., 2009. However, the method introduced in Chehata et al., 2009 performs better than ours on artificial ground, with recalls of 96% and 88%, respectively.

Although we found no other research performing point-based classification of subtidal, intertidal, and supra-tidal habitats, we can compare our findings to those in Tulldahl et al., 2012, where the authors also observed that the use of waveform data improves seabed maps and obtained an OA of 86% for their classification of seabed substrates and aquatic macrovegetation. Their approach provides a better mapping of low underwater vegetation on soft substrate (100% versus 82.5% of PA in our case) but a

less accurate detection of hard seabed substrate (68% versus 76% of PA in our study). Again, although we have less accurate results for some classes, our method seems to provide more balanced and homogeneous performances among different classes.

Our results also corroborate those from Collin et al., 2012, where 19 land-water continuum habitats were classified with an OA of 90%, and the authors concluded that the best classification results were obtained when combining spectral information and elevation. However, Collin et al., 2012 used digital models of waveform features that they obtained by rasterizing their data, and they relied on a maximum likelihood classifier. Although our metrics are similar when using green waveforms only, our classification has the advantage of preserving the spatial density and repartition of the data.

Other studies such as Butler et al., 2020, Grande et al., 2009, and Chust et al., 2010 used 2D lidar-derived data and imagery along with machine learning classifiers to map similar coastal habitats as the ones we attempted to map. They obtained performance scores in the same range as ours, with OAs between 84% and 92%. The authors did not use waveform data in these studies and observed low accuracy when classifying only digital elevation models obtained with lidar surveys, therefore requiring the additional processing of imagery. Our approach has the advantage of requiring only one source of data out of the two sources often used in existing literature, which facilitates both acquisition procedures and processing. Globally, our results are in line with Collin et al., 2012, Chehata et al., 2009, and Neuenschwander, 2009, which all state that bathymetric lidar waveforms are well-suited for benthic habitats mapping and observe the same complementarity between spectral and elevation information for habitat mapping. Our method offers an OA similar to existing research in lidar data classification for habitat mapping, while extending the application to a wider range of habitats—both marine and terrestrial—and avoiding information loss through rasterization. Although the recall obtained for some classes is lower than results previously presented in other studies (Tulldahl et al., 2012; Chehata et al., 2009), this method also has the advantage of offering homogeneous performances and low inter-classes recall differences, contrary to these aforementioned existing research results.

The random forest models trained showed low overfitting, as the extended application results illustrated. The classification of boats located outside of the training and test data collection area, for example, illustrated that the classifiers obtained could be applied to other datasets accurately. Natural, semi-natural and anthropic habitats were well distinguished, and vegetation was precisely isolated, which opened perspectives for ecological assessments of those coastal areas.

Feature extraction method design

We defined 16 features to extract from the portions of the waveforms that correspond to layers of ground or seabed covers. These were efficiently retrieved both on land and underwater. To perform feature ex-

traction, we relied on a previous land-water distinction made by the sensor's constructor software before the data was delivered to us. This is a rather easy step that is well covered in the literature and often performed in pre-processing to generate PCs since the effects of light refraction in water need to be corrected to derive PCs with correct positions. We consequently preferred building on that already strong aspect of lidar processing and focusing on distinguishing more detailed classes below and above the waterline, rather than focusing on the waterline itself.

However, our approach did not handle extremely shallow waters, where the surface component and the bottom return overlap in the waveforms. In these cases, the peak detection employed did not distinguish the seabed from the water surface and no features were retrieved. There was consequently a 24 m wide band without bathymetric waveform features in our bispectral waveform features dataset. We also noticed cases of confusion between seabed return and noise in the water column component of the waveform, which resulted in a mis-located detected seabed. These issues could be handled by improving the way the different waveform components are isolated: using waveform decomposition (Shen et al., 2017) or deconvolution (Zhou et al., 2017; Wu et al., 2011) could produce better results on that aspect.

Another aspect of our feature extraction method that could be improved is the correction of the attenuation of the signal in water that we perform on bathymetric waveforms. This exponential correction produces extremely high values of backscattered intensities underwater, which make little sense physically. On the other hand, topographic waveforms are not corrected: their typical intensity order of magnitude is thus several times smaller, which might disrupt the classifier when building some tree nodes. The fact that the permutation-based feature importance assessment revealed that the uncorrected intensities were among the highest-ranked predictors and that several parameters expressing statistical variations of corrected intensities were discarded suggests that this correction hinders the classification process.

Contribution of the bispectral information

The combination of infrared and green datasets produces the most accurate classification overall in both our experiments with bi-spectral data (Letard et al., 2022b; Letard et al., 2022a). The addition of the infrared features to the green parameters improves the distinction of the different types of grounds, and results in an increase of 3% of the overall accuracy, precision and recall.

The increase in accuracy obtained when adding NIR information was even higher in our first experiment on this study area, where we distinguished 21 different classes. In both cases, an explanation of the shortcomings of the green waveforms on topographic classes can be found in the sensor specifications and the data. Indeed, the HawkEye III was particularly designed for bathymetry extraction, consequently, its green lasers are set to be powerful enough to reach the seabed up to several dozens of meters in coastal waters. Over land, this laser power is so high that most of the waveforms originating from highly reflec-

tive surfaces are saturated. The green wavelength alone might consequently not capture fine variations of backscattered intensities over land to allow the separation of similar environments such as plane habitats, different types of herbaceous vegetation, etc. The shapes of the saturated waveform returns are also affected: the shape of a saturated peak around its maximum is not assessable. This can explain why there was a lot of confusion between topographic habitats when using green waveforms only to study more detailed classes - e.g. separating tar from concrete, or differentiating between low, medium and high salt marsh vegetation.

Additionally, the infrared laser models a more concise and precise surface, while the larger size of the green laser may sometimes result in the mixing of different land covers into one single return, but is able to penetrate through water (Philpot, 2019). We suggest that the combination of two types of spot sizes and wavelengths optimises the information collected on a given area in terms of albedo, water content and surface rugosity, which all impact the waveform and characterise natural surfaces. Having two different lasers also provides a good understanding of the vertical complexity of the scene. Surfaces are sampled with varying sizes of laser beams thanks to the dual-laser system. A wider footprint hits a wider portion of the surface at a time: it can mix information about several layers of covers and create intermediate points between the canopy and the ground. This phenomenon is documented in all laser scanning systems (Brodu et al., 2012). The smaller footprint of the infrared laser may not always penetrate through dense covers but creates fewer mixed points. The combination of both sources of data results in local differences in elevation and a greater vertical density over more complex surface covers and non-planar areas. Bispectral lidar thus gives a more thorough review of the vertical structure of the environment, which can explain the improvement of the classification accuracy (+0.2%) when adding DZ to the predictors. Using DZ is also a way of including spatial context data, which is particularly contributive in PCs classification (Brodu et al., 2012), to the model, without making it too dependent on the training area or involving neighbours. Finally, the few predictors discarded after the importance analysis - green waveforms' skewness, IR waveforms' AUC, skewness, maximum and mean - show that both wavelengths contain relevant information. They also confirm the theoretically more exhaustive nature of the green waveform. Indeed, infrared waveforms seem to contain fewer essential details on the surveyed scene, as most of the features dropped after this step concerned the characteristics of the infrared return, whereas nearly all descriptive parameters of the green returns were useful to the random forest model.

Contribution of the elevations

Our first experiments included elevations as predictors since most of the existing approaches for coast habitat classification with lidar data used DTMs. Elevations contributed greatly to the improvement of the classification OA but were largely outperformed by spectral predictors when used alone. This was

expected since many different habitats coexist at similar altitudes and are mainly differentiated by their reflectance. However, the classification accuracy of surface types whose elevation is an intrinsic quality - e.g. salt marsh - greatly benefitted from the use of elevation as a complement to waveform features.

Despite these observations, we decided not to keep the Z coordinate as a predictor as it greatly limits the generalization abilities of the resulting classifier, which is too specific to the absolute elevations of the area it was trained on. Considering we obtained high accuracies without this information in our later experiments, we decided that this feature was not essential. Furthermore, the elevation values extracted with our approach were not always consistent with those provided by the original PCs in marine areas. This can be explained by peak detection artefacts caused by high noise in the water column in very shallow areas namely. To remove artefacts due to water quality or sea foam, a post-processing step could be implemented, and the neighbouring elevations could be used to regularize the processed PC obtained.

Classification algorithm and prediction confidence

Overall, our observations corroborate existing research on the classification of lidar data using random forest algorithms (Yan et al., 2015; Hansen et al., 2021). In both our experiments with random forests, the model was quickly applied to a great number of features and points and shows a good ability to exploit information and detect between 17 and 21 classes. The average prediction probability for 17 classes is 77%. This criterion and the qualitative results in Figures 4.8 and 4.9 outside the test dataset show that the classifier is robust to overfitting and is generalisable to a wider scene.

The average prediction confidence being high, it can be used to filter the results and preserve the overall quality of the map even if it means compromising on the 3D density in some areas, as introduced by Brodu et al., 2012. Table 4.4 and Figure 4.9 show how points can be removed by applying a confidence criterion in order to prioritise solid predictions, and thus improve the overall accuracy, as in Brodu et al., 2012. The points affected by this filtering step give indications of the strengths and weaknesses of the random forest model. Misclassified points with high prediction probability are evidence of training errors. This is the case of the false detections of submerged rock along the surf zone, which Figure 4.12 shows more clearly.

In Figure 4.12, the false detections of submerged rock are located along the surf zone points. The submerged rocks detected next to the rocks in the foreground are true positives. These mislabelled points have a confidence value higher than 90%, as Figure 4.9b) reveals, which means that their descriptive features correspond to the usual range of statistics describing submerged rock. In this case, further investigation into the most distinctive feature of submerged rock could help identify the origin of the issue. A hypothesis could be that the similar DZ (i.e. water depth) of all the samples of submerged rock in the scene introduces a bias that causes the surf zone to be confused for submerged rock.

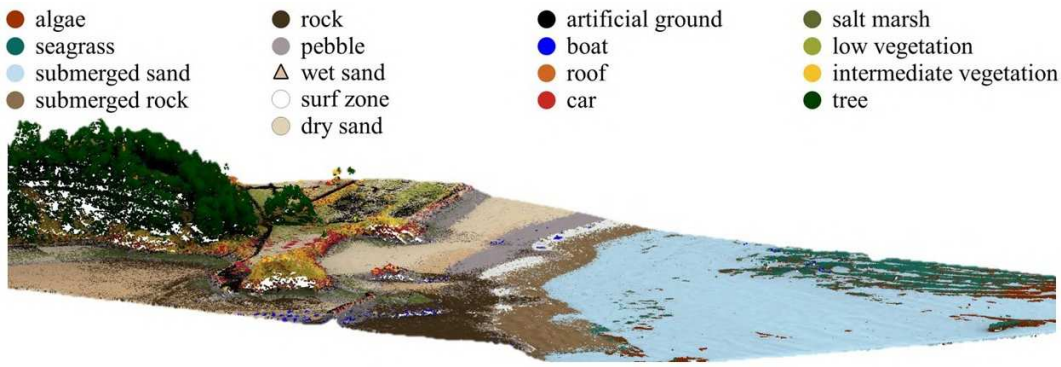


Figure 4.12: Extract of the resulting 3D classification: the land-water interface and its habitats.

The prediction confidence is also lower for areas at the interface between very distinct environments (Figure 4.9b), where the waveforms might integrate mixed information due to the size of the laser spots. The same issue was documented in Brodu et al., 2012 in 3D terrestrial laser scanning data classification. These areas were less documented in our training data, and the prediction probability underlines this lower confidence. On the other hand, errors made with low confidence reveal poorly represented ranges of values among the training dataset, that are difficult to place in the possible labels. They reveal less about the training process than about the quality of the dataset.

Brodu et al., 2012 identify that some areas of 3D PCs can have lower classification confidence due to a smaller point density compared to the rest of the scene. In their case, the classifier relies on geometrical features, and this difference in the dataset's constitution explains why their features are not as distinctive. In our case, outliers and erratic values of intensity or elevation in the data can be difficult to classify correctly, since they are not included in the training data, that are rigorously selected for their representativity of each class. In this respect, our results illustrate how important the quality of the lidar dataset is for classification tasks. Vertical or radiometric calibration issues can severely impact the possibility of detecting the nature of the surface, as we can see in Figure 4.8, and as suggested in Brodu et al., 2012. Flight lines can be recognized through the classification obtained around the salt marsh and on the beach between the boat moorings and the dune. In our dataset, these lines had saturated intensities difficult to interpret, and the elevations at the interface between flight lines differed by several centimetres.

The possibility to analyse the prediction confidence is thus a great indicator of the potential biases in the method. It also offers the possibility to improve the classification accuracy (Brodu et al., 2012): more than 10% of OA can be gained thanks to confidence-based filtering (see Table 4.4).

Further developments to allow the classification of each surface cover backscatter could improve the classification's resolution. Indeed, in this work, we did not consider multiple covers individually but

merged them as a group. Classifying each waveform peak independently could improve the vertical density of the map and further increase the quality of the results. An improved separation of the different echoes could also limit the difficulties encountered in very shallow waters. In the next section, we explore another approach to get closer to these objectives.

4.3 Semantic segmentation of lidar waveforms with U-Time

The method introduced in the previous section illustrated the limits of waveform classification techniques relying on pre-detection of the backscatters. In this Section, we process bi-spectral lidar data and propose an approach able to detect and classify waveform components without supervised peak detection and feature extraction. More precisely, we focus on the ability of U-time (Perslev et al., 2019), a neural network inspired by the well-known U-Net developed for image analysis, to classify waveform samples. In practice, we aim at:

- (1) identifying real peaks from noise and signal not corresponding to backscatter and
- (2) classifying them to discriminate coastal ecosystems efficiently and automatically detect the presence of water or vegetation.

We propose to use a neural network adapted to lidar waveform processing as a classification technique that allows the detection and characterization of various coastal ecosystems, and that correctly separates subtidal, intertidal and supratidal areas. This network is calibrated and validated using very high-resolution UAV optical imagery and lays the foundations for future integrated approaches of land-water continuum mapping.

When we first developed this approach, we were not able to extract the infrared waveforms from the survey data, due to a technical issue in the pre-data delivery processing. Consequently, in this study, only the green shallow channel full-waveforms are used. As we were unable to exploit the NIR waveforms, we used the NIR PC intensities instead.

Since then, we managed to retrieve the NIR waveforms as well, and used them in our first approach, as presented in Section 4.2. However, we decided not to incorporate them into our semantic segmentation workflow considering the results we obtained with green waveforms only, and preferred moving towards different approaches.

4.3.1 Data preprocessing

Green lidar waveforms labelling

In order to train our waveform processing algorithm and assess its performance, 1500 green waveforms per habitat were selected and their echoes were located and labelled.

Waveforms already contained **echo positions** obtained via the machine's constructor peak detection algorithm. These positions practically consist of the points present in the delivered PCs. We filtered these peak locations and used them as ground truth after labelling them depending on the nature of the ground in the area (green PC files and green waveform files were linked using the GPS time information). Peak locations corresponding to noise points were removed using the noise classification flag present in the PCs delivered by Leica's survey software.

We then selected areas of interest for each type of ground to discriminate using the UAV orthomosaics, the orthoimage from 2014 and the ground-truth data, and selected representative waveforms in these areas. The 2014 orthoimage was used as past ground-truth to confirm the ground type since there is a 1.5-year gap between the lidar acquisition and the UAV and photoquadrat ones. For each representative area, the echoes were assigned to the label corresponding to the type of cover observable both in 2014 and 2021.

For underwater objects, several echoes corresponding to different cover types were present. The distinction between the seabed echoes and the sea surface return was made using the flag corresponding to the sea surface present in the PC for refraction correction purposes. This detection is considered to be trustworthy as (1) the surface position is easily distinguishable in shallow water areas using the NIR PC and (2) this is the flag used by the French Marine hydrographic and oceanographic service to process their lidar PCs and produce their underwater maps. No very shallow water areas were selected to create the labelled dataset in order to avoid return location uncertainties.

Waveforms were individually visually checked before and after echo selection and labelling to make sure no outliers were noticeable in the dataset. In practice, **nine covers** were identified as well as a "water surface" class since it is relevant for underwater waveforms. Finally, a "non-useful" class was created for all discrete values not assigned to a specific object. All the echo classes are summed up in Table I.

In total, 13500 labels – 1500 per cover type – were generated, in the form of series of 512 values alternating between 0 (= non-useful) and the label of the relevant class at the position of an echo. For each cover type, **1000 waveforms** were dedicated to **training** and validating the classification and **500** were kept to **test** the trained algorithm. There are consequently two datasets: a 9000-item training dataset, and a 4500-item test dataset, one item consisting of a couple of waveform and label time series.

Green lidar waveforms preprocessing

In order to achieve more generalizable results, we converted the backscattered intensities into **pseudo-reflectance** values. Indeed, lidar intensities depend on many acquisition-related parameters and cannot be considered as an absolute measurement of a surface's albedo (Kashani et al., 2015). Since the laser impulse is registered for each waveform, we divided the backscattered signal by the maximum of the impulse to obtain pseudo-reflectances, which would probably be less dependent on our dataset. To be passed as input of U-time, all waveforms were normalised between 0 and 1, and only the 512 first samples were kept since the last samples of the waveforms usually only contain noise.

Infrared lidar data analysis

To construct our **bispectral dataset**, the **NIR backscattered intensity** was extracted from the NIR PC at each position where a green waveform was selected for the train and test datasets. To get this parameter, the NIR PC was first cleaned up by removing noisy points observable meters above the surface. The resulting PC was then gridded at a 5 m resolution, to obtain a digital intensity model. No interpolation was made in order to preserve the data, empty cells were consequently left empty. Finally, the value of the digital infrared intensity model was extracted for each position at which a green waveform was used. The 5 m resolution was chosen to account for the impossibility of finding a green point at the exact same position as an infrared point, due to the differences between both lasers and due to their non-cofocality. Although they could not be converted into pseudo-reflectance values, the infrared intensities were also normalized between 0 and 1 before being passed to the network.

4.3.2 Semantic segmentation of waveforms

Neural network architecture

Before entering into details, let us recall that the main idea behind deep neural networks is to model complex relations between inputs (waveforms here) and outputs (labels associated with each point) through the succession of a large number of elementary operations performed with neurons. The parameters associated with each neuron are learned during the training phase, aiming at finding the optimal parameters that minimize the loss function, i.e. the discrepancy between estimated labels and real ones, with ad-hoc optimization tools.

When spatial/temporal relations exist between input data, as with waveforms, convolutional networks are often used to capture multi-scale relationships between inputs. In image processing, efficient networks performing convolution/deconvolution of the data have enabled impressive results for image

understanding and classification. In this work, we rely on a **1-dimensional** (1D) version of a **convolutional/deconvolutional network**, named **U-time**, to process the waveforms. The chosen architecture is a network inspired by the U-Net neural network, characterized by its so-called **skip connections** linking the encoder (convolution) and the decoder (deconvolution) parts of the model.

To adapt the network to our time series, we only use 1D convolutions, contrary to the original architecture that was created to process 2-dimensional images. The **encoder** part consists of four convolution blocks of a fixed-size kernel, each encompassing two sequences of convolution, batch normalization and activation with a rectified linear activation unit (ReLU). The two last operations enable us to keep only the main interesting part of the information. Max pooling, i.e.reducing by two the size of the filtered waveforms, is performed at the end of each of these blocks. By doing so, the convolution of the next level (with the same size kernel) embeds large-scale information.

The **decoder** part consists of four sequences of upsampling, skip connection and convolution block. Both parts are linked by a convolution block. Our architecture consequently has a depth of four max-pooling layers, meaning that at the end of the encoding part, the waveform is synthesized into 32 samples. This seems adapted to the problem, considering there are one to five echoes detected, which are often located in the second third of the signal, surrounded by noise at the beginning and the end. The size of the convolution kernel is 3. More details about the architecture are presented in Figure 4.13.

Training

Two different networks were trained: a **monospectral model**, processing green waveforms only, and a **bispectral model** that uses both the green waveforms and the infrared intensity values to perform echo detection and classification.

Both networks were trained on batches of 30 items using weighted **categorical cross-entropy** as the loss function and **stochastic gradient descent** for the optimization process. Categorical cross-entropy is defined by the following expression:

$$\mathcal{L} = -\sum Y_{true} \log\left(Y_{pred}\right)$$

where Y_{true} are the true class probabilities and Y_{pred} the predicted class probabilities. Both Y_{pred} and Y_{true} are normalized so they sum to 1 for each waveform.

The set of weights resulting in the lowest **validation loss** was kept for each configuration (mono- or bi-spectral).

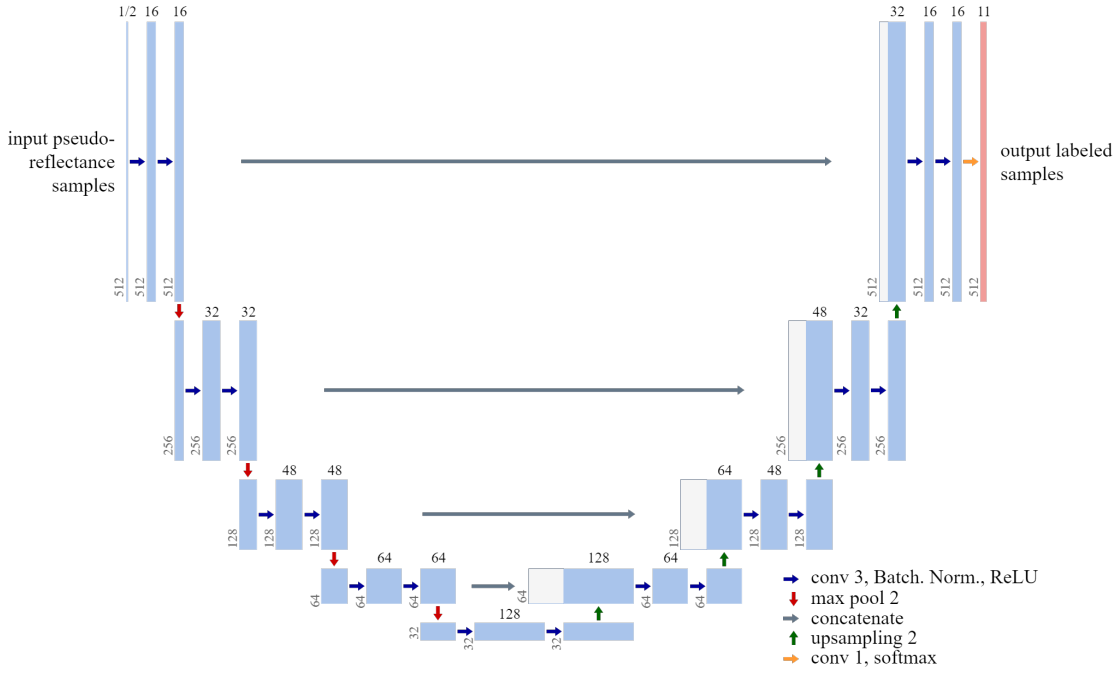


Figure 4.13: Diagram of the 1-dimensional U-time neural network architecture implemented to detect and label returns corresponding to sea- or land-covers in green waveforms.

Post-processing

The network tends to label several adjacent samples as an echo. However, in the labelled data, echoes are located in the form of single waveform samples. Consequently, the predictions were post-processed to keep only one sample as the location of a detected echo. Each time several contiguous samples obtained a label different than zero, the sample with the highest prediction probability was kept as is, while the others were re-labelled as zero (class "nothing").

Quantitative assessment of the results

Accuracy (ratio of good classification, best when its value is 1), **precision** (fraction of correct classes among each ground truth class, best when its value is 1), **recall** (fraction of correct classification for each estimated class, best when its value is 1), **F1-score** (combination of precision and recall, best when its value is 1), and **confusion matrices** were used to assess the quality of the predictions quantitatively. Different assessments were performed to identify the strengths and weaknesses of our approach.

The quality of the **peak detection** was assessed by considering this task as a binary problem (presence or absence of a return). By analyzing the predicted location of the samples having a label different than

zero regardless of the accuracy of the label, we can assess the performance of our architecture for waveform components detection by computing accuracy, precision, recall and F1-score. Considering the width of the peaks, the fact that the training labels correspond approximately to their maximum and the usual considerations for waveform processing, a margin of error can be accepted for the location of the peaks. The peak detection assessment was therefore also conducted with a tolerance of ±2 samples offset.

The second aim of our approach is to detect the land- or sea covers in the waveforms. The different objects do not necessarily need to be placed correctly in the waveforms since the simple information of their presence or absence can already be used for mapping purposes. **Classes detection** is therefore a relevant criterion to evaluate our approach. It was assessed by considering the classes predicted in each waveform, regardless of their occurrence or positions and aims at determining whether the neural network detects the different objects correctly. Accuracy, precision, recall and F1-score were computed for this task too.

The last evaluation criterion is the quantification of the **quality of the global classification**, i.e. assessing the prediction made on each sample for each waveform. Confusion matrixes were computed to get detailed information on the behaviour of the networks and the main confusions impacting both peak detection and classification.

4.3.3 Results

Monospectral network

The first network was trained only on the **green waveforms**. It reached its lowest validation loss (0.11) on the 106th training epoch. The weights obtained at this epoch were kept to test the model's performance. The following observations were made based on the testing dataset.

Qualitatively, the results show good identification of the different ecosystems. The parts of the waveform that are labelled as a return are consistent with the waveform components. As expected, before post-processing, continuous ranges of samples are often labelled as echoes of the same class, covering the whole peak. Since the ground-truth dataset is labelled with one sample indicating a peak, post-processing is useful if we want to assess quantitatively the quality of the detection and classification. Examples of results obtained in the testing dataset are presented in Figure 4.14.

Peaks detection Although most of the returns seem to be correctly placed when visualizing the predictions, metrics presented in Table 4.6 show that less than half the returns predicted actually exist. However, more than half of the echoes are properly detected.

If we choose to tolerate an error of 2 in the predicted location of the echoes (i.e. consider that if an

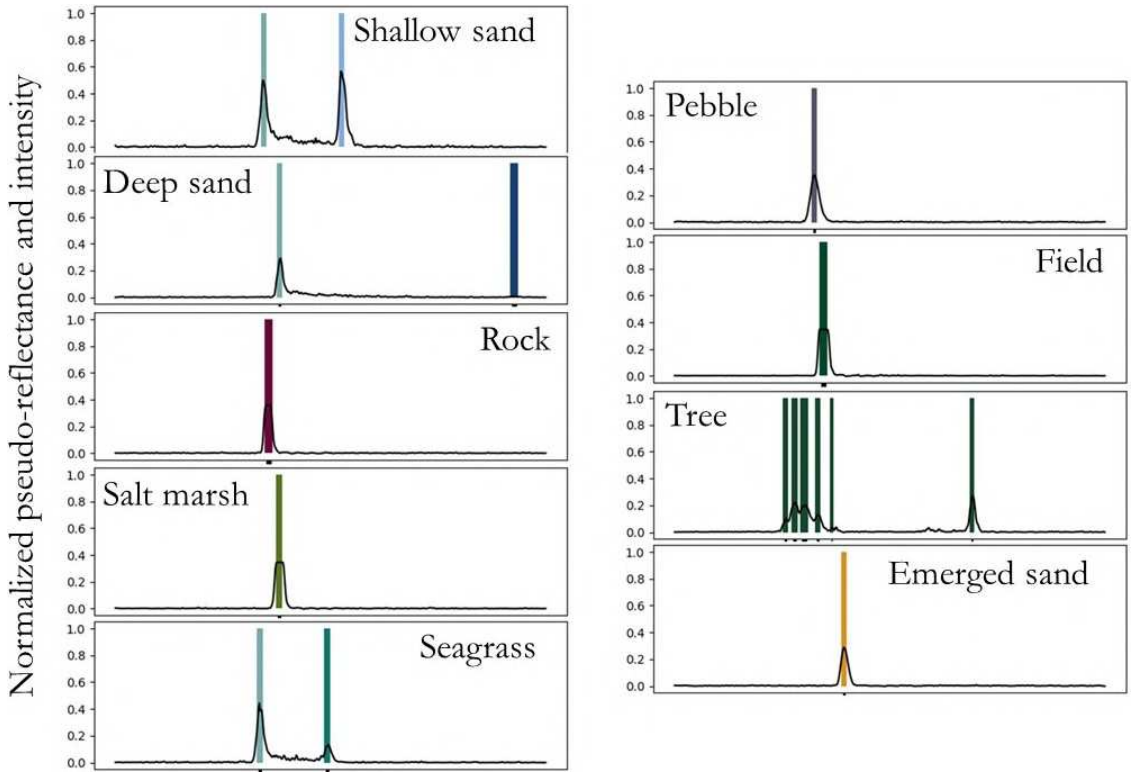


Figure 4.14: Examples of predictions made by the green waveforms-based U-time. The black signal corresponds to the waveform. Each coloured vertical line refers to a predicted peak. All nine cover types are pictured.

Metric	No location error tolerance	Location error tolerance of 2
Precision	0.42	0.63
Recall	0.62	0.94
Accuracy	0.99	0.99
F-score	0.50	0.76

Table 4.6: Performance metrics of the peak detection obtained with the green waveforms-based neural network

echo was predicted two samples further or earlier than the truth it is a correct prediction), the statistics improve significantly. Let us recall that this margin of error does not massively impact waveform analysis, since peaks are wider than 2 samples and the locations indicated in the training data correspond to the middle/maximum of the peaks (our goal here is not to improve nautical charting, but explore semantic information extraction possibilities).

With this error margin, more than **60% of the echoes predicted are real**, and **94%** of the existing returns are **identified** by the model. There is still an overestimation of the presence of returns, but few of the real echoes are missed. Furthermore, **98%** of the echoes truly located 2 samples away from the detection were **originally detected** by the network but erased during post-processing. In practice, it means that our network predicts these echo positions, but not with the highest probability of the local neighbourhood.

Classes detection In terms of land- or sea-cover identification, the monospectral network performs well. On average, **89%** of the covers predicted at a given location are **present in reality**, and **99%** of the objects hit by the laser beam are **identified**. Few covers are missed, but some are inaccurately predicted, as Table 4.7 reflects.

General classification The average accuracy of the classification after post-processing is **0.61**. The confusion matrix obtained (Figure 4.8) gives further indications of the performances of the network for each type of sample. Considering the values present in the first column (predicted = 0), the major source of error is the **mislocation** or **misdetection** of the echoes. Indeed, the highest percentages of confusion are obtained when a label 0 is predicted instead of the actual label. Moreover, the percentage of misclassification often corresponds to the difference between 1 and the accuracy obtained for a class.

The few cover type errors can be summed up as follows. *Shallow sandy seabed*, *trees*, and *seagrasses* are

Metric	Average value
Precision	0.89
Recall	0.99
Accuracy	0.97
F-score	0.94

Table 4.7: Performance metrics of the class detection obtained with the green waveforms-based neural network.

		Predicted labels										
	Label	0	1	2	3	4	5	6	7	8	9	10
True labels	0	1	0	0	0	0	0	0	0	0	0	0
	1	0.25	0.74	0	0	0	0.006	0	0	0.002	0	0
	2	0.53	0	0.47	0	0	0	0	0	0	0	0
	3	0.68	0	0	0.29	0.001	0	0.002	0.007	0.002	0.002	0
	4	0.55	0	0	0.002	0.44	0	0	00.004	0	0	0
	5	0.39	0.002	0	0	0	0.6	0	0	0.002	0	0
	6	0.31	0	0	0.002	0	0	0.65	0	0	0.03	0
	7	0.52	0	0	0	0	0	0	0.48	0	0	0
	8	0.33	0.001	0	0	0	0	0.001	0	0.66	0.001	0
	9	0.28	0	0	0	0	0	0.02	0	0	0.69	0
	10	0.26	0	0	0	0	0	0	0	0	0	0.74

Table 4.8: Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the green waveforms based neural network.

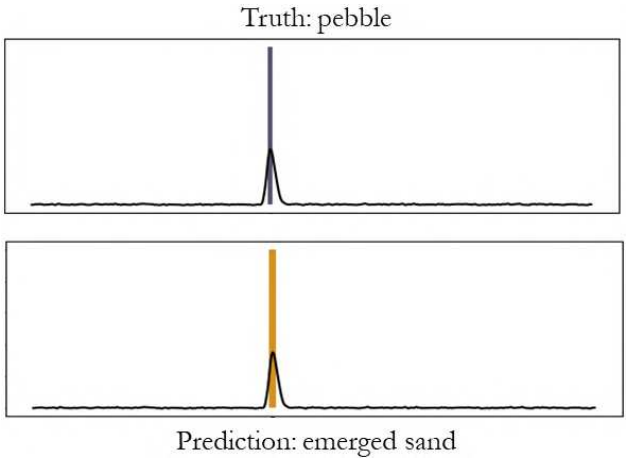


Figure 4.15: Examples of a prediction error made by the green waveforms-based U-time. The black signal corresponds to the waveform. The coloured vertical lines refer to a predicted peak. Here, the peak is located correctly, but the label is wrong.

sometimes mixed up. *Rock* is also mistaken for other types of terrestrial covers having similar saturated echoes (*pebble*, *tree*, *salt marsh* vegetation, or *field*), and inversely. Bare ground types can be misidentified too. Figure 4.15 illustrates this confusion between bare ground types.

The confusion matrix obtained on the post-processed results **with a tolerance** for a location error of 2 for the peaks (Figure 4.9) confirms the observation made above. The major explanation for the low accuracies obtained for most of the classes is the **improper location** of the returns by a few samples. Cover types errors remain rare, but there are still missed returns, and false detections, mostly of *deep sandy seabed* and *trees*. The average accuracy reaches **92%**.

Bispectral network

The second network was trained on **green waveforms** and corresponding **NIR intensities**. Its lowest validation loss (0.07) was observable on the 82nd epoch of the training process, providing the weights for the final model. The testing dataset was processed with this model and accounts for this configuration's performances for the detection and classification of waveform components.

Once again, visual appreciation of the predictions shows proper identification of most of the ecosystems and the returns. Overestimation of the number of samples to be labelled for a given echo is still observable and justifies the need for post-processing. Examples of results are presented in Figure 4.16.

Peaks detection The bispectral network shows **slightly poorer** precision, recall, and F-score for peak detection when no margin of error is tolerated than the monospectral model. 62% of the detected echoes

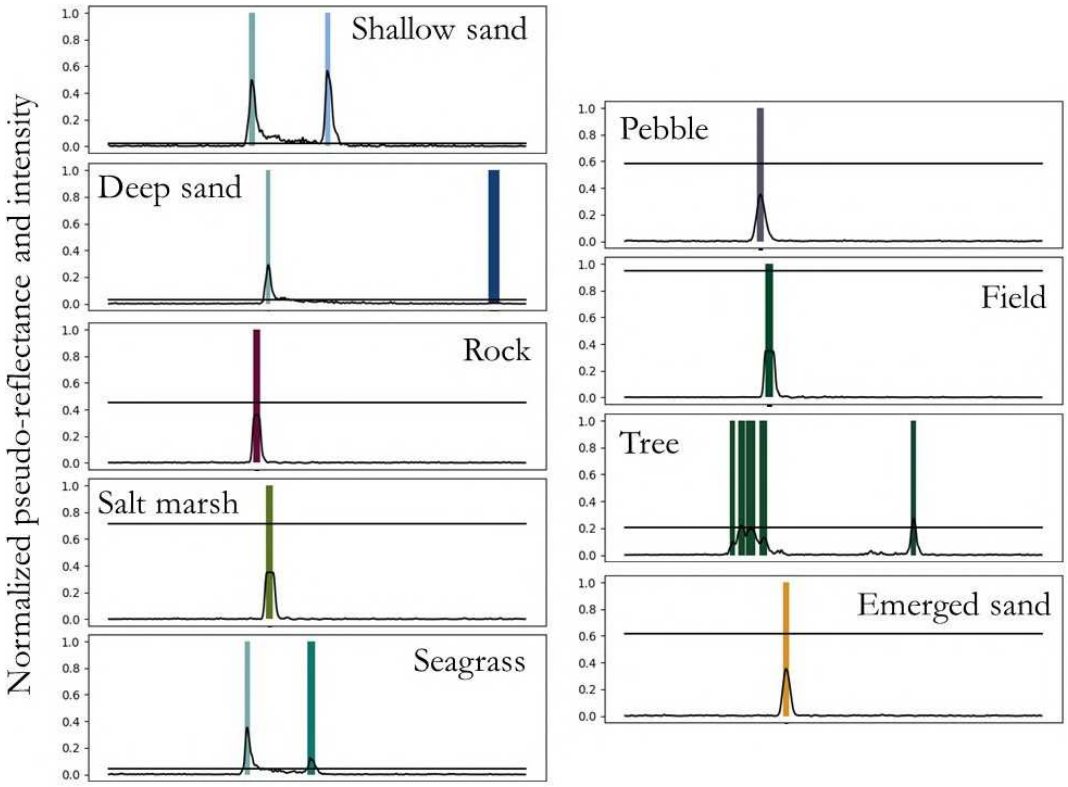


Figure 4.16: Example of class confusion in the green waveforms-based U-time predictions. The black signal corresponds to the green waveform. The black horizontal lines correspond to the near-infrared intensity values matched with each green waveform. Each coloured vertical line refers to a predicted peak.

		Predicted labels										
	Label	0	1	2	3	4	5	6	7	8	9	10
True labels	0	1	0	0	0	0	0	0	0	0	0	0
	1	0	0.98	0	0	0	0.01	0	0	0.004	0	0
	2	0.06	0	0.94	0	0	0	0	0	0	0	0
	3	0.2	0	0	0.75	0.03	0	0.002	0.01	0.002	0.002	0
	4	0.07	0	0	0.05	0.88	0	0	0.004	0	0	0
	5	0.03	0.002	0	0	0	0.96	0	0	0.002	0	0
	6	0.02	0	0	0.01	0.002	0	0.89	0	0.002	0.07	0
	7	0.07	0	0	0.01	0	0	0	0.92	0	0	0
	8	0.09	0.001	0	0.003	0.001	0.002	0.001	0	0.9	0.001	0
	9	0	0	0	0.006	0	0	0.04	0	0	0.95	0
	10	0	0	0	0	0	0	0	0	0	0	1

Table 4.9: Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the green waveforms-based neural network, with a tolerance of a location error of ± 2 samples for the returns.

do not exist in the truth dataset. And 2% less returns are properly detected. Table 4.10 enumerates the performance statistics of this second network for peak detection.

If, as previously, we set an **error margin** of ±2 for the location of the peaks, the **statistics increase**, but **slightly less** than for the first configuration. Almost **60%** of the echoes are **true positives**, and the network suitably detects **94%** of the **true returns**. As for the first network, 98% of the echoes actually located within 2 samples of the predictions were originally labelled as such, but erased during post-processing.

Class detection The bispectral network **predicts land- and sea-covers somewhat better**, as proven by the metrics presented in Table 4.11. Only 9% of the covers predicted are false positives, and **99%** of the **habitats are found**. The **average accuracy** of the identification of classes is **98%**. The global class detection metrics can be read in Table 4.11.

General classification The **global classification** task was performed with an average **accuracy of 0.61**. More details about classification errors can be found in the confusion matrix presented in Figure 4.12. Confusion between the label 0 and the others is again the biggest source of errors and appears to be

Metric	No location error tolerance	Location error tolerance of 2
Precision	0.38	0.59
Recall	0.60	0.94
Accuracy	0.99	0.99
F-score	0.46	0.73

Table 4.10: Performance metrics of the peak detection obtained with the infrared intensities and green waveforms-based neural network.

Metric	Average value
Precision	0.91
Recall	0.99
Accuracy	0.98
F-score	0.95

Table 4.11: Performance metrics of the class detection obtained with the infrared intensities and green waveforms-based neural network.

True labels \ Predicted labels	0	1	2	3	4	5	6	7	8	9	10
0	1	0	0	0	0	0	0	0	0	0	0
1	0.3	0.69	0	0	0	0.002	0	0	0.004	0	0
2	0.53	0	0.47	0	0	0	0	0	0	0	0
3	0.69	0	0	0.3	0	0	0	0	0.003	0.003	0
4	0.42	0	0	0	0.58	0	0	0	0	0	0
5	0.41	0	0	0	0	0.59	0	0	0	0	0
6	0.40	0	0	0	0	0	0.56	0.002	0	0.003	0
7	0.47	0	0	0	0	0	0	0.53	0	0	0
8	0.41	0	0	0	0	0	0	0	0.59	0.001	0
9	0.34	0	0	0	0	0	0.01	0	0	0.65	0
10	0.27	0	0	0	0	0	0	0	0.0007	0	0.73

Table 4.12: Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the infrared intensities and green waveforms based neural network.

dragging down the accuracy scores of most of the classes.

Less confusion between classes other than "nothing" is observable. *Underwater sand*, *seagrass* and *tree* are still mistaken for each other, as well as *rock*, *trees* and *dry sand* or *pebble*, *field* and *dry sand*. Confusion between *tree* and *sea surface* also occurrs, which was not the case for the first network. However, globally, there seems to be **less confusion between terrestrial covers**. The separation between subtidal and supratidal areas is somewhat better, except for the false tree detections in marine environments.

Similarly to the green-based predictions, with tolerance for a margin error of 2 samples in the peak locations, the average accuracy increases. It reaches **93%**, a 1% improvement, with green and infrared-based predictions. As shown in Figure 4.13, it also confirms that the slight offset in the location of some peaks accounts for most of the accuracy loss for ten out of eleven classes. Class confusions remain similar, and there are still inopportune detections of *tree* or *deep sandy seabed*, on top of missed echoes (for seven out of eleven classes).

		Predicted labels										
	Label	0	1	2	3	4	5	6	7	8	9	10
True labels	0	1	0	0	0	0	0	0	0	0	0	0
	1	0	0.99	0	0	0	0.006	0	0	0.004	0	0
	2	0.04	0	0.96	0	0	0	0	0	0	0	0
	3	0.23	0	0	0.76	0.005	0.002	0	0	0.005	0.005	0
	4	0.06	0	0	0.015	0.92	0	0	0	0.006	0.002	0
	5	0.03	0	0	0	0	0.97	0	0	0	0	0
	6	0.01	0	0	0.006	0	0	0.9	0.004	0	0.075	0
	7	0.07	0	0	0.002	0.006	0	0.004	0.92	0	0.002	0
	8	0.12	0	0	0.001	0	0	0	0	0.88	0.002	0
	9	0	0	0	0	0	0	0.03	0	0	0.97	0
	10	0	0	0	0	0	0	0	0	0.001	0	0.996

Table 4.13: Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the infrared intensities and green waveforms based neural network, with a tolerance of a location error of ± 2 samples for the returns.

4.3.4 Discussion

Detection of returns

U-time detects almost **entire peaks** and not only their maximum halfway location, as they are indicated in the training data. However, this peak detection is consistent with typical waveform behaviour: peaks constitute a significant pseudo-reflectance variation compared to their surroundings in the waveforms. Since noise has approximately always the same shape and pseudo-reflectance and is labelled as "non-useful" in all the waveforms, U-time attributing to the rest of the peaks labels different than 0 is expected.

To assess the quality of our peak detection, we performed post-processing to obtain results comparable to the labelled data. The post-processed results showed poor *precision and recall for peak location*. Since the peaks in our training data were located around their maximum, we know that an offset of about 2 samples in their predicted location will still correspond to the peak in the waveform data and thus not alter their semantics. Considering this, we obtain metrics that show the proposed architecture is able to detect peaks in waveforms accurately and efficiently since little preprocessing was needed, and no impulse function modelling or curve fitting was required.

Moreover, the fact that 98% of the detected peaks located within the margin of error was labelled as

peaks before pre-processing asserts the ability of the proposed approach to correctly detect peaks. However, it is clear that the **main limitation** of this approach is the **waveform labelling procedure**. Indeed, using discrete peak locations is not adapted to convolutional neural networks with this approach, since they are mostly based on patterns of variation, and will therefore perform better in identifying the whole range of samples corresponding to the peaks. Labelling the peaks continuously in the training data would avoid needing post-processing and allow the use of the intersection over union as a relevant performance metric.

Furthermore, since the files storing the waveforms are structured to contain up to four returns, in some cases, all the actual returns might not be indicated. For example, our approach seems to overestimate the number of tree returns, but in some cases, the detected peaks seem to be relevant. Figure 4.17 illustrates such additional detections of *tree*. Our method could therefore be helpful to improve echo extraction from waveform data as in (Chauve et al., 2007), but the extra returns' accuracy has to be evaluated first.

For the peak detection task, the proposed approach showed satisfying performances with limited pre-processing, contrary to traditional full-waveform analyses. The green-only based U-time showed somewhat better results, but the small difference makes both networks equivalent.

Classification

The **detection of the different classes**, regardless of their positions in the waveforms, is more robust and accurate and shows **little confusion**. There are **more false detections than missed detections**, and further analysis and visualization of the predictions show that most of the false detections could be avoided. Indeed, most of the false detections are *tree* returns in marine environments and *deep sandy seabed* returns before the water surface or after the seabed return. Figure 4.17 gives an example of false detection of deep sand before the detection of a *water surface* return. By adapting the network architecture, these issues could probably be solved.

A **recurrent neural network** could be useful to make future predictions depend more on previous predictions along the pseudo-reflectance time series. For example, it could mitigate the apparition of false detections of *deep sandy seabed* when a seabed return has already been identified earlier in the time series. **Region-based** convolutional neural networks could also determine the useful range of the waveform and avoid untimely peak detection outside of it.

Another possible improvement is the **addition of other types of data** to further distinguish different environments. Geometrical features, traditionally used to classify PCs (Zhang et al., 2013; Yan et al., 2015), could make better use of the rich spatial context inherent to lidar data. However, they imply pre-processing: it would therefore be interesting to see if the resulting improvements could advocate for a more complex processing chain.

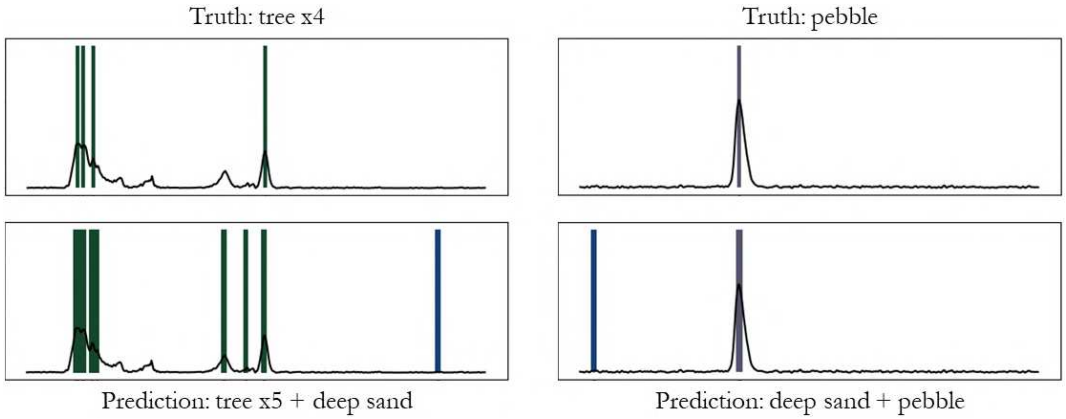


Figure 4.17: Examples of false predictions made by the green waveforms-based U-time. The black signal corresponds to the green waveform. Each coloured vertical line refers to a predicted peak.

Using both green and infrared waveforms is another solution. Indeed, since green wavelengths penetrate water and infrared do not, infrared intensity is an efficient predictor for the presence or absence of water. The bispectral neural network already shows a slight improvement in the separation of marine and terrestrial habitats, although tree returns are still detected in marine waveforms. Relying on more detailed infrared information would provide a better distinction between land- and sea-covers. However, further investigation on how to match NIR and green waveforms is required. Indeed, fusion encoders would need to be well designed to avoid the network to consider both wavelengths to be properly matched temporally, while they are not. Otherwise, training could be biased if a false equivalence of both signals in time and space is inferred.

Nonetheless, the results of our study show that there is not a massive difference in performance between the green waveforms-based network and the bispectral configuration, which is an interesting observation. The general classification is performed with an average accuracy above 92% with an error tolerance, and habitat detection in waveforms shows precision, recall, accuracy, and F-scores of at least 0.9 for both networks. Our approach therefore allows the classification of coastal and estuarine habitats based on lidar waveform processing, even with **only one wavelength**. It also obtains higher performance metrics than existent methods for similar coastal environment mapping (our experiments on seagrass meadows of Corsica with handcrafter features reported an 86% accuracy for seagrass and shallow sand mapping (Letard et al., 2021b), and Collin et al., 2012 obtained 90% of accuracy on their classification of coastal ecosystems). The green waveforms-based network also identifies correctly **almost 100% of the *water surface*** returns for shallow and deep waters, a promising perspective for full-waveform lidar processing for coastal and estuarine areas management.

Indeed, if the approach is valid on various water depths, it could make full-waveform lidar analysis more accessible and limit the need for a second wavelength dataset, as TB lidar acquisitions are expensive and land-water separation is traditionally made by using both green and NIR PCs (Lague et al., 2020). However, considering the tendency of our networks to label complete peak shapes and not punctual peak locations, this approach may be unsuitable to very shallow waters. Indeed, semantic segmentation doesn't allow class overlapping. Overlapping echoes may thus be impossible to delineate with U-time.

4.4 Conclusion

In this chapter, we investigated processing options to extract semantic information from topo-bathymetric lidar waveforms. We started by using proven methods for waveform classification and adapting them to bi-spectral datasets and a wider range of terrestrial and marine classes. Using a random forest algorithm and handcrafted waveform features, we obtained 3D classifications with accuracies above 85% and a mean prediction confidence of 77%. To further enhance waveform-based classification, we developed a semantic segmentation approach with a limited need for pre-processing compared to the more traditional method we first used. To achieve that, we trained two temporal convolutional networks, inspired by the U-Net architecture: one processing only green waveforms, the other processing green waveforms and infrared intensities. Our networks enable accurate sea- and land-cover classification through green lidar waveform processing. By tolerating an error of ±2 samples in the detected peaks' locations (this error implies no semantics change, and corresponds to 0.12 cm in reality), the waveforms classification reaches an average accuracy of 92% and 93% for the monospectral and the bispectral networks, respectively. In both experiments, we exploited the prediction probability and concluded it was a major tool for post-processing and a reliable way of filtering over-detections or classification errors.

In our first approach, we observed a significant performance increase when coupling both green and NIR wavelengths. However, a major advantage of the deep semantic segmentation method is its ability to perform excellently with green data only. Water surface returns were correctly identified in almost 100% of the cases by the green waveforms-based network, which also showed the best peak detection performances. The addition of infrared intensities did not show a significant improvement in waveform semantic segmentation.

Overall, our results with both traditional machine learning and deep neural networks illustrate once more the significant usefulness of topo-bathymetric lidar for diverse and complex environment surveying. Classified or segmented waveforms pave the way for enhanced ecological assessments of land-water areas by providing 3D information about the surface covers. Indeed, combining vertical structure information with the knowledge waveforms provide on the physical properties of the environment could serve

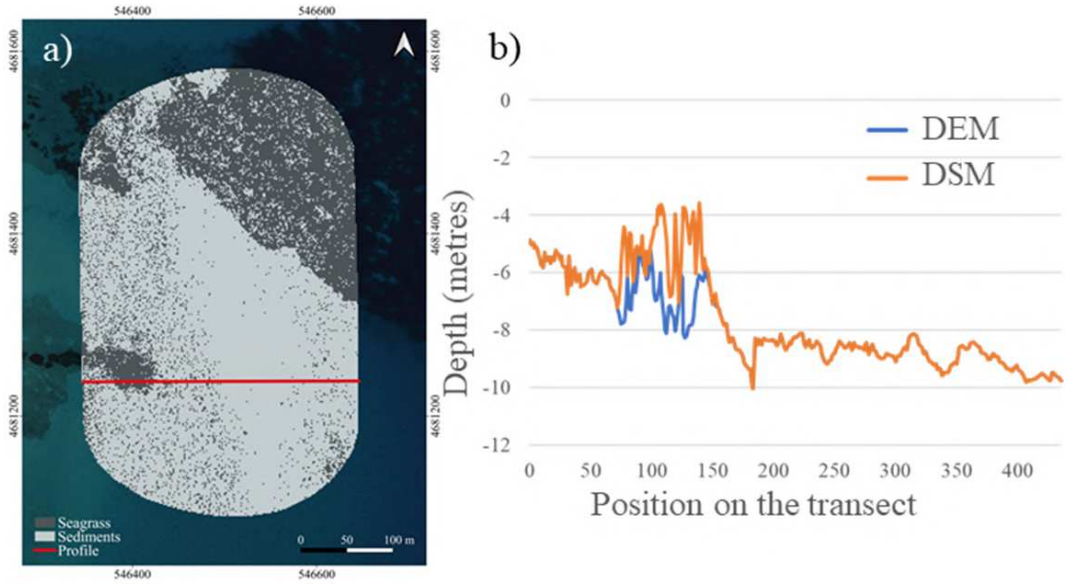


Figure 4.18: Examples of false predictions made by the green waveforms-based U-time. The black signal corresponds to the green waveform. Each coloured vertical line refers to a predicted peak.

as proxy data to evaluate ecosystem functions or water properties. As a matter of example, studying vegetation thickness or volume becomes possible, which is a key parameter of biomass assessments. Figure 4.18 below illustrates this opportunity[3]. It features preliminary results of seagrass meadow height extraction in Corsica. Although it should be refined with refraction correction, it already shows the benefits of a finer separation of underwater vegetation and seabed.

Confronting both methods, we conclude that globally, deep neural networks have significant advantages over a more classical machine learning approach:

- They require less pre-processing, and thus involve a lower uncertainty propagation;
- They do not require feature engineering;
- They seem to learn more abstract features allowing them to extract equivalent information from green waveforms alone and bi-spectral information.

This latter point is particularly interesting in the context of topo-bathymetric waveforms analysis. Indeed, full-waveform lidar data is massive and requires adapted storage and computing power to be used. The possibility of needing only one of both wavelengths to be recorded as waveforms is thus attractive. Furthermore, as we introduced in Chapter 1, studying bathymetric areas with lidar typically necessitates

3. Further details on the method used to obtain this result can be found in Letard et al., 2021b

the use of the NIR wavelength due to green surface return uncertainty. It is thus interesting to experiment with new data processing methods that could lift this requirement.

Consequently, in the next chapter, we focus on tackling the main limitations of this approach: its inadequacy to localize precisely actual backscatters, and its inapplicability to extremely shallow waters. Once we manage to improve these points, a spatial application of a deep learning-based waveform processing method will become possible for uninterrupted land-water interface modelling, and, potentially, for ecological assessments.

CHAPTER

5

EXPLOITATION OF PHYSICAL MODELS FOR WAVEFORM PROCESSING WITH NEURAL NETWORKS

Contents

Full-waveforms processing usually comes with two challenges: detecting intercepted surfaces, and characterizing them. Two main families of approaches can be used to automatically extract the nature of the surfaces illuminated: feature engineering combined with classical machine learning, or deep neural networks trained to learn consistent representations of the data. In this Chapter, we further explore the knowledge extraction possibilities offered by deep neural networks for lidar waveform processing presented in the previous chapter by investigating straightforward simultaneous component detection and identification from raw lidar signals and bathymetric waveform inversion. To this end, we base our experiments on physical models of lidar waveforms, both to allow the training of our models through access to simulated labelled data and to increase the quantities of training samples. At the end of sections 5.3 and 5.4, we discuss the application of these models to real data acquired with an Optech Titan lidar sensor.

5.1 Context

In the previous chapter, we saw how crucial backscatter detection is for the extraction of semantic information from lidar waveforms. In U-time, we proposed a unified approach to perform both echo detection and classification simultaneously. However, this approach requires to classify each time step and to localize the echoes with a single position. This is not optimal, as backscatters modify the local shape of the waveform on multiple time steps. Although attributing the class 0 ('non useful') to noise is a solution, it is not optimal since the associated optimization problem is difficult. In a similar way, changing the pre-processing in order to label all the echo is an alternative, but this would make the characterization of overlapping returns impossible - at least not with this semantic segmentation approach. Therefore, here we turn towards object detection and classification frameworks, which seemed more adequate for waveform processing.

Additionally, our first experiments with neural networks to analyse lidar waveforms encouraged us to go further and exploit them for different extraction tasks. For example, waveforms contain information on water's optical properties (Lu et al., 2022) or vegetation's physical characteristics. In chapter 2, we saw that the extraction of such knowledge is mostly based on regression approaches trained with in-situ measurements. However, it is in practice difficult to access such data, as simultaneous field samplings and airborne surveys are time-consuming and expensive. Another possibility reviewed in chapter 2 was to inverse radiative transfer models (RTMs). This task is however complicated, as the parameters are interrelated and there are thus many possible solutions. Such approaches were also outperformed by neural networks-based regressions in (Zhang et al., 2023). In the second half of this chapter, we thus explore the possibility of developing a non-iterative environmental parameter estimation method. We also wish to analyse whether waveforms collected without synchronous field measurements can still be exploited for

environmental parameter analyses. In particular, we investigate the possibility of using physical models designed for waveform simulation for algorithm optimization.

Whether for waveform component detection and identification or for parameters estimation, a major problem arises: how to extract precise labels? Indeed, in the case of **object detection**, two issues remain:

- It is difficult to assess with precision the beginning and end of a backscatter without introducing bias. This difficulty exists for peaks that are easy to spot with the naked eye but is increased for less prominent ones.

- Some echoes are impossible to spot with certainty by the human eye. Consequently, if some components are missed by the operator - for example, weak or overlapping returns - the algorithm is trained on unstable information. It can thus falsely label mixed backscatters as a single return and have trouble identifying it with its mixed features. In the end, it will be neither stable nor reliable.

In the case of **environmental parameters estimation**, similar issues arise to label the position of the water surface and the water bottom. However, the main difficulty lies in the obtention of the right parameter values to predict. It is in fact impossible to associate a value with each parameter without synchronous measurements. As reviewed in chapter 2, some methods thus use MODIS observations at different wavelengths and convert them into the desired spectral domain. The issue is that this approach is not adapted for shallow coastal waters or narrower inland water extents. Here, we want to assess the possibility of using a similar network for different types of water.

This issue of labelled data availability is addressed in this chapter. We suggest using a radiative transfer model rigorous enough to realistically reproduce typical waveforms. We then generate as much labelled data as desired, and design and train our own algorithms on them. In a second step, we apply the resulting networks to real waveforms acquired with a narrow FOV topo-bathymetric system to evaluate how they handle the domain change.

5.2 Main principles

To reach our objectives, we developed a simulation model for bathymetric waveforms and for single-return topographic waveforms. Since waveform components' shapes depend on the emitted laser impulse, we choose to include it as an input on the side of the waveform to process. However, emitted laser impulses are not registered for all sensors. Yet, they contain crucial information as they influence the global parameters of the waveform peaks (see chapter 1). We thus make the hypothesis that their main characteristics - namely their width, skewness and asymmetry - can be evaluated on a planar emerged surface intercepted at a low off-nadir angle. Consequently, we simulate such waveforms as well and use them

as inputs to our networks. Our simulations are made on the parameterized for a topo-bathymetric sensor such as the one we used for real data collection: the Optech Titan.

5.2.1 Training on simulated waveforms

We relied on a physical model of lidar waveforms to simulate training, validation, and test data. As explained in Chapter 1, each waveform component is the result of the convolution between the incident laser pulse and the response function of the intercepted surface. The waveform is then the sum of all individual components, plus a noise component. In this section, we detail how we simulate the incident laser pulses, the surface response functions, and the noise component.

Incident laser pulses We first simulate the pulses emitted by the sensor. These functions depend on the peak laser power P_{em}, the time duration of the pulse - which corresponds to its width - w_e, the time t and the time step at which the pulse is emitted t_e. Our model includes four different impulse functions $I(P_{em}, w_e, t, t_e)$ to generalize to different types of laser beams:

- A Gaussian impulse function (Jutzi et al., 2006);

$$I_g(P_{em}, w_e, t, t_e) = P_{em} \frac{2}{w_e} \sqrt{\frac{\ln(2)}{\pi}} \exp(-4 \ln(2) \frac{t - {t_e}^2}{w_e^2})$$

- An extreme value distribution;

$$I_{EV}(P_{em}, w_e, t, t_e) = t_e + P_{em} \exp\left(-\exp\left(-\frac{t}{w_e}\right) - \frac{t}{w_e} + 1\right)$$

- A generalized extreme value function;

$$I_{GEV}(P_{em}, w_e, t, t_e) = t_e + P_{em} \frac{\exp\left(-(1 + 0.15\frac{t}{w_e})^{\frac{-1}{0.15}}\right)(1 + 0.15\frac{t}{w_e})^{-1-\frac{1}{0.15}}}{\exp(-1)}$$

- A log-normal distribution

$$I_{logn}(P_{em}, w_e, t, t_e) = P_{em} \frac{1}{t w_e \sqrt{2\pi}} \exp\left(-\frac{\ln(t - t_e)^2}{2w_e^2}\right)$$

These functions are illustrated in Figure 5.1:

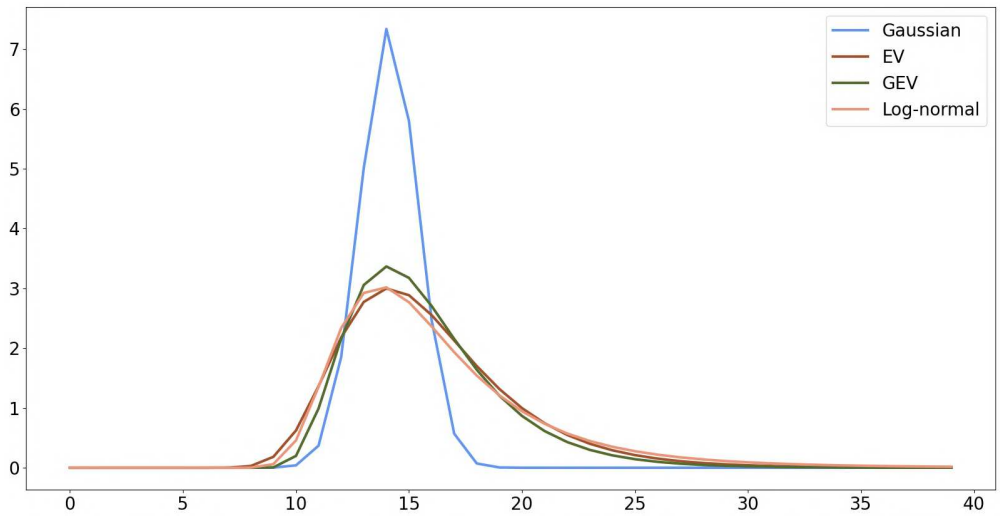


Figure 5.1: Impulse functions used for waveform simulation.

The main differences between these different impulse models are their skewness and asymmetry. In particular, Gaussian impulses are completely symmetric and much less skewed than the other functions. GEV and EV are close, but GEV has a slightly shorter rising time and shorter tail. The EV distribution has the longest rising time, with log-normal impulses close behind. GEV, EV, and log-normal functions are all asymmetric, and thus have a shorter rising-time than tail. GEV and EV are particularly similar to the shape of the pulse function of the Optech Titan sensor used in this document for real data processing.

To define the incident laser beams, we take into account the effects of non-nadir incidence through pulse stretching modelling, which increases the width of the emitted pulse depending on the range, the beam divergence angle, and the incidence angle. To this end, we follow the pulse stretching modelling introduced in Guenther et al., 2000. It relies on the estimation of the time delay generated by non-nadir surface interception, Δt:

$$\Delta t = \frac{2H}{c}\left[\frac{1}{\cos(\theta + \frac{\gamma}{2})} - \frac{1}{\cos(\theta - \frac{\gamma}{2})}\right]$$

where c is the speed of light, H the operating altitude, θ the beam incidence angle, and γ the beam divergence angle.

Since the water depth is negligible compared to H, we use the same expression to compute Δt for topographic surfaces, water surface, and water bottom returns and simply use the corresponding values of θ.

This time delay is caused by the difference in range between the closest portions of a non-planar surface and its furthest portions. Depending on the value of Δt compared to the width of the emitted pulse

w_e, it causes the incident pulse to be stretched by a factor τ:

- If $\Delta t < 2w_e, \tau = 0.1\Delta t$
- If $\Delta t \geq 2w_e, \tau = w_e \times \left(\frac{0.5\Delta t}{w_e} - 0.4\right)$

In our model, each intercepted surface's response function is thus convolved with an incident impulse of width $w_e + \tau$ to produce the waveform components.

Intercepted surface response function For the concerned applications, we simulate four types of components: topographic surfaces, water surfaces, water columns, and seabeds. Contrary to the other components, water columns do not necessitate the definition of a response function, as they have a very specific behaviour, explained in Chapter 1. However, for all other surfaces, a response function representing the geometry of the cross-section and its material must be defined.

To introduce sufficient variability and representativity in our datasets, we take into account different settings for the seabed and topographic surface components. We model these surfaces using Gaussian pulses with varying widths representing surface geometries and amplitudes symbolizing various materials. These amplitudes correspond to the received powers P_{rec} and P_b backscattered by topographic and benthic surfaces.

Since we aim at simulating small footprint lidar data, we consider water surfaces to be locally plane and thus simulate them with Dirac distributions δ, which we multiply with the value of the received power after reflection on the water surface.

To compute the power backscattered by each component, our simulator relies on 4 main physical equations, introduced in Chapter 1:

- Equation 1.4 describing the power received after reflection on a single topographic surface:

$$P_{rec} = \frac{\rho P_{em} T_{atm}^2 A_r \eta_e \eta_r \cos\theta}{\pi R^2}$$

- Equation 1.6 describing the water surface backscatter power for bathymetric waveforms:

$$P_s = \frac{L_s P_{em} T_{atm}^2 A_r \eta_e \eta_r}{\pi R^2}$$

- Equation 1.7 describing the water column component of bathymetric waveforms:

$$P_c = \frac{\beta_\pi \exp\left(-2K_d r_w\right)(1 - L_s) P_{em} T_{atm}^2 A_r \eta_e \eta_r}{n_w^2 R^2}$$

- Equation 1.8 describing the received power after reflection on the seabed for bathymetric waveforms:

$$P_b = \frac{\rho_b \exp\left(-2K_d r_w\right)(1 - L_s) P_{em} T_{atm}^2 A_r \eta_e \eta_r}{\pi n_w^2 R^2}$$

Globally, we make the hypothesis that the water column is homogeneous and that the factor of loss linked to the FOV is included in the loss factor L_s. Indeed, since we simulate a topo-bathymetric lidar survey, the FOV is probably too narrow to capture all energy and provide a precise estimate of the diffuse attenuation coefficient.

Noise To generate the final waveform, we sum all components with a noise component μ. We incorporate two noise functions:

- a Gaussian white noise $\mathcal{N}(0, 1)$
- a Gaussian white noise summed with a sinusoidal noise of fixed period (30 time steps)

$$\sin(t \times \frac{2\pi}{30}) + \mathcal{N}(0, 1)$$

These noise time series are multiplied by a factor N, which corresponds to a percentage of the maximum waveform energy.

Final waveforms In the end, bathymetric waveforms W_{bathy} are obtained with the following equation:

$$W_{bathy} = I(P_{em}, w_e + \tau_s) * (\delta(t_s) \times P_s) + I(P_{em}, w_e) * (P_c) + I(P_{em}, w_e + \tau_b) * (\mathcal{N}(t_b, w_b) \times P_b) + \mu N$$

where τ_s and τ_b are the stretching factors at the water surface and the water bottom, w_b is the width of the seabed surface response Gaussian denoting its geometry, and t_s and t_b are the time steps at which the water surface and the water bottom are intercepted.

Topographic waveforms W_{topo} are obtained with:

$$W_{topo} = I(P_{em}, w_e + \tau_t) * (\mathcal{N}(t_t, w_t) \times P_{rec}) + \mu N$$

where τ_t is the stretching factor at the topographic surface, t_t is the time step at which this surface is intercepted, and w_t is the width of its surface response function, denoting its geometry.

Simulation of training, validation and test datasets In our processing approach, we do not assume that the emitted pulse is always available. We thus simulate simple topographic waveforms acquired over

Label	Definition
1	Extremely turbid waters
2	Extremely shallow turbid waters (overlapping echoes)
3	Turbid waters
4	Turbid waters with high energy loss at the air/water interface
5	Deep turbid waters (weak echoes)
6	Extremely shallow clear to moderately turbid waters (overlapping echoes)
7	Clear to moderately turbid waters
8	Clear to moderately turbid waters with high energy loss at the air/water interface
9	Deep clear to moderately turbid waters (weak echoes)
10	Very deep clear waters with highly reflective seabeds
11	Shallow clear to turbid waters with low albedo seabeds

Table 5.1: Definition of the 11 types of bathymetric waveforms simulated.

planar, tar-like surfaces with low incidence angles to approximate the impulse function. When processing real data, such waveforms can be used to estimate the laser pulse function, as they are very close to it in shape. To train our networks, we thus need arrays of simulated bathymetric and topographic waveforms and their inherent parameters. To limit bias and favour generalizability, we train our networks to be robust to various environmental contexts but also to different sensor parameters. For a given set of bathymetric waveform parameters, we thus simulate signals for each impulse function type, and then disturb each of them with each type of noise implemented, for three different levels of noise. The initial bathymetric waveform parameters - K_d, L_s, R_b, β_π, depth, θ and w_b - and the impulse parameters - P_{em} and w_e - are drawn randomly within fixed ranges. For each bathymetric waveform, a topographic waveform is simulated with the same impulse function parameters and the same noise function and noise level. The topographic waveform parameters are fixed: the reflectance of the ground is set at 0.05 to denote tar, θ is set at 0.17 rad, and the width of the Gaussian surface response function is 0.1 m.

To reflect the complexity of benthic environments, we simulated eleven types of bathymetric waveforms in different physical contexts. They are presented in Table 5.1.

For each type of bathymetric environment, the parameters of the model are drawn randomly from a uniform law within pre-defined ranges of parameters. Each waveform depends on the values of depth, K_d, β_π, L_s, R_b, θ, P_{em}, w_e, w_b, the type of impulse function, the type of noise, and the noise level N. Some parameter ranges are common to the whole simulation, they are presented in Table 5.2.

However, other parameters related to the physical context vary within different ranges depending on

θ	P_{em}	w_e	w_b
0.08 to 0.6 rad	2 to 4 mJ	0.2 to 0.8 m	0.1 m to w_e

Table 5.2: Simulation parameters common to all waveform types.

Label	Depth	K_d	β_π	R_b	L_s
1	0.2 m to 2.5 m	1.5 m^{-1} to 5 m^{-1}	0.1 to 0.4	0.03 to 0.85	0.05 to 0.85
2	0.2 m to 1 m	0.7 m^{-1} to 1.5 m^{-1}	0.003 to 0.009	0.03 to 0.85	0.05 to 0.85
3	1 m to 3 m	0.7 m^{-1} to 1.5 m^{-1}	0.003 to 0.009	0.03 to 0.85	0.05 to 0.3
4	1 m to 3 m	0.7 m^{-1} to 1.5 m^{-1}	0.003 to 0.009	0.03 to 0.85	0.3 to 0.85
5	3 m to 5 m	0.7 m^{-1} to 1.5 m^{-1}	0.003 to 0.009	0.03 to 0.85	0.05 to 0.6
6	0.2 m to 1 m	0.1 m^{-1} to 0.7 m^{-1}	0.0002 to 0.003	0.03 to 0.85	0.05 to 0.85
7	1 m to 4 m	0.1 m^{-1} to 0.7 m^{-1}	0.0002 to 0.003	0.03 to 0.85	0.05 to 0.3
8	1 m to 4 m	0.1 m^{-1} to 0.7 m^{-1}	0.0002 to 0.003	0.03 to 0.85	0.3 to 0.85
9	4 m to 10 m	0.1 m^{-1} to 0.7 m^{-1}	0.0002 to 0.003	0.03 to 0.85	0.05 to 0.5
10	6 m to 20 m	0.1 m^{-1} to 0.4 m^{-1}	0.0002 to 0.002	0.5 to 0.85	0.05 to 0.4
11	1 m to 2.5 m	0.1 m^{-1} to 1 m^{-1}	0.0002 to 0.008	0.03 to 0.2	0.05 to 0.5

Table 5.3: Specific simulation parameters associated to each waveform type.

the type of environment simulated. They are summed up in Table 5.3.

For each set of simulation parameters drawn randomly, waveforms are simulated with each type of impulse function, each type of noise, and each level of noise, resulting in 24 simulated waveforms per random drawing. Table 5.4 illustrates the result obtained for one drawing of simulation parameters.

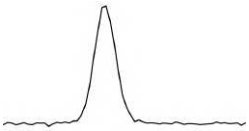
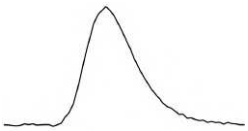
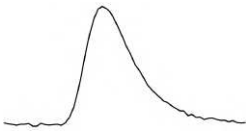
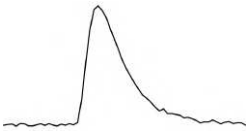
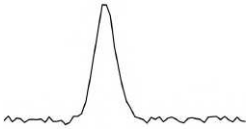
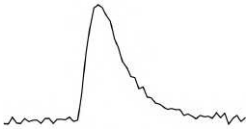
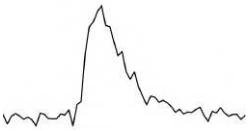
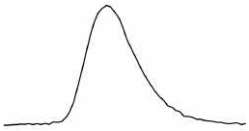
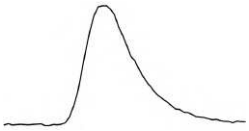
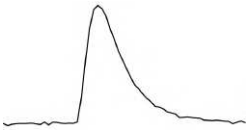
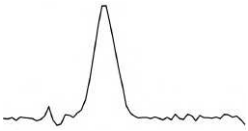
	Gaussian impulse	EV impulse	GEV impulse	Log-normal impulse
Low Gaussian noise				
Medium Gaussian noise				
High Gaussian noise				
Low Sin-Gauss noise				
Medium Sin-Gauss noise				
High Sin-Gauss noise				

Table 5.4: Illustration of the different waveforms obtained for a single set of simulation parameters.

The training dataset is composed of 200 random drawings of bathymetric waveform parameters, resulting in $200 \times 24 \times 11$ training couples of bathymetric and topographic waveforms. The validation dataset has half the amount of items, and the test dataset is as large as the training set.

5.2.2 Application to real data and domain adaptation

After having trained and evaluated the models on simulated data, we apply them to real waveforms to evaluate the applicability of the developed approaches. These data were acquired with an Optech Titan topo-bathymetric lidar equipped with a 1064 nm NIR laser and a 523 nm green laser. Only green waveforms were recorded. They are samples with a time interval of 1 ns, corresponding to 0.1275 m in water. The sensor does not start and end recording received power at fixed time intervals. Consequently, the waveforms obtained have varying lenghts, and are typically shorter in shallow waters. More details about the sensor can be found in Chapter 3 and in Lague et al., 2020.

The impulse function is estimated by selecting a waveform acquired over a road with a low incidence angle. The same example of topographic waveform is used to process all bathymetric waveforms acquired during the same survey.

The trained models are first applied to the raw Titan waveforms. In a second step, we explore the possibility of using ad-hoc techniques, and in particular domain adaptation.

Domain adaptation is a specific case of transfer learning that consists of adapting a machine learning model trained on data from a source domain to make it generalize and perform well on data from another related domain. It is necessary when a model is applied to a dataset that has a different distribution than the training dataset used to build this model. In such cases, the network tends to perform poorly on the new distribution. In our case, the domain adaptation process involves projecting insufficiently labelled data into the domain of another set of correctly labelled data to benefit from the knowledge of models learned on this annotated dataset.

There are several possible methods to perform this type of domain adaptation task. Here, we use optimal transport (OT) to transport real waveforms in the simulated waveforms' domain.

Optimal transport for domain adaptation

Optimal transport is an old mathematical problem that consists of finding the best correspondence between production in mines and transport towards manufactures. It can be viewed as a problem of distribution matching and enables to find a transport map that transfers a dataset in a source domain into a target one (or vice versa). For that, a cost matrix between all elements of source and target distributions is computed. From this matrix, the lowest transportation cost between one distribution and the other is

estimated.

In this study, we use optimal transport in two different manners. The first aims to perform domain adaptation, while the second consists of transporting outputs back to their original domain. These two ways are detailed below.

1 **Input waveforms adaptation.**
The goal is to transport real waveforms into the domain of simulated ones. For that, we consider the simulated dataset as a distribution of waveforms, and the real dataset as another distribution of waveforms. We then use optimal transport to find a transformation between both domains.

2 **Domain adaptation of the predictions.**
When facing a classification problem, the outputs of the model applied to the transported waveforms can be directly exploited. However, in this case, we estimate information related to the location within the waveform. In this situation, we have positions within the waveform derived from the transported space (simulated data). To obtain a consistent result (positions in the domain of real data), estimating the position in the original waveform is necessary. Optimal transport can also be utilized here, but unlike the previous case where we transported the entire set of source waveforms to the set of target waveforms (both distributions consisting of a set of waveforms), here we transport each real transported waveform (to be applied to the neural network) back to its original. In this case, each waveform is treated as a distribution, and the transport solution will estimate the transformation allowing the transition from one waveform to another, thereby allowing us to adjust the estimated output positions on the original waveform.

Within OT, each time step of the real waveform is associated with a linear combination of time steps of the transported waveforms and vice versa. Figure 5.2 extracted from the Python POT library website[1] illustrates the transport matrix between two distributions. The lightest colours in the matrix represent the optimal correspondences between both series.

To keep discrete box locations, we first normalize the coefficients of the transport matrix across lines. We then determine time step correspondences using the highest normalized transport matrix coefficient and keep the value of this coefficient as a matching score.

For example, if the solution has a form

$$Tt_i = \alpha_1 Tr_{j-1} + \alpha_2 Tr_j + \alpha_3 Tr_{j+1}$$

1. https://pythonot.github.io/

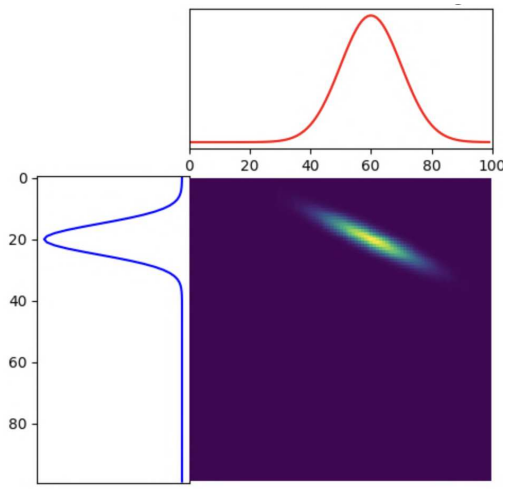


Figure 5.2: Illustration of the transport matrix between two distributions, from `https://pythonot.github.io/`.

with T_t the transported positions, Tr the real positions, i the position in the transported time series, j the position in the real time series, $\alpha_1 < \alpha_2$, and $\alpha_3 < \alpha 2$, we take j as the correspondence for i, with a *matching score* α_2. In Figure 5.2, T_t is the blue distribution, and T_r the red one. In such a context, the 20^{th} element of T_t would be associated principally with the 60^{th} element of T_r, but also with the adjacent positions, as the transport matrix illustrates.

To summarize, the global approach used in this chapter is to base the design and optimization of neural networks for waveform processing on physical models to tackle the issue of labelled data unavailability. Optimal transport is then used to perform domain adaptation and process real data. In Section 5.3, we present a neural network designed to perform waveform component detection and identification. In Section 5.4, a network to estimate parameters related to the water column from bathymetric waveforms is introduced.

5.3 Simultaneous waveform pattern detection and identification

In this section, we investigate object detection using deep neural networks as a solution for bathymetric waveform component detection in challenging settings. To this end, we develop a convolutional neural network inspired by the existing image processing network YOLO, and adapt it to bathymetric waveform processing. The network is trained using simulated data; its performances are thus discussed regarding metrics obtained on simulated test data. Preliminary results of its application to real acquisitions are also presented.

5.3.1 Methodological background: the YOLO approach

In image processing, object detection and identification is a common problem that has resulted in the development of multiple dedicated solutions and is now well dealt with (Zhao et al., 2019). Options allowing real-time detection to perform object tracking in videos exist, along with slower approaches favouring higher precision in the prediction of object locations and higher accuracy in their identification.

Globally, two main types of frameworks exist to solve the problem of object detection in images (Zhao et al., 2019):

- **Region proposal-based networks**: also called two-stage detectors, they first scan the whole image to predict regions of interest (ROIs) and then classify these ROIs. The network generating region proposals and the CNN used to perform classification are trained separately. Thus, the time spent handling these different elements is a bottleneck for real-time applications (Zhao et al., 2019).

- **Regression/Classification-based networks:** these networks are also referred to as single-stage detectors, as they only contain one step. They perform a global classification and regression task by directly predicting bounding box coordinates and class probabilities from the image pixels. They thus reduce the time expense and are thus more suited to real-time application (Zhao et al., 2019).

In the second category - regression/classification-based - the YOLO approach is a popular framework (Zhao et al., 2019). YOLO was introduced in Redmon et al., 2016 and stands for "You Only Look Once". As its name suggests, its principle is to encode the image with convolution operations only one time and perform unified object detection and classification.

The global principle of YOLO is illustrated in Figure 5.3. It first consists of splitting the input image into a grid. Each grid cell is then responsible for the object it contains. From a one-shot convolutional encoding, several bounding boxes are predicted for each grid cell, with an objectness score and a class probability. A threshold is applied to the objectness score to keep only boxes containing objects.

In practice, YOLO predicts, for each cell of the grid, the offset of the bounding box's planar coordinates compared to the bottom left corner of the cell. In Redmon et al., 2017 and Redmon et al., 2018, the authors introduce several improvements to the initial approach. Instead of predicting bounding boxes for only one scale, YOLOv3 has three outputs consisting of three grids of different sizes. This allows it to better handle the presence of objects of varying sizes in the images. The three different prediction layers each specialize in handling different ranges of box width and height using anchor boxes. Anchor boxes are boxes used to initialize the three different prediction layers. They are supposed to represent different widths and heights of boxes present in the training data and help attribute each box to its responsible prediction layer during training. Each grid cell of each output layer predicts three possible bounding boxes. They are then pruned depending on their objectness score and prediction confidence.

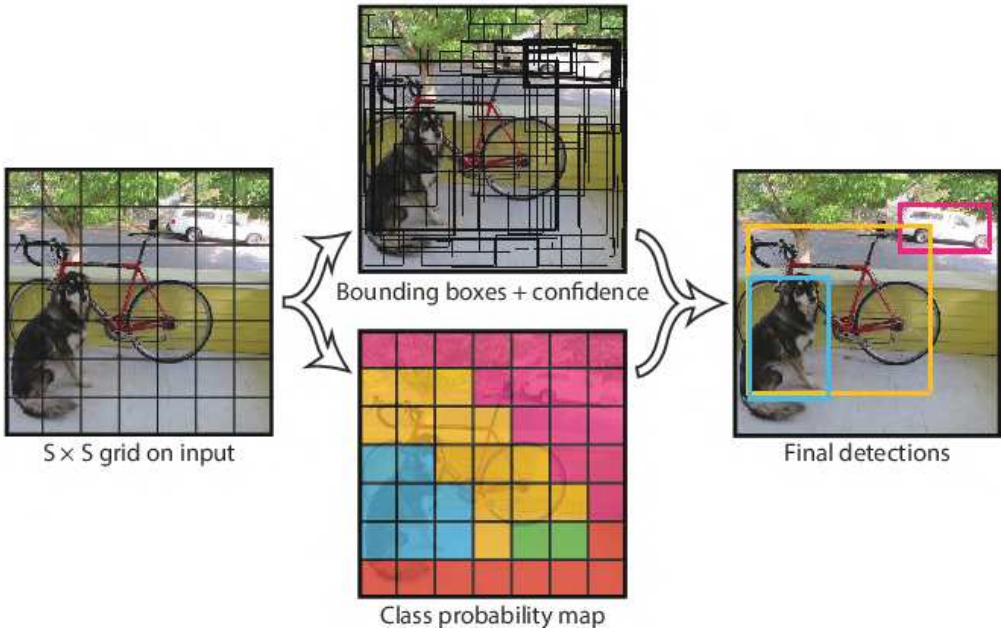


Figure 5.3: Illustration of the approach used in the YOLO network, from Redmon et al., 2016.

YOLOv3 relies on a convolutional backbone named Darknet, which implements the principles of Inception networks (Szegedy et al., 2015) and residual networks (He et al., 2016). It is optimized by summing errors on bounding box x, w, and label. In this section, we exploit this version of the network and adapt it to 1D data.

5.3.2 A YOLO-inspired network for waveform pattern identification

Several changes have to be incorporated into YOLOv3 to adapt it to waveform processing. The following paragraphs explain the resulting design and the corresponding architecture, input, and output changes. Figure 5.5 features a simplified illustration of the resulting architecture.

Backbone There are several possibilities to encode the information available in lidar waveforms. Since lidar waveforms are specific types of time series, they can be processed with traditional MLPs and CNNs, but also with networks dedicated to sequential input processing. Among them, we investigated LSTM- and GRU-based recurrent neural networks. We also experimented with Transformers. In the end, we observed better performances with convolution-based networks and thus decided not to use RNNs, whether based on GRU or LSTM modules. However, as expected considering their increasing popularity in the machine learning community, we found interesting properties of Transformers in terms of

performance and optimization, and thus keep them as a solution to further investigate in the future to reduce the computation cost of our approaches.

In the end, we base our advanced developments for waveform processing on convolutional encoders built from convolutional blocks organized as illustrated in Figure 5.4.

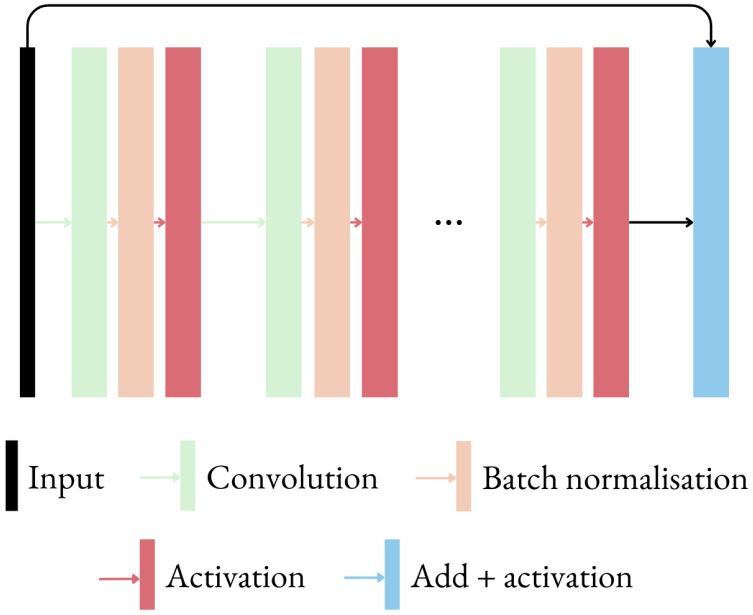


Figure 5.4: Illustration of the convolutional block used as the base of our waveform encoders.

We thus kept the convolutional encoding used in YOLOv3. The main change we applied was to turn all 2D convolutions into 1D convolutions.

We also adapted some aspects of the Darknet architecture, namely the downsampling factor: its value is 16 for us, versus 32 in the original YOLOv3. This means that at the end of the encoding, feature maps have a length equal to one-sixteenth of the input waveform size.

We evaluated how adding dot-product self-attention or cross-attention could help. Qualitatively, we did not find any added value. However, we observed that replacing skip connections with cross-attention resulted in a significant performance decrease. Further quantitative evaluation would thus be required to choose whether to add attention and in which form.

Data fusion with parallel encoders To accommodate for our dual input - bathymetric and topographic waveforms - we doubled the original YOLOv3 encoder to have two parallel encodings for each waveform, as Figure 5.5 illustrates.

Originally we used strictly identical encoders for the bathymetric waveform and its topographic counterpart. However, we observed that it biased the application to real data: the network seemed to attribute more weight to bathymetric waveform peaks located around the topographic waveform's only backscatter, namely to locate the water surface.

We thus changed the topographic waveform encoding. To help the network focus on the shape of the backscatter rather than its temporal location, we changed the way it is fed to the network. Now, we only input the portion of the topographic waveform located close to its maximum. The input topographic waveform has a size of 64, with 32 samples on each side of the maximum. We found that our network performs better this way and generalizes better to real data.

This resulted in changes in the topographic waveform encoder, which has the same depth as the bathymetric waveform but performs fewer pooling operations. After encoding, both feature vectors have the same sizes - reduced in length by a factor of 16 - and are concatenated before being further convolved.

Prediction layers We also had to adapt the prediction layers of YOLOv3. Instead of predicting x and y coordinates and the height and width of each bounding box (BB), we predict only one coordinate x and an associated width.

In the end, we have three prediction layers corresponding to three grids: a 16x16 grid, a 32x32 grid, and a 64x64 grid. Each of them predicts three vectors per grid cell. These vectors contain the following elements:

- x: the offset of the BB centre compared to the left corner of the cell;
- w: the width of the BB;
- o: the objectness score of the BB, ranging from 0 to 1 and representing a probability that the bounding box predicted actually represents an object;
- a list $c_1, \ldots, c_n$, with n the number of classes, containing the class probability of the object.

A predicted BB is considered to be an object if its objectness is above a user-defined threshold. If it overcomes the threshold, it is attributed the class for which it has the highest probability.

Optimization We kept the same optimization and training process as defined in YOLOv3. We thus use anchors during weight optimization. Similarly to the outputs, we adapted the dimensions of the anchors. Initially, anchors are typical boxes, defined by width and height. In our version, they are defined as typical time ranges and are thus only characterized by a width. We used anchors adapted to the typical size of our waveform components. The 64x64 grid thus specializes in widths below 10 samples, the 32x32

grid typically deals with boxes between 10 and 20 samples large, while the 16x16 grid handles the largest components with widths up to 50 time steps. In our labeled data, we consider the box to start and to stop when the signal is at one-tenth of the peak's maximum.

During training, validation, and testing, predictions are considered valid depending on several parameters. The main criterion is their maximum IoU with any of the true boxes. For each predicted BB, its IoU with all labelled components is computed. The prediction is attributed to the labelled box with which it has the highest IoU. If this IoU is below a user-defined threshold, the detection is considered to be a false positive. Otherwise, the object is considered a true positive. We will name this threshold the *IoU threshold* from now on.

Predictions can also be filtered depending on the predicted objectness of each bounding box. We will name this threshold the *object threshold* from now on. Lastly, when several overlapping boxes are predicted for the same class, they can be filtered depending on their objectness, too. To this end, the user sets an IoU threshold. If multiple predictions of the same class have an IoU over this value, only the one with the maximum objectness score is kept. Further on we will refer to this threshold as the *NMS threshold*, NMS being short for non-maximal suppression.

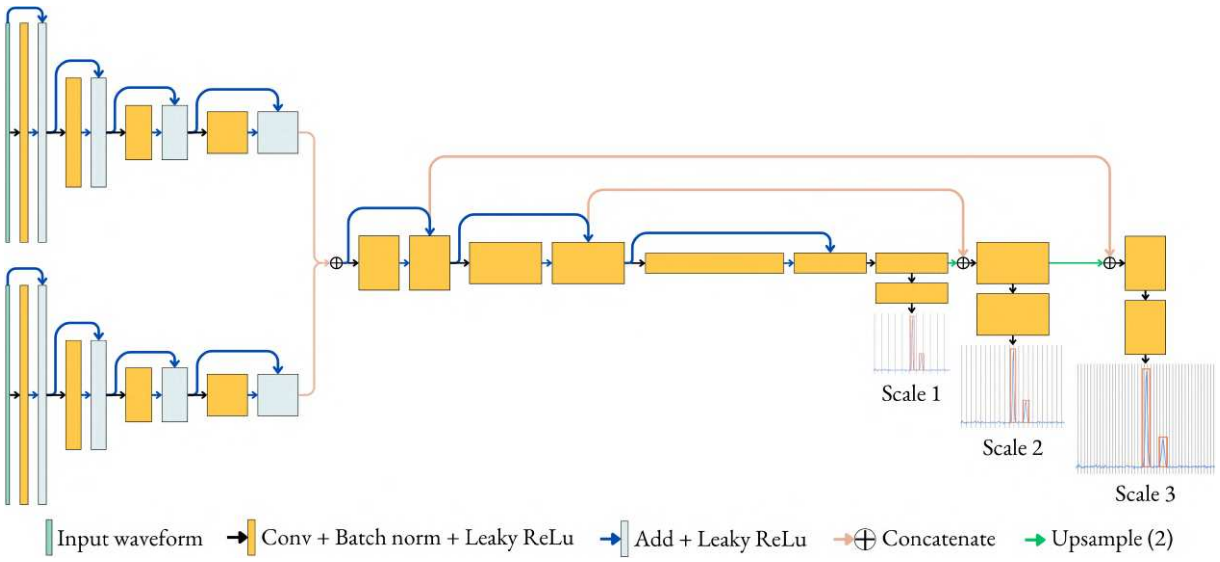


Figure 5.5: Architecture 1DYoloFusion of our bathymetric waveform component detection and identification network.

5.3.3 Training on simulated data and consequent challenges

Simulated data distributions are more homogeneous than natural data distributions, as it is impossible to simulate natural variations exactly. Consequently, even though our simulated data includes several types

of laser impulsions and noises to represent a maximum of cases, there is a risk of overfitting when training on simulated data, since they are all generated with similar background patterns. If this occurs, the network potentially learns patterns on the training data that make it good at predicting outputs on them, but not at processing the validation data. Instead, we want the network to learn the best transformations and representations of lidar waveform to predict accurate component location, width, and type.

Our first training of FuseYolo1D showed, as expected, an issue of overfitting. Typically, this results in high validation loss compared to training loss, and an ever-decreasing training loss. Both are observable in the validation and training loss curves in Figure 5.6.

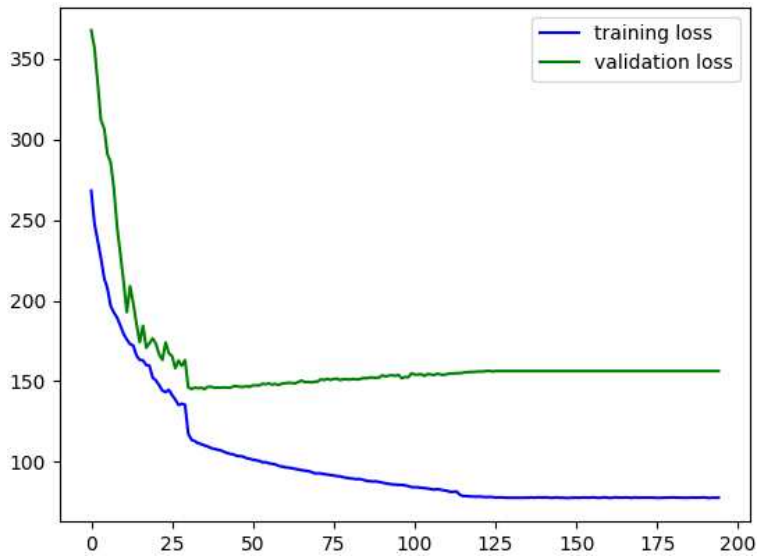


Figure 5.6: Training and validation loss during network optimization without data augmentation.

To tackle this issue, the complexity of the training dataset should be increased to capture the underlying features of the data instead of simple patterns. In our case, simply increasing the number of input training samples does not solve the problem, since it only provides the network with more similar training items. We thus explored data augmentation techniques to make the network more robust and thus more accurate on validation waveforms.

In image processing, training data are often augmented using random rotations, cropping, resizing, or histogram equalization. However, applying these transformations to our data makes little sense. Resizing the waveforms would bias all the temporal information, and mirror-like rotations would also discard all sequential information, which is the base of the waveform data.

Consequently, we explored more advanced data augmentation methods:

- Mixup (Zhang et al., 2017), which consists of randomly augmenting the dataset with linear com-

binations of different training inputs and their respective outputs;

- Cut-out (DeVries et al., 2017), which consists of randomly setting portions of the training items to 0;

- CutMix (Yun et al., 2019), which randomly replaces portions of training inputs with the corresponding portions of other training inputs.

All these modifications aim at preventing from overfitting by learning on more complex data. Figure 5.7, originally published in Yun et al., 2019 illustrates the results of each method on a dog image from ImageNet, an image classification benchmark dataset.

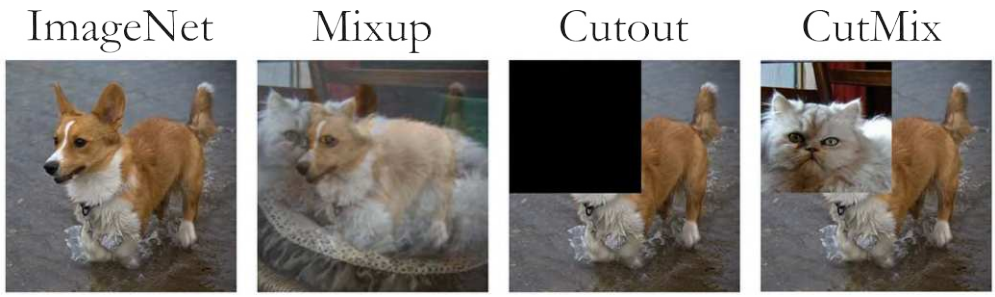


Figure 5.7: Illustration of the principles of Mixup, Cut-out, and CutMix, from Yun et al., 2019. In Mixup, two different input images are mixed with a linear combination; in Cut-out, a portion of the image is set to 0; in CutMix, a portion of the input image is replaced with the corresponding part of another one.

In the case of waveforms, the mixup process is slightly different, as we do not necessarily mix data of different classes, but simply select randomly a waveform of the training dataset to mix with each training waveform. Additionally, the labels are not mixed as in classification applications but concatenated. Similarly, when applying cutout, the output associated with each waveform must be adapted as well, to discard the components originally labeled in the part set to 0. Lastly, for CutMix, components labelled in the portion of the waveform replaced by another must also be replaced by the corresponding boxes of this second waveform.

Applying data augmentation to our training dataset helps the neural network's convergence, as illustrated by the different training and validation loss curves presented in figure 5.8. However, the training process could still be improved by applying the same augmentations to the validation dataset, as the resulting curves display lower loss for the validation data than for the training one, which suggests that the training data is harder to fit than the validation data in this current setting.

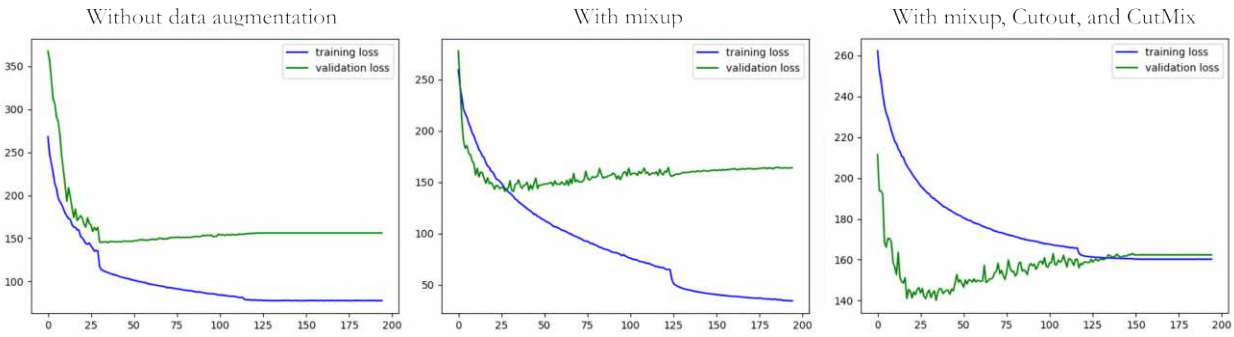


Figure 5.8: Training and validation loss during optimization depending on the data augmentation settings.

5.3.4 Results obtained on simulated waveforms

Evaluation metrics

To evaluate the quality of the predictions made by our 1DYoloFusion architecture on the test dataset, we compute several metrics adapted to object detection:

- Average precision, aP (without taking duplicates into account);
- Average recall, aR (without taking duplicates into account);
- Mean Intersection over Union of the predictions and their closer true box, mIoU_p ;
- Mean Intersection over Union of the validated detections and their associated true boxes, mIoU_d;
- First percentile of the distribution of detected components' signal-to-noise ratio (SNR), $\text{SNR}_{0.1}$;
- Percentage of missed components;
- Percentage of duplicate detections.

We consider duplicates to be validated predictions that have the same associated true box as another detection. When multiple predictions have the same associated true box, the one with the maximal IoU with the associated true box is kept, while the others are considered duplicates.

Overall results

The obtained metric averaged on both classes over all types of waveforms are presented in Table 5.5 below. Since a prediction is a true positive only if it has an IoU over a given threshold with a true box, we analyze the influence of this threshold on the metrics. By default, the IoU threshold used is 0.5, but we also assess

Class	IoU threshold = 0.5					IoU threshold = 0.75				
	aP	aR	$mIoU_p$	$mIoU_d$	$SNR_{0.1}$	aP	aR	$mIoU_p$	$mIoU_d$	$SNR_{0.1}$
Water surface	0.46	0.99	0.49	0.85	24.47	0.40	0.95	0.49	0.88	24.5
Water bottom	0.77	0.81	0.65	0.78	10.87	0.59	0.61	0.65	0.84	10.78

Table 5.5: Average performance metrics obtained by the network for each type of bathymetric waveform component.

the performances obtained when applying a threshold of 0.75 when distinguishing between *predictions* and *detections*. Here, we used an object threshold of 0.7 and a NMS threshold of 0.1.

Globally, several observations can be made. Water surface detection has a much lower average precision than average recall. This illustrates the tendency of our network to detect more surface components than actually exist. However, the average precision of 0.99 means that 99% of the water surface annotations in the test dataset are correctly detected. Given the way we computed recall and precision - i.e. without taking duplicates into account - the average precision of the surface detections suggests many false positives.

Water bottom detection features more balanced metrics. The average precision symbolizes a lower tendency to predict false additional seabed components. The average recall of 81% illustrates the capacity of our YOLO-based network to detect most seabed components. More water bottom components than water surface components are missed, which is expected since weak seabed returns occur more frequently than invisible surface backscatters.

Globally, predictions have a relatively high mean IoU with true components. This suggests that there are more predictions located close to real components than completely away from them. Additionally, the IoU of validated predictions is relatively high, even though $SNR_{0.1}$ is relatively low.

Increasing the IoU threshold does not impact significantly the lowest SNRs that can be detected. However, its results in a decrease of average precision and average recall for both classes, the most significant impact being on water bottom. Indeed, aP and aR lose 6% and 4% respectively for water surface, while they face a 18% and 20% decrease for water bottom.

Performance depending on the type of waveform

Although the developed approach allows the simultaneous detection and classification of waveform components, our main interest is to evaluate its detection capacity in challenging settings that usually make

Label	aP	aR	$mIoU_p$	$mIoU_d$	$SNR_{0.1}$	Missed (%)	Duplicates (%)
1	0.48	0.99	0.50	0.79	24.5	0.48	17.6
2	0.68	0.99	0.62	0.78	24.73	0.48	19.9
3	0.42	1.00	0.48	0.89	25.7	0.0	9.9
4	0.37	1.00	0.43	0.89	24.4	0.0	9.7
5	0.32	1.00	0.37	0.86	24.4	0.0	14.6
6	0.67	0.99	0.61	0.75	24.2	0.95	18.9
7	0.46	0.99	0.50	0.87	21.3	0.3	13.6
8	0.39	1.00	0.45	0.89	24.4	0.0	9.5
9	0.39	1.00	0.42	0.86	25.0	0.0	21.6
10	0.49	1.00	0.48	0.87	24.7	0.0	16.2
11	0.41	1.00	0.48	0.88	24.7	0.0	10.7

Table 5.6: Performance metrics of the water surface detection for each type of waveform.

other methods fail. Indeed, since we focus on bathymetric return detection in these experiments, the classification is not challenging at all. Consequently, we focus on the ability to detect echoes in bathymetric waveforms in the following paragraphs.

To this end, we analyze the abilities of our method to separate surface and bottom components in overlapping echoes - i.e. extremely shallow waters - and to detect weak water bottom returns with peak intensities close to the noise level.

The following tables present detection metrics depending on the types of waveform defined for data simulation (cf. Table 5.1). Labels 2 and 6 correspond to waveforms containing only overlapping water surface and water bottom returns. Labels 5, 9, and 10 regroup waveforms containing only strong water surface and weak water bottom returns.

Table 5.6 first summarizes the results obtained for water surface detection. The metrics were computed for an IoU threshold of 0.5, an object treshold of 0.7, and an NMS threshold of 0.1.

Globally, the number of duplicate detections of water surface components is high. It is at its lowest for very turbid shallow waters and clear shallow waters, and at its highest for extremely shallow waters and very deep clear waters. However, as expected considering the average recall for the whole class, missed detections are scarce: they never overcome 1% of the true components and are 0 for most types of waveforms.

The most varying factor is the average precision, still denoting a high tendency to falsely detect additional water surface echoes. Similarly to the percentage of duplicates, it appears as the main differentiating factor between overlapping and clearly separate water surfaces. Indeed, extremely shallow waters feature

much higher average precisions than other classes ($\simeq$ 67.5% versus 32% tp 48%). There seem to be three main categories of waveforms sharing similar ranges of aP:

- extremely shallow waters, with 0.99% of aR and 67.5% of aP;
- waters with prominent surface echoes - extremely turbid shallow waters, moderately turbid shallow waters, and deep clear waters -, with 0.99% of aR and 47.6% of aP;
- waters with less strong surface returns - shallow turbid waters, moderately turbid waters with high surface loss factor, deep moderately turbid waters, and shallow waters with high albedo -, with an aR of 99% and an aP of 38.3%

In the end, these three categories are consistent with the definition of the waveforms they symbolize. For example, overlapping echoes have a strong surface component and deep clear waters also feature a prominent surface return compared to the amplitude of the seabed return. They may include fewer patterns that are likely to be considered similar to surface returns.

However, although they have higher average precisions, extremely shallow waters and extremely turbid waters feature a mean mIoU_d of 77%, while the other categories have a mean mIoU_d of 88%.

Overall the lowest values of SNR detected do not vary significantly depending of the type of waveform. They are only slightly lower for clear to moderately shallow waters with low energy loss at the water surface, which are typically the easiest waveforms to process.

In table 5.7, the detection metrics are detailed for the water bottom class. The same parameters as those used for surface detection metrics were used: an IoU threshold of 0.5, an object treshold of 0.7, and an NMS threshold of 0.1.

For the detection of water bottom components, different tendencies emerge. Overall, there are different categories of settings characterized by decreasing performance metrics. Extremely shallow waters and shallow waters with low loss of energy at the air/water interface or with dark seabeds (labels 2, 3, 6, 7, 11) have the highest metrics: their aP, aR, and mIoU_d are the highest, and they have the lowest percents of missed annotations.

Shallow waters with high loss of energy at the air/water interface (labels 4 and 8) come second, with around 21% of missed annotations and mean aP and aR of $\simeq$78%.

Water bottom components of extremely turbid waters and deeper waters of varying turbidity (labels 1, 5, 9, and 10) come last. They have globally lower aP ($\simeq$58%) and aR ($\simeq$65%), and between 28% and 45% of missed annotations.

Overall, a major difference with water surface detection is the absence of duplicates for water bottom predictions. However, the proportion of missed components is much higher, with a peak at 45% for deep clear waters.

Label	aP	aR	$mIoU_p$	$mIoU_d$	$SNR_{0.1}$	Missed (%)	Duplicates (%)
1	0.64	0.67	0.57	0.75	15.9	32.6	0.0
2	0.94	0.96	0.78	0.82	23.1	4.11	0.0
3	0.92	0.92	0.76	0.80	15.3	8.04	0.0
4	0.77	0.79	0.64	0.74	10.4	20.54	0.0
5	0.69	0.72	0.60	0.73	9.3	27.9	0.0
6	0.94	0.95	0.78	0.81	22.9	4.70	0.0
7	0.94	0.95	0.77	0.81	19.4	4.94	0.0
8	0.76	0.79	0.65	0.76	11.5	21.37	0.0
9	0.53	0.66	0.46	0.74	9.8	33.51	0.0
10	0.46	0.55	0.40	0.76	12.2	45.3	0.0
11	0.94	0.94	0.77	0.81	13.4	5.83	0.0

Table 5.7: Performance metrics of the water bottom detection for each type of waveform.

Deep clear waters are the category for which we obtain the lowest performance. They include waters with K_d ranging from 0.1 m^{-1} to 0.4 m^{-1}, and cover depths between 6 m and 20 m. They are a particularly challenging category, as they contain mostly weak returns, potentially sometimes impossible to distinguish from noise. However, although they obtain weak performance metrics, the visual results on weak echoes show that our method is still able to detect components with low SNR.

Illustration of the results obtained with this method

In this section, we present visual results of waveform component detections obtained with our YOLO-derived network. These qualitative results allow us to better comprehend the performances of the network. We present visual results for each category of waters: turbid waters, moderately turbid waters, extremely turbid waters, deep clear waters, and shallow waters with low albedo seabeds.

These illustrations highlight the ability of the network to separate water surface and water bottom in extremely shallow waters for different noise levels and different impulse widths. They also show the possibility of identifying very weak seabed returns, without background noise detections. Lastly, they also show the ability of the network to handle different configurations: surface return weaker than the seabed backscatter, strong geometrical pulse stretching or low reflectance of the water bottom.

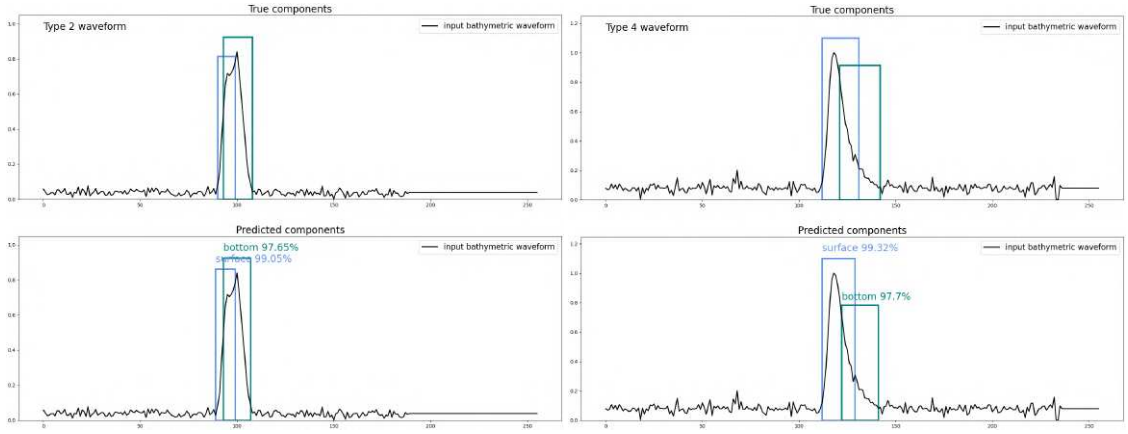
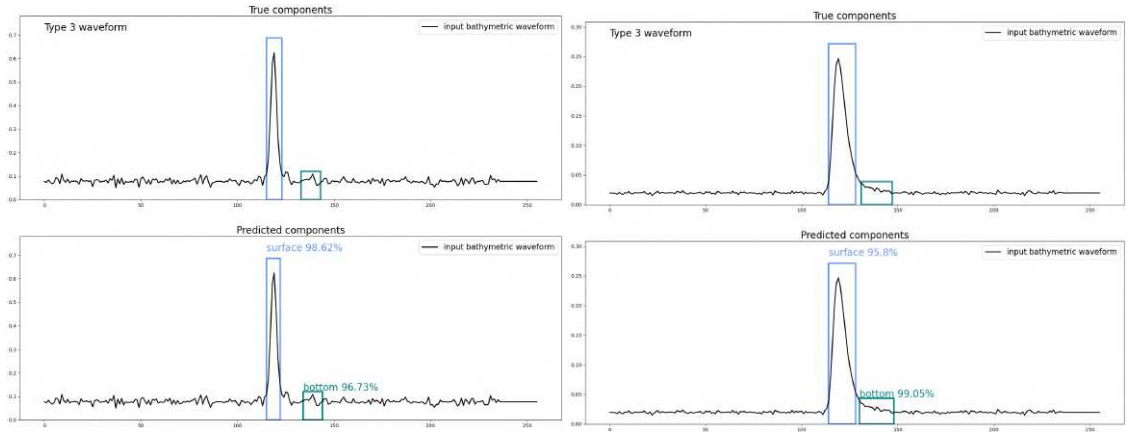
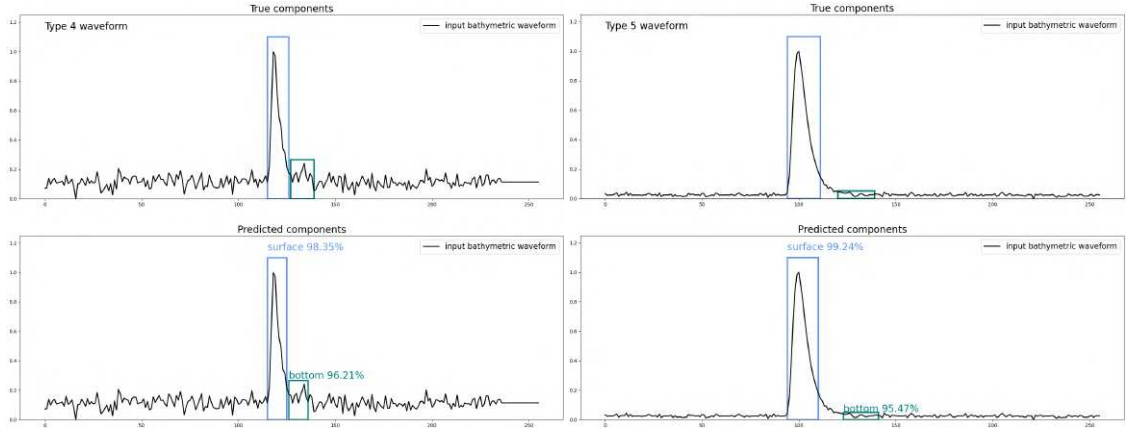
Very turbid waters




Table 5.8: Illustration of the results obtained in very turbid waters.

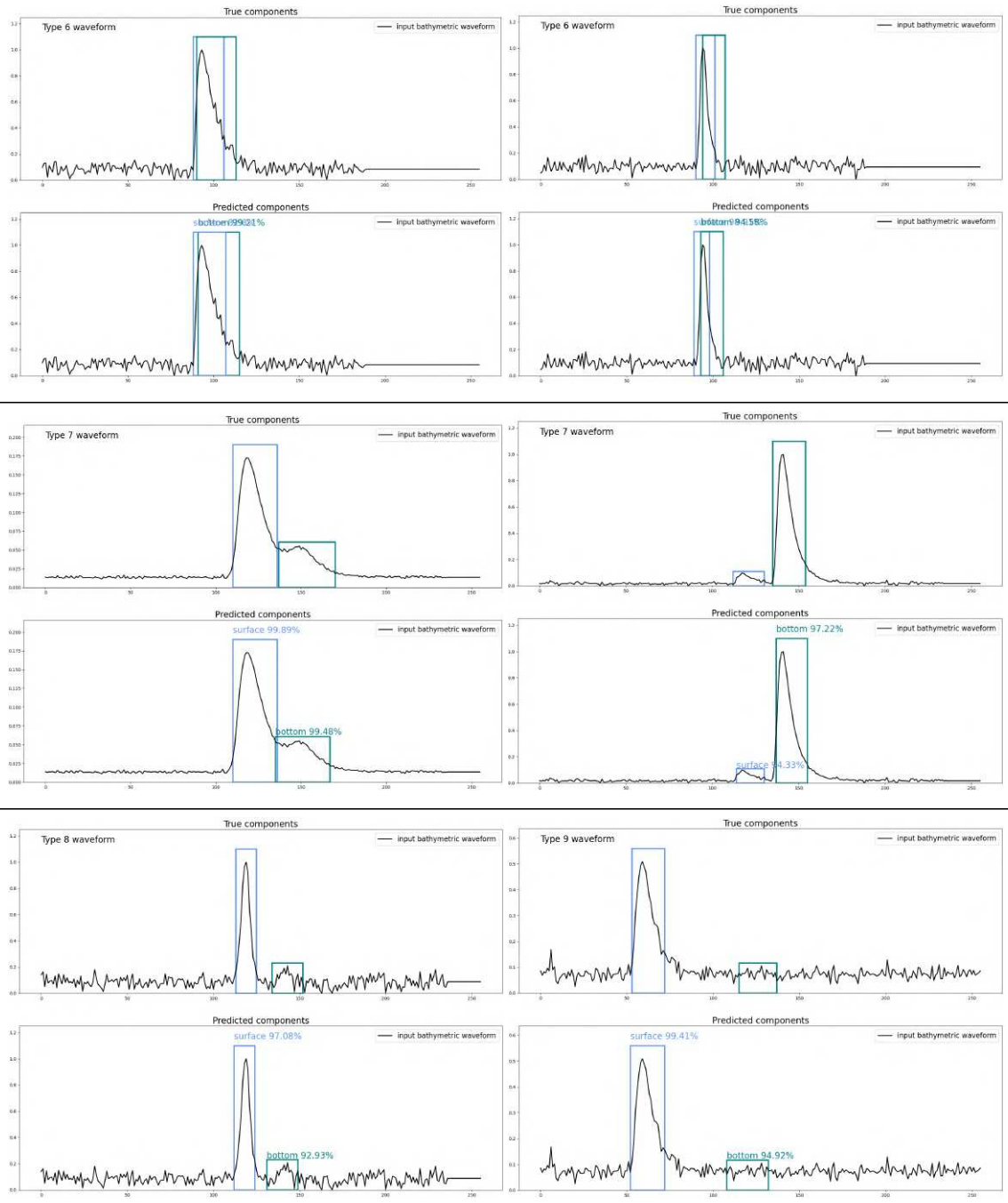
Clear to moderately turbid waters


Table 5.9: Illustration of the results obtained in moderately turbid waters.

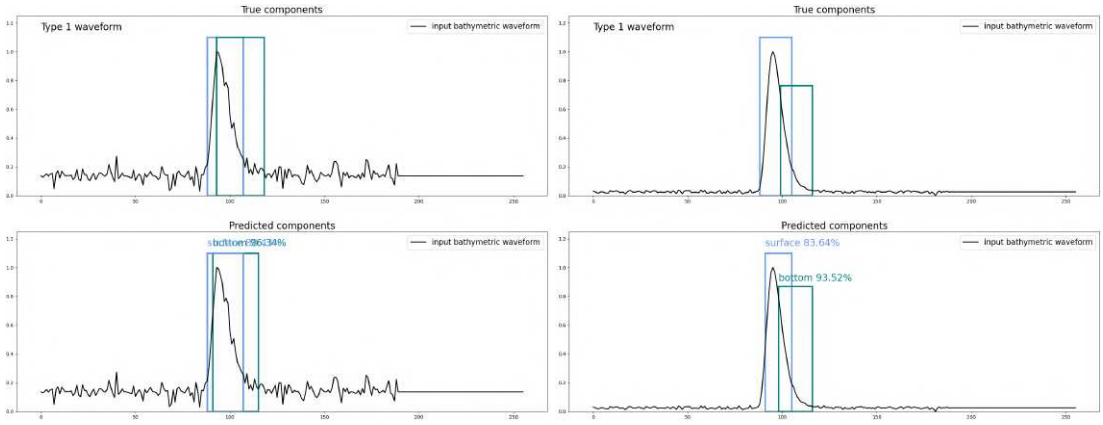
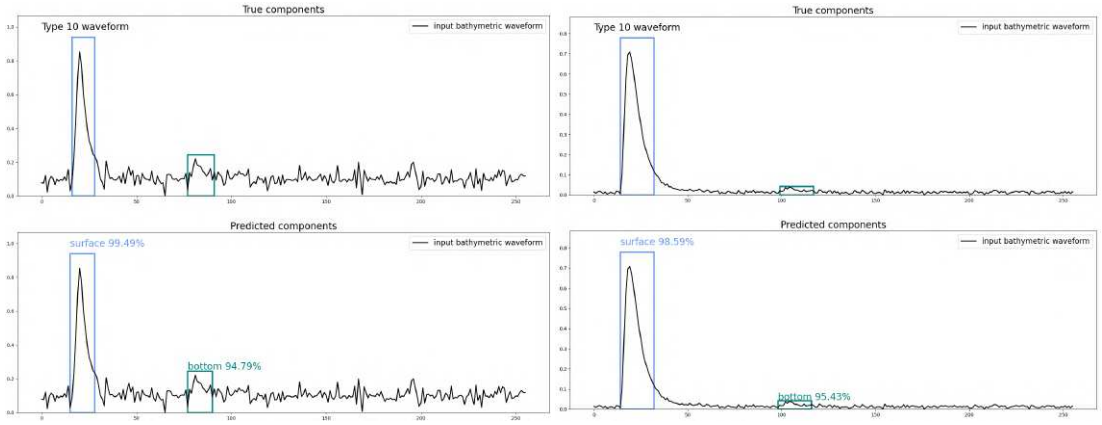
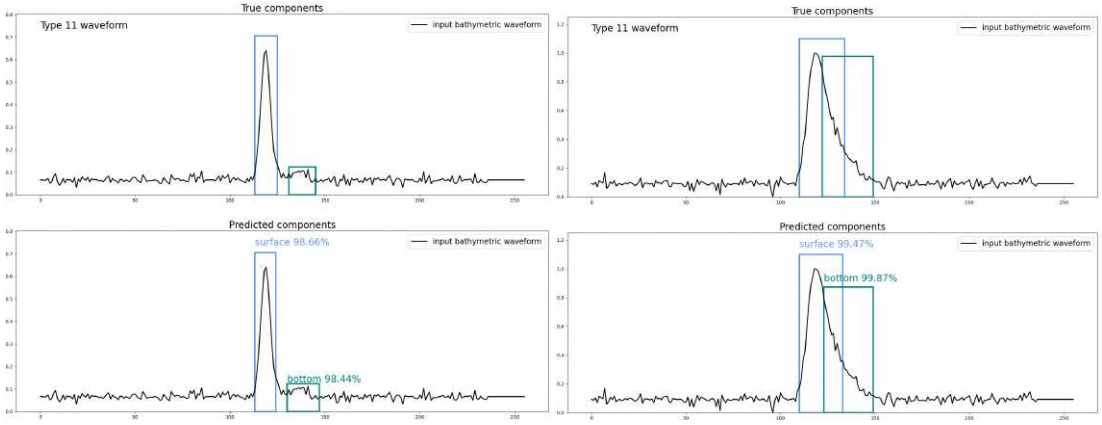
Extremely turbid waters

Very deep clear waters

Shallow waters with dark seabeds


Table 5.10: Illustration of the results obtained in extremely turbid waters, very deep clear waters, and shallow waters with dark seabeds.

To visualize the predictions, we also added a filter on the class-specific confidence score. Let us recall that, contrary to the objectness score used to validate a prediction or not, this score is defined in Redmon et al., 2016 as:

$$P_r\left(\text{Class}_i\right) \times \text{IoU}\ {}^{\text{truth}}_{\text{pred}} = P_r\left(\text{Class}_i \mid \text{Object}\right) \times P_r\left(\text{Object}\right) \times \text{IoU}\ {}^{\text{truth}}_{\text{pred}}$$

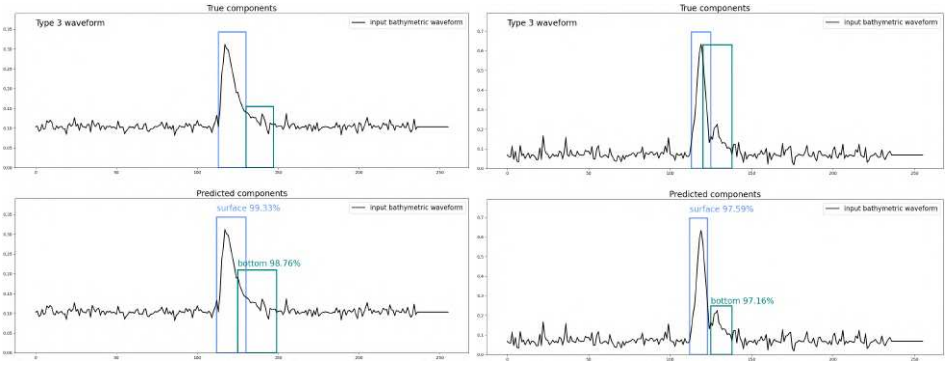
As it is typically not used for prediction validation, we did not use it when computing performance metrics. However, the qualitative results suggest that it might be a good addition to the filters used to consider a prediction a detection or not. Indeed, in the illustrations presented in Tables 5.8, 5.9, and 5.10, over-detections of water surface components are not visible, which suggests that they were removed by applying a threshold of 0.8 on the class-specific confidence score.

Despite this visual improvement when further filtering predictions, some errors persist. We illustrate them is the next paragraphs.

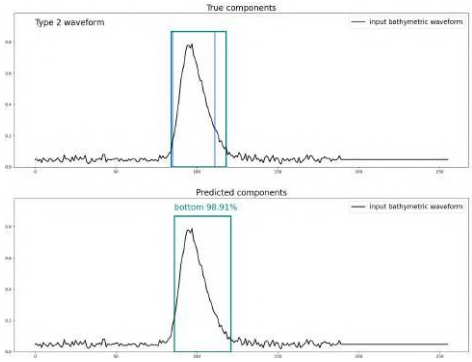
Typical errors

Several types of mistakes can be found by analyzing qualitatively the test predictions. The main types of errors are the following:

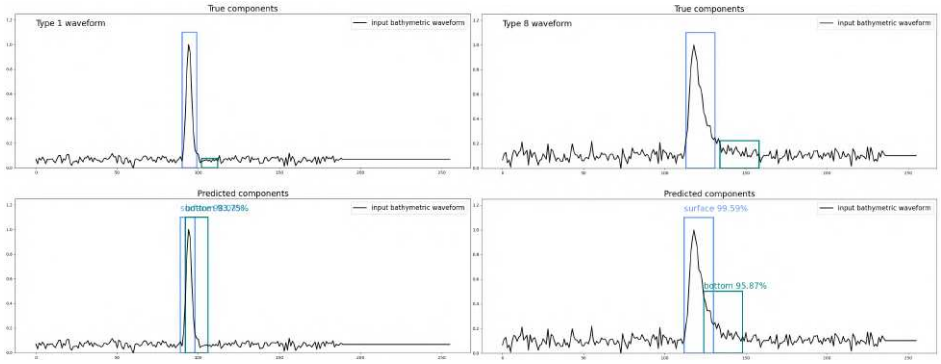
- **Overestimated or underestimated component widths:** the most common mistake is in the estimation of the component width, illustrated below. This is expected, since the mIoU_d metric does not reach 100%. The width is more frequently under-estimated than it is over-estimated. This error is also not symmetric: the network tends to underestimate the tail of the components more than their rising time.



- **Un-detected water surface in extremely shallow areas:** visually, we observe that overlapping components with large widths tend to result in a clean detection of the water bottom, but sometimes at the expense of the water surface (see illustration below). It should be noted that it is much less frequent on narrower components.

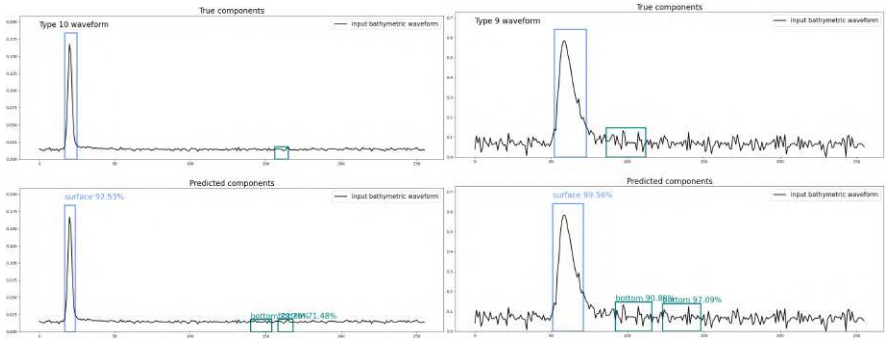


- **Mislocated water bottom echo with a low IoU with the true component**: another common source of error is the mislocation of the seabed return. We observe that most of the time, the detected component intersects the true box, but is largely offset and thus does not overcome the IoU threshold used, as featured in the following figure.

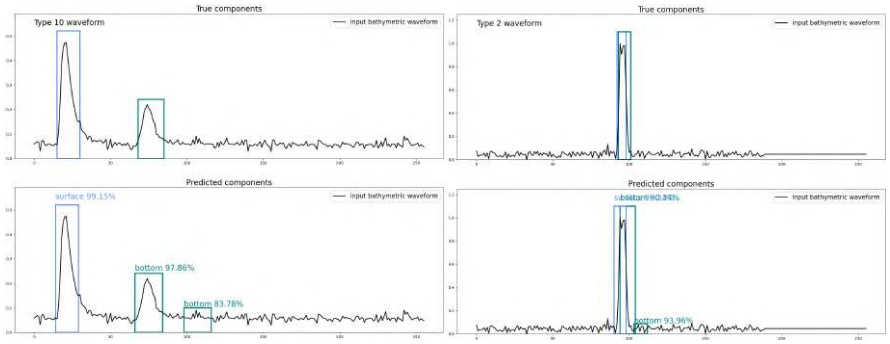


- **False supplementary detection of water bottom, but annotated seabed correctly detected:** in some cases, correct detections of seabed components are accompanied by false positives close to them. We distinguish two subcases.

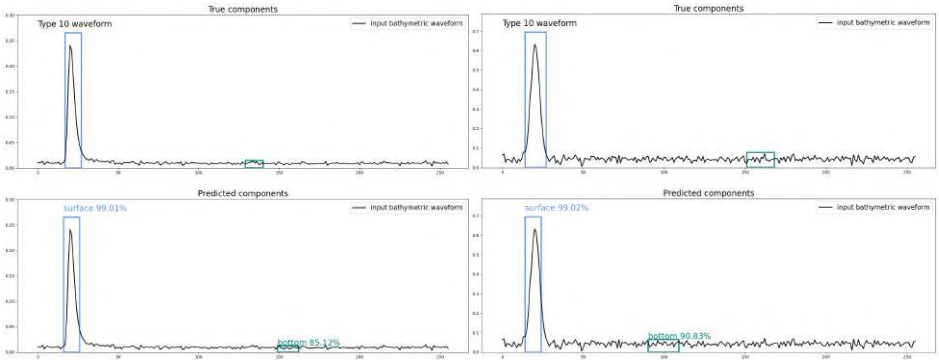
  – When only **weak seabed returns** exist: this error occurs more frequently when there are no strong seabed returns in the waveform, as illustrated below.



– We have also observed several cases of supplementary false detections even when a **strong seabed return** is detected, as illustrated below.



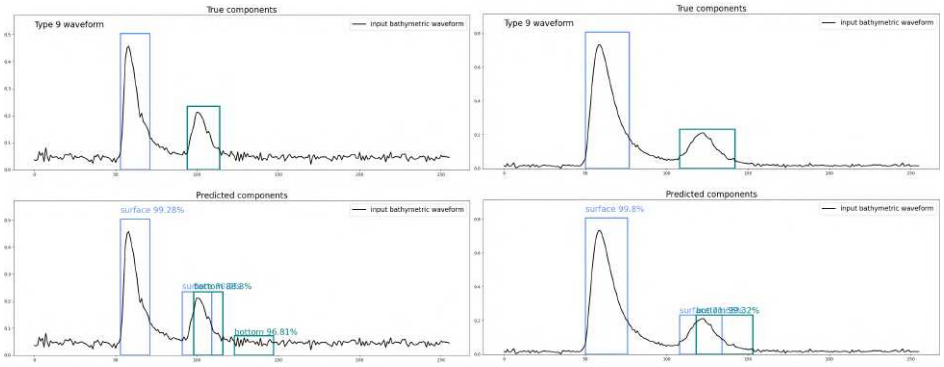
- **False detections of water bottom and missed annotated seabed:** sometimes, the seabed component is completely missed, and falsely predicted at another location in the waveform. It is mostly the case in deep waters, where the returns are very weak and may have similar features with the noise. Illustrations below show this type of errors.



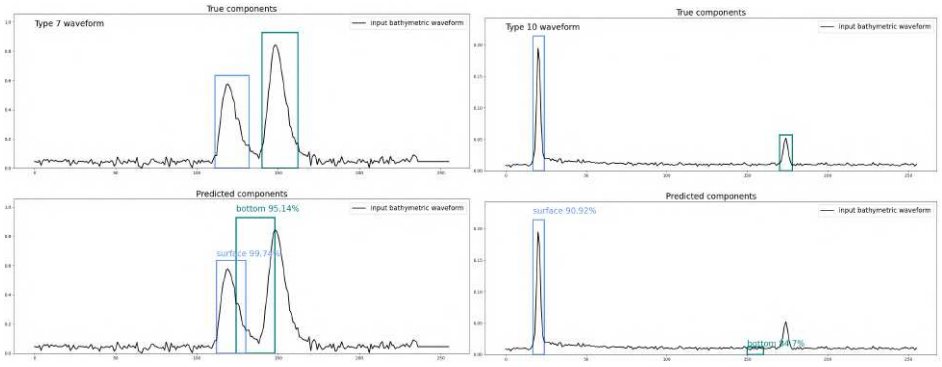
Rarely observed errors

Less frequent errors (observed once or twice over the 400 predictions visually checked) include:

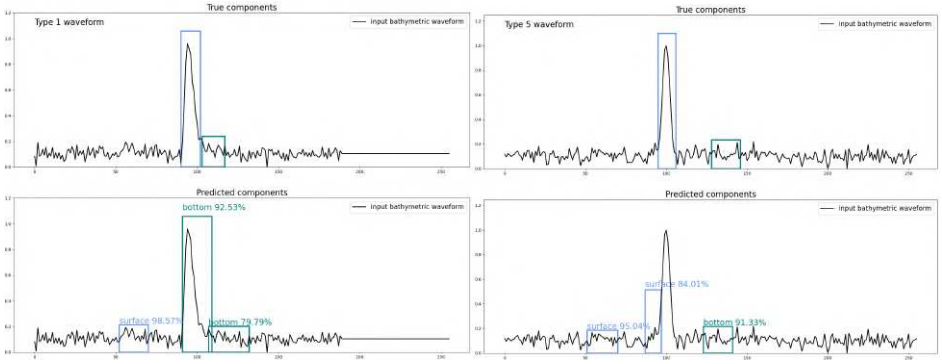
- **Seabed component mistaken for a mixed backscatter.** The examples below show large and prominent seabed returns mistaken for mixed surface/bottom components.



- **Prominent seabed return missed or mislocated.** We also observed the following predictions, in which strong water bottom components were missed.



- **Prominent surface echo mislocated and duplicated.** The last type of uncommon mistake is the prediction of the water surface in the noise that precedes actual information. It was not systematically paired with an error on the water bottom identification, as shown below.



These rare errors could be linked to the data augmentation procedures we used, which mixed different waveforms without taking into account spatial or temporal relevance.

Detection of the water bottom backscatter depending on its characteristics

Signal-to-noise ratio An interesting metric to evaluate our ability to predict weak water-bottom components is the signal-to-noise ratio (SNR). We used the following expression to compute it in dB:

$$\text{SNR} = 10\, log_{10}\left(\frac{P_{signal}}{P_{noise}}\right), \text{ where } P_{signal} = \frac{\sum_{c_0}^{c_1} \text{signal}^2}{c_1 - c_0 + 1}, \text{ and } P_{noise} = \frac{\sum_{c_0}^{c_1} \text{noise}^2}{c_1 - c_0 + 1}$$

As expected, when the noise is too strong compared to the signal, the network is not able to correctly predict the position of the water bottom component in the waveform. This is reflected by the percentage of detected components for very low SNR values, which is presented in Figure 5.9. In Figure 5.9, we compare the proportion of detected water bottom components depending on the SNR of these components for two IoU thresholds. The values presented are averaged over bins of SNR values increasing with 0.5 dB.

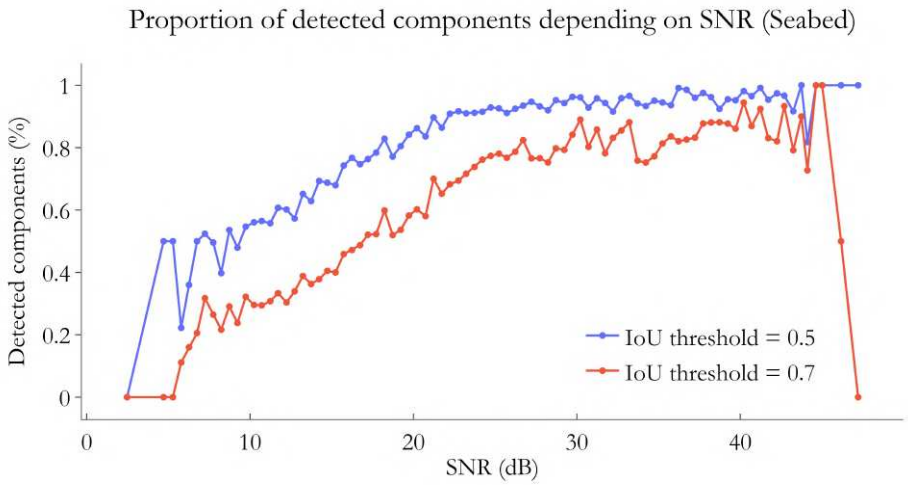


Figure 5.9: Percentage of water bottom components detected depending on component signal-to-noise ratio.

Overall, fewer components are detected for the higher IoU threshold, as already observed in Table 5.5. For the 0.5 threshold, we observe that the proportion of detected seabed returns converges close to 95% for SNR values above $\simeq$ 22 dB, where the signal power is approximately 100 times stronger than the noise power. However, half components are systematically detected for SNRs over 10 dB, and even for lower values close to 5 dB, although performance is less stable between 5 dB and 10 dB.

When using an IoU threshold of 0.7, the proportion of detected components has a less stable convergence, and varies around a lower level, which is around 80%. It also starts oscillating above 70% for higher

SNRs: $\simeq$ 22 dB versus $\simeq$ 15 db for the lower IoU threshold.

Component width Another parameter that may interfere with the network's ability to identify water bottom components is their width. As explained in chapter 1, waveform component width depends on several factors, including the geometry of the intercepted surface and the duration of the incident pulse. In water, depending on the turbidity and depth, the incident pulse can be stretched spatially and temporally. In real datasets, the network could thus be faced with varying component widths. Figure 5.10 illustrates the mean percentage of detected water bottom components depending on their true width. Widths were binned with a step of 5 and are expressed in waveform time steps.

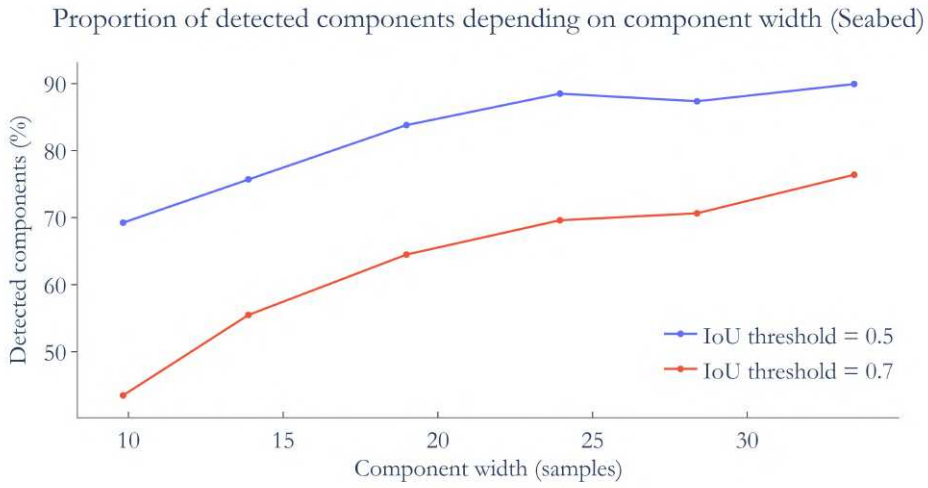


Figure 5.10: Percentage of water bottom components detected depending on component width.

Once again, we compare two IoU threshold values: 0.5 and 0.7. The stricter threshold of 0.7 has a similar trend than is 0.5 counterpart, but globally ranges between lower proportions ($\simeq$ 42% to $\simeq$ 75% versus 70% to 90% for the 0.5 thresold).

Globally, the proportion of detected components increases with their width. This is expected, as the main criterion to consider a prediction true is its IoU with the annotated component. Wider components have a greater chance to intersect a greater portion of the true component.

It is interesting to note that very short backscatters are already well detected with a 0.5 IoU threshold: 70% of the components with widths below 10 time steps (corresponding to 10 ns in our simulations) are correctly identified.

Component depth Lastly, we analyse the proportion of detected water bottom components depending on their depth. The results are presented in Figure 5.11, obtained by binning component depths with a step of 0.3 m.

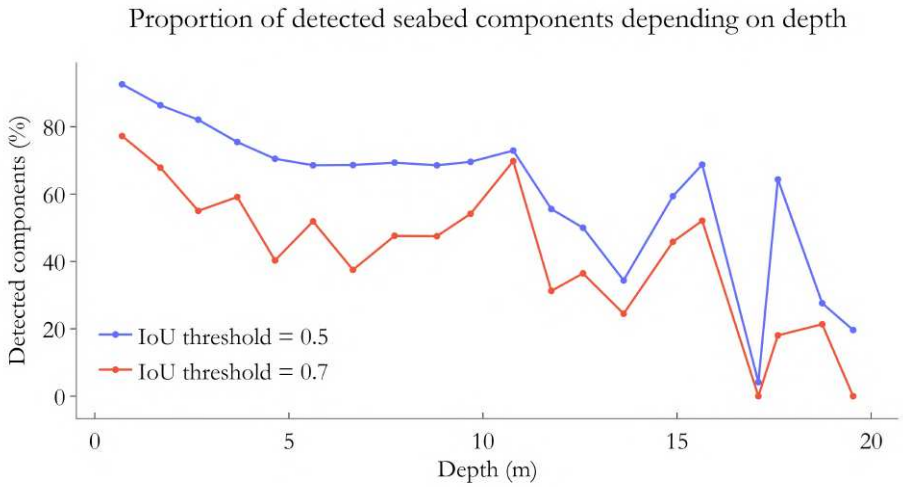


Figure 5.11: Percentage of water bottom components detected depending on component depth.

Similar to what we observed when analyzing the SNR (see Figure 5.9), the 0.7 IoU threshold shows less stability than the 0.5 threshold. However, in both cases, the proportion of components detected decreases with increasing depth. For depths over 10 m, the values fluctuate highly, but tend to feature lower values than at smaller depths, where they approach 90% for the 0.5 threshold.

It is interesting to note that, although there is a significant decrease with depth, the network still detects $\simeq$ 20 % of the components at depths of 19 m when using an IoU threshold of 0.5. Similarly, extremely shallow waters (with depths below 1 m) are detected in 79% of cases with an IoU threshold of 0.7 and in more than 85% of cases with a lower threshold. Yet, extremely shallow waters are particularly difficult to handle as they feature a large number of overlapping surface and bottom returns.
Our main results can be summed up as three key points:

- Excellent results are observed for the detection of the water surface component, although it is sometimes missed in extremely shallow waters;

- Overlapping water component results are detected with average precision and recall of 95%, while 70% of seabed returns having SNRs over 15 dB are detected, illustrating the capacity to locate weak

backscatters. These performances are seemingly not affected by the seabed geometry. Additionnaly, depths up to 19 m are handled, depending on local conditions.

- Globally, the network tends to not completely align the detected bounding boxes with the true components. Indeed, increasing the IoU threshold used for prediction validation results in performance decreases.

5.3.5 Application to real data

We applied the trained network to two different sets of real waveforms. Both were acquired East of Sables d'Or les Pins, the area studied in chapters 3 and 4. The first area contains shallow waters in which returns are easily detected in flight. The second, on the contrary, features deeper waters, for which seabed returns were not detected in flight. We present preliminary results of these first, uncompleted tests in the following paragraphs.

Direct application

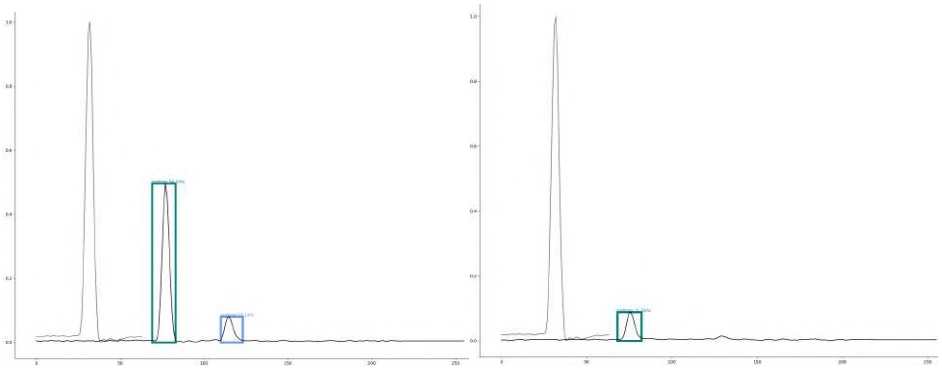
When processing real data, several differences must be taken into account. First, data need to be normalized as they were in the simulated data. To this end, both topographic and bathymetric inputs are normalized between 0 and the maximum power of both inputs. This enables us to maintain the relationships of relative reflectance between both waveforms, while providing the network with a reduced range of values.

We also observed that overall, the Titan waveforms start very close to the first return and can be short in shallow waters. This is explained by the recording mechanism of the sensor. To have more homogeneous lengths, instead of applying a significant zero padding at the right end of the waveforms, we also pad them on their left side by replicating the noise observed at the end of the signal. We found that it improved the results we obtained when applying the network to real data, and transformed data.

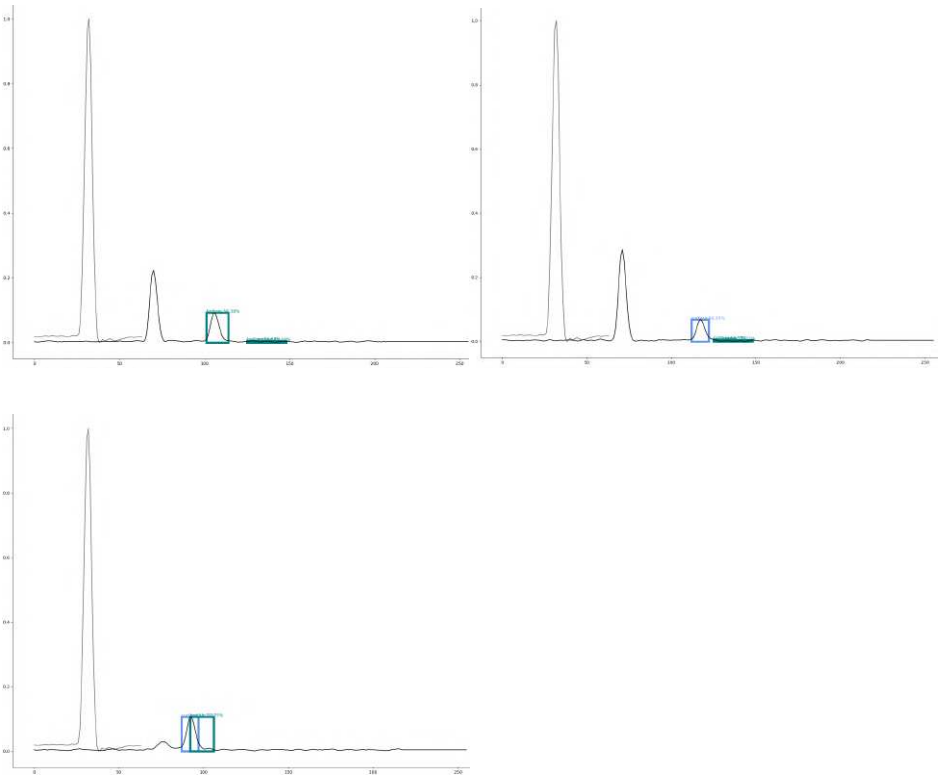
For prediction, we used a NMS threshold of 0.45 and an objectness threshold of 0.5.

When applying the network directly to the data, we observed poor detection performance. There were 3 main cases:

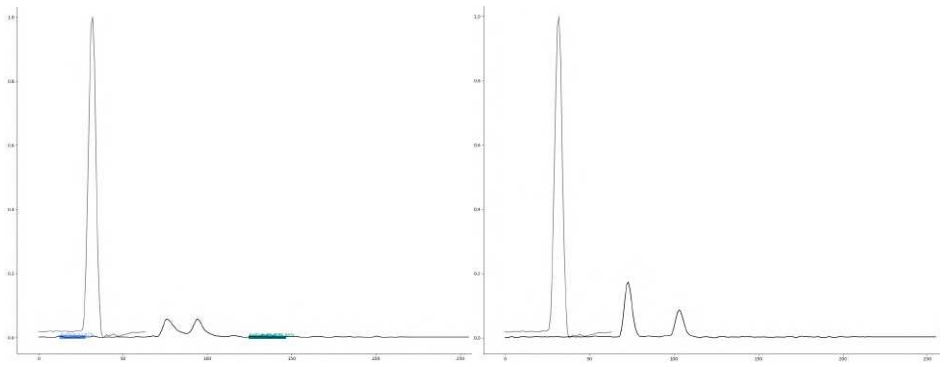
- At best, peaks were detected but mis-classified, as illustrated below.



- The water surface return is often ignored, even if it is prominent. In such cases, the water bottom component is either detected correctly - with additional false positives - or mistaken for the water surface. Sometimes, weak surface returns were ignored and seabed components were considered to be mixed surface and bottom returns. These three cases are illustrated below.



- Other cases featured completely irrelevant detections, or no detections at all even though prominent echoes were present, as the following examples illustrate.



Consequently, we decided to apply domain adaptation to better evaluate if these poor performances were linked to a too-big domain change or to the network itself.

Domain adaptation

We first used our complete simulated dataset to find the transformation between simulated and real data. However, this resulted in inaccurate results, such as concatenations of different types of waveforms with a much higher number of returns than the original waveform contained. Consequently, we restrained our OT computations to simulated waveform types consistent with each real dataset. In practice, real waveforms originating from shallow waters of Sables d'Or les Pins were transported in the domain of the type 7 waveforms (shallow moderately turbid waters, see Table 5.1). The second type of real waveforms - deeper waters - were transported into the domain of the type 9 waveforms (deep moderately turbid waters, see Table 5.1).

Using domain adaptation based on the Earth Mover's Distance (EMD), the results seem more relevant. Investigating the main changes between real and transported waveforms reveals the importance of reflectance domain changes. Indeed, it seems that the biggest change, that may explain poor performances with DA, is the much lower power of the real bathymetric waveforms relative to the topographic return used to estimate the incident pulse.

Most of the time, transported waveforms feature relevant detections, but with wrong classifications. However, the issue of irrelevant detections in noise and ignored prominent echoes is solved. There are also multiple very close detections of the same components frequently. This suggests the need to adapt the different thresholds used for prediction filtering (they may need to be stricter).

When transporting the predictions back into the domain of the simulated waveforms, we observe that surface returns are sometimes still missed, but fewer water bottom components are ignored.

Figure 5.12 illustrates the difference between the three steps - direct application, domain adaptation, and transport back to the original domain - for a shallow water waveform.

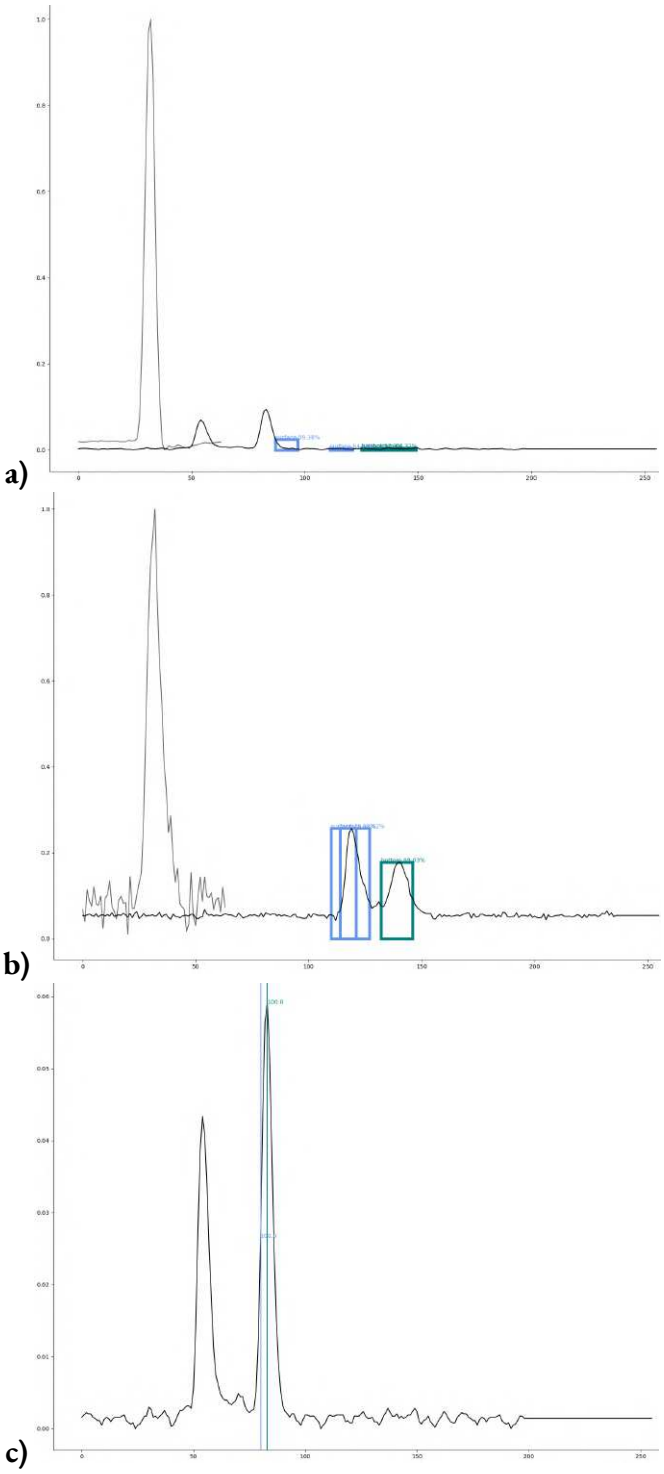


Figure 5.12: Illustration of our workflow for component detection in real waveforms: a) direct prediction, b) prediction on transported waveform c) original waveform with transported predictions. The dark black lines are the waveforms, in figures a) and b) lighter lines represent the associated topographic waveform.

Spatialization

To evaluate visually if the network makes spatially relevant predictions, we projected the detected components into a 3D point cloud using the beam firing vector of the original waveforms. This time, we applied an object threshold of 0.78 to see if multiple detections are discarded.

In the shallow area, the results show high spatial consistency, as Figure 5.13 illustrates. However, noisy over-detections persist in the water column and below the seabed.

Figure 5.13: Point cloud obtained with our detection method in coastal shallow waters. The point cloud is colored depending on the attributed class: green = seabed, blue = water surface.

It is also interesting to note that the correspondence score computed when transporting predictions back into the real waveforms domain can be used to filter out noisy detections in the water column. Figure 5.14 shows the result of the application of a 0.4 correspondence score threshold.

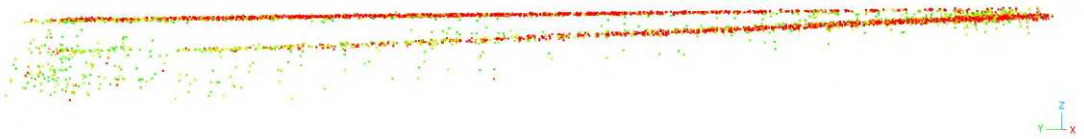


Figure 5.14: Filtered point cloud obtained with our detection method in coastal shallow waters. Only points with a correspondence confidence higher than 0.4 are kept. The points are coloured depending on their correspondence confidence; green symbolizes lower confidence than red.

On the second area, which features greater depths, the results also show spatial consistency, but the seabed is not detected. Figure 5.15 shows the resulting point cloud, obtained with an object threshold of 0.1 to maximise the number of detections included.

For this area, the detection of the water surface is particularly confident. Figure 5.16 shows the point cloud obtained by keeping only points with a correspondence score of 1.

Similarly to the first area, applying a correspondence score threshold helps to remove some of the noisy detections in the water column. However, even with a strict threshold, some of them persist.

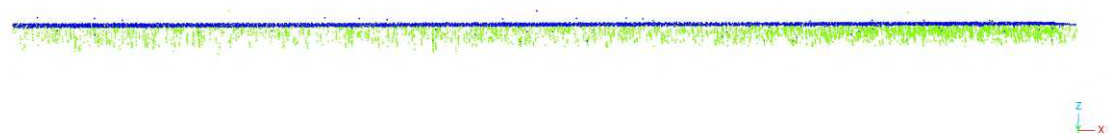


Figure 5.15: Point cloud obtained with our detection method in deeper coastal waters. The point cloud is colored depending on the attributed class: green = seabed, blue = water surface.

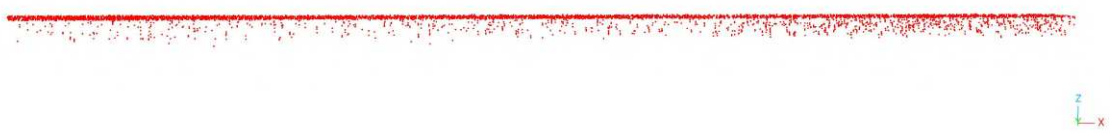


Figure 5.16: Filtered point cloud obtained with our detection method in deeper coastal waters. Only points with a correspondence confidence of 1 are kept.

This result suggests that further investigations around domain adaptation and the application to real data should be led to improve knowledge transfer on real datasets.

5.3.6 Discussion

In this section, we have investigated improved detection and identification of bathymetric waveform components using a deep neural network based on the YOLOv3 architecture (Redmon et al., 2018). Since our goal is to investigate if this methodological direction would be interesting to pursue, we started with an easy classification task including only two types of echoes: water surface and water bottom. Our preliminary results are promising: the network is robust to variations in environmental settings and seems to be applicable to real cases.

Detection performance

Although there are, at first, multiple duplicate detections of the water surface, they can be easily filtered using the class-specific score that the network outputs for each predicted bounding box. It is also interesting to note that there are few background noise detections. In our test dataset, we observe false detections of water bottom components in noise portions mostly when the seabed is difficult to identify. Such false positive are not systematic and they do not appear before the water surface return, which suggests that

the network does not randomly pick out noise as water bottom components.

One of the main strengths of the network is its ability to separate water surface and water bottom returns in extremely shallow waters. In some cases, the water surface is missed, but the detection of the seabed in depths below 1 m is strong ($\simeq$ 95% of average precision and recall). This is particularly promising for bathymetric waveform processing, as such settings are typically difficult to deal with.

The network also manages to identify very weak returns, as Tables 5.9, 5.8 and 5.10 illustrate. Analysing the SNR of the annotated components also shows that weak returns are well detected: above 15 dB, more than 70% of the annotations are correctly detected on average when using an IoU threshold of 0.5. This is promising, as it could increase the amount of lidar backscatters retrievable from bathymetric lidar surveys. Similarly, the network could improve backscatter detection in low albedo shallow water bottoms. Such settings are frequent in coastal areas, where rock can be covered in low albedo algae. Figure 1.7 in Chapter 3 illustrated how such configurations impacted topo-bathymetric surveys, resulting in wholes in the point clouds.

Lastly, the robustness of the network to varying water bottom geometries is another strength. Indeed, it has been shown that geometrical pulse stretching on the seabed leads to a bias in the estimated position of this component in waveforms (Bouhdaoui et al., 2014). Although we did not model all pulse stretching effects in our simulated data, this suggests that further developments could be interesting for the extraction of sloped seabed returns, for example.

Architecture choices and potential improvements

We chose to explore the contribution of object detection approaches with YOLO because it is one of the dominant architectures in this domain (Zhao et al., 2019). YOLO is mainly praised for its fast application, allowing real-time detection, which is interesting for massive dataset processing. In Redmon et al., 2018, the authors highlight the strong performances of YOLOv3 with IoU thresholds of 0.5, which are similar to other state-of-the-art methods, but faster. However, YOLO is globally known to struggle to get perfectly aligned bounding boxes, and thus have lower IoUs than two-step region proposal and classification frameworks (Redmon et al., 2018; Zhao et al., 2019). Practically this results in performance drops when increasing the IoU threshold used to validate predictions during testing. We observed similar phenomena when analysing the results of our network. Indeed, figures 5.9, 5.10, and 5.11, as well as Table 5.5 showed a significant performance decrease for water bottom detection when increasing the IoU threshold from 0.5 to 0.7. We also observed that globally the predicted bounding boxes were narrower than the annotated boxes.

Several improvements could be explored to tackle this issue. The one we consider the most promising is to adapt our network so it handles asymmetric components better. Indeed, when visually analyzing

the results, we observed that misestimations of component width were not similar for the peaks' rising times and tails. YOLO was built for object detection in images, in which getting aligned bounding boxes is more essential than perfectly estimating their centre (image components also do not necessarily have a native centre). However, for waveform detection, the component's centre is more important than the global bounding box, as it is used to determine the position of the backscatter. Extracting component width is mostly useful if component feature extraction is an objective. However, since the network performs simultaneous classification, this objective takes second place compared to the correct location of backscatters. Since waveform components are mostly asymmetric due to the incident pulses - it is commonly admitted that the pulse is more efficient if it has a short rising time, no matter is tail (Guenther et al., 2000) - the hypothesis made in YOLO that box width is symmetric relative to box centre is not valid for waveform processing. This may explain the offset we observed for some components. In the future, the architecture should thus be improved to predict a "left width" and a "right width" describing the spreading of the backscatter on each side of its location.

If this modification does not improve the results, a spatial-temporal approach taking into account neighborhing waveforms during encoding could be investigated. Indeed, spatial stacking of waveforms has the advantage of increasing the SNR, which is helpful for decomposition (Mader et al., 2023; Mader et al., 2019; Mader et al., 2021). Instead of stacking waveforms - which causes the spatial resolution of the derived products to be lower - we suggest using them for feature learning, but still predicting detections for each waveform.

In the meantime, our network can already be used as a primary detection that can be refined using other techniques if needed. In such a setting, it would have the advantage of reducing the search area for iterative decomposition methods. In the future, using a topographic waveform simulator (Gastellu-Etchegorry et al., 2016), it could be extended to a wider range of component classes, both topographic and bathymetric, to allow topo-bathymetric waveform processing on large extents. However, large-scale applications would first require improvements in the workflow used to apply the network to real datasets.

Application to real data

Our results highlight the contribution of physical models to the development of tools adapted to waveform processing: while we were not able to identify waveform returns in extremely shallow waters with U-time because they were not possible to label, we managed to handle these cases with this network. By simulating waveforms, we were able to train a network that is robust to different types of bathymetric environments. A key aspect is now to evaluate the possibility of transferring the knowledge acquired by the network to real data processing. Although we only started exploring the application to real waveforms, our preliminary results suggest that further developments are necessary.

Indeed, during testing, we were able to predict weak returns in moderately shallow waters with an average precision of 66%. However, when applying the same network to real data and using domain adaptation, we did not manage to retrieve weak seabed returns of a moderately turbid coastal area (see Figure 5.15). Furthermore, in very shallow waters, although the network localized correctly the water surface and the water bottom, it misclassified a large amount of seabed components (see Figure 5.13). Yet, during testing on simulated data, we did not observe such classification errors.

These observations suggest a problem of transfer learning. In these preliminary experiments, we investigated optimal transport between raw waveforms as a solution to domain adaptation. Other solutions to better apply networks trained on simulated data to real data could be investigated. For example, the network's encoder could be pre-trained in the manner of an autoencoder on real waveforms, before being injected with a very low learning rate into the optimization of the detection network. This could encourage the network to rely on features learned on real waveforms to derive prediction rules. Another possibility could also be to perform optimal transport between waveform representations rather than raw waveforms, similar to what is done in Damodaran et al., 2018.

If improving transfer learning does not solve the prediction errors at application time, one last possibility is to consider encoder architecture changes. We experimented with attention layers when building our YOLO-derived network. Although integrating it into the current architecture did not demonstrate particular improvements, further research on this side could be relevant. Indeed, the attention mechanism is useful to help the network focus on relevant features, by performing feature weighting based on similarity between layers. Multi-head attention could be useful to attend to different aspects of the input waveform and limit irrelevant predictions such as those evocated previously. If adding it to the current YOLO-based architecture does not work, replacing the convolutional encoder with a Transformer encoder (Vaswani et al., 2017) could be explored. This would allow us to better capture long-range dependencies in the waveforms and not only local interactions between samples, which could increase the overall performance and encourage the learning of more global features.

5.4 Bathymetric waveform inversion and environmental parameters estimation

In this chapter, we investigate the potential offered by deep neural networks for advanced lidar waveform exploitation when used in combination with physical models. With our YOLO-based network, we investigated the possibility of enhancing overlapping and weak echo detection in bathymetric waveforms to extract more information than we previously did. In this section, we explore another possibility offered by bathymetric lidar waveforms: the estimation of environmental parameters, including optical properties

of water bodies (WOP).

Deriving inherent optical properties (IOP) of water bodies from lidar waveforms has been a research subject since the advent of airborne lidar bathymetry (Guenther, 1985). Several solutions have been proposed, and all have shown that bathymetric waveforms could effectively be used to approach IOPs. Up to this day, most of the existing methods rely on waveform decomposition and the analysis of the water column component's characteristics or on regressions between waveform features and IOP or WOP measurements.

Independently on the method used, estimations of IOPs and WOPs published in the scientific literature rely on the access to measured samples of these variables. However, measuring WOPs is strenuous as it requires performing both in-situ measurements and airborne lidar acquisitions simultaneously, with a large enough number of punctual records of the WOP values over the studied area.

Some approaches exploit satellite imagery of such properties at different wavelengths to approximate them at the wavelength of the green laser used and rely on them for model fitting. For example, NASA's Moderate Resolution Imaging Spectroradiometer (MODIS) provides estimations of the diffuse attenuation coefficient at 490 nm - $K_{d_{490}}$. However, the products derived from MODIS are not adequate to study inland waters such as lakes and rivers, and their resolution - between 250 m and 1,000 m - makes them unsuitable for observing land-water interfaces. Using satellite-derived approximations of WOPs to train an estimation model would thus not fit our applicative objectives.

In this section, we thus aim to build a deep neural network for WOPs and IOPs estimation, as well as the identification of water surface and seabed positions from bathymetric waveforms. We wish to assess the possibility of estimating these parameters in one shot, without relying on in-situ measurements or satellite observation to perform regression.

5.4.1 Methodology

To estimate WOPs and IOPs, as well as depth, water surface and water bottom position, we designed a convolutional neural network. In total, we predict seven different parameters:

- The diffuse attenuation coefficient, K_d;
- The backscattering coefficient at $180°$, β_π;
- The loss factor at the air/water interface (which includes loss liked to the FOV), L_s;
- The water bottom reflectance, R_b;
- The position of the water surface, S;

- The position of the water bottom, B;
- The depth, D.

Based on the physical principles of lidar waveforms, we consider these parameters to be linked to the incident pulse and its angle of incidence on the water surface, which may cause geometrical pulse stretching. Consequently, our network takes the following parameters as inputs:

- The considered bathymetric waveform;
- The angle of incidence of the incident beam on the water surface, θ_b;
- A green waveform acquired at a low off-nadir angle on a plane, tar-like topographic surface;
- The angle of incidence at which this other waveform was acquired, θ_t;
- The reflectance of the ground it intercepted, ρ_t;

Network architecture

Our network contains two parallel convolutional encoders, which each encode one of the input waveforms, as shown in Figure 5.17. The convolutional blocks used in these encoders have the structure described in Section 5.3.2 and thus incorporate batch normalization, ReLu activation, max pooling, and residual connections. At each new layer, the number of convolution filters is doubled.

Each encoder applies four convolutional blocks to its input, halving the size of its outputs four times. After encoding, each feature map thus has a first dimension equal to 16. The feature maps resulting from each encoding are concatenated to form a new vector of representations with a first dimension of 32. This vector is passed through three convolutional blocks with a progressively reduced number of filters and no max pooling. The resulting vector is concatenated with θ_b, θ_t, and ρ_t.

The network then combines four separate decoders to favour the use of features relevant to each type of parameter:

- A convolutional decoder with ReLu activation to estimate K_d;
- A convolutional decoder with sigmoid activation to predict β_π;
- A second convolutional decoder with sigmoid activation to predict L_s and R_b;
- A third convolutional decoder to estimate S, B, and D.

All decoders consist of four convolutional blocks with one convolution filter, their respective activation function and max pooling, and a perceptron per parameter to predict.

The values of K_d, β_π, L_s, and R_b are directly predicted. However, S, B, and D are predicted as a proportion of the waveform's length. This choice results from an empirical observation that the network was harder to optimize when predicting S, B, and D directly as time indices.

Each output of the network has an individual loss. K_d, β_π, L_s, and R_b are optimized with a classical MAE loss. For S, B, and D, the MAE is computed on the network's output multiplied by the input size (256) to convert the predictions into time indices and compare them with the true values. The resulting architecture is illustrated in Figure 5.17.

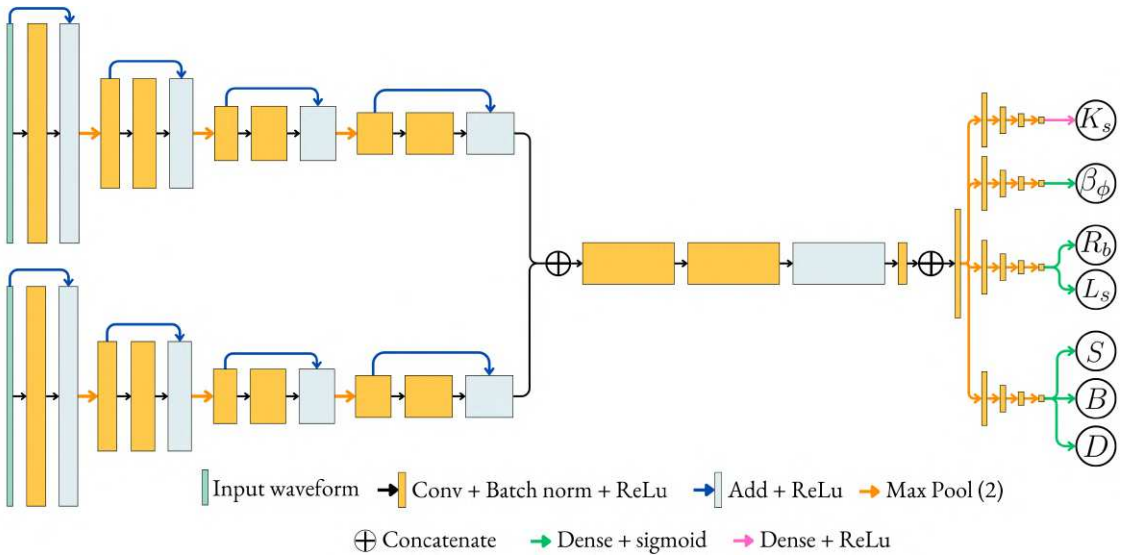


Figure 5.17: Architecture of the deep neural network used for environmental parameter estimation from bathymetric waveforms.

Optimization on simulated data

The network is optimized on a reduced set of waveform types. Indeed, to first assess the validity of our approach, we do not include extremely turbid waters (label 1 in Table 5.1) and extremely shallow waters, which are rarely used for parameter estimation (labels 2 and 6 in Table 5.1).

For each simulated bathymetric waveform, a topographic waveform was simulated on a plane ground with a reflectance of 0.05 and an off-nadir angle of $10°$ using the same simulated laser pulse.

The network is optimized with an Adam optimizer, mini-batches of size 1000, and a learning rate of 10^{-4} until it reaches its lowest validation loss.

	K_d	β_π	L_s	R_b	S	B	D
Mean error	-0.01 ±0.26 m^{-1}	0.0005 ±0.0021	0.002 ±0.095	0.04 ±0.02	0.001 ±0.185 m	0.007 ±1.31 m	-0.02 ±1.3 m

Table 5.11: Mean estimation error for each parameter.

Evaluation metrics

To evaluate our results, we analyze the prediction errors of each parameter. We also explore the variation of these errors depending on the waveform simulation parameters and analyze error distributions for each parameter.

5.4.2 Results obtained on simulated data

General performance

Table 5.11 gives a general overview of the performance of the network. Mean errors are relatively low, but most of the parameters present a high standard deviation, suggesting the need to perform a more thorough analysis.

Since the errors are differences between true values and predicted values, they must be considered depending on the typical range the parameter covers. For example, β_π has a much lower error than K_d, but considering β_π ranges from 0.0002 to 0.003 and K_d from 0.1 to 1, it does not reflect a better estimation performance.

Globally, K_d, S, and L_s are predicted with relatively low errors compared to their typical values in our data. However, β_π, R_b, B, and D seem to have less stable results.

Analyzing the distribution of the prediction errors for each parameter gives a more detailed idea of the estimation quality. Figure 5.18 contains histograms of the prediction errors for each variable investigated.

The histograms obtained for K_d, S, B, and D follow Normal laws centred close to 0 and with varying standard deviations. This is an interesting property as it means that error can be reduced by averaging spatially consistent predictions.

However, L_s has a more bimodal distribution, which is more asymmetric. The histograms of β_π and R_b confirm the lower quality of estimation we reach with the network. They cover large ranges of values and are multimodal.

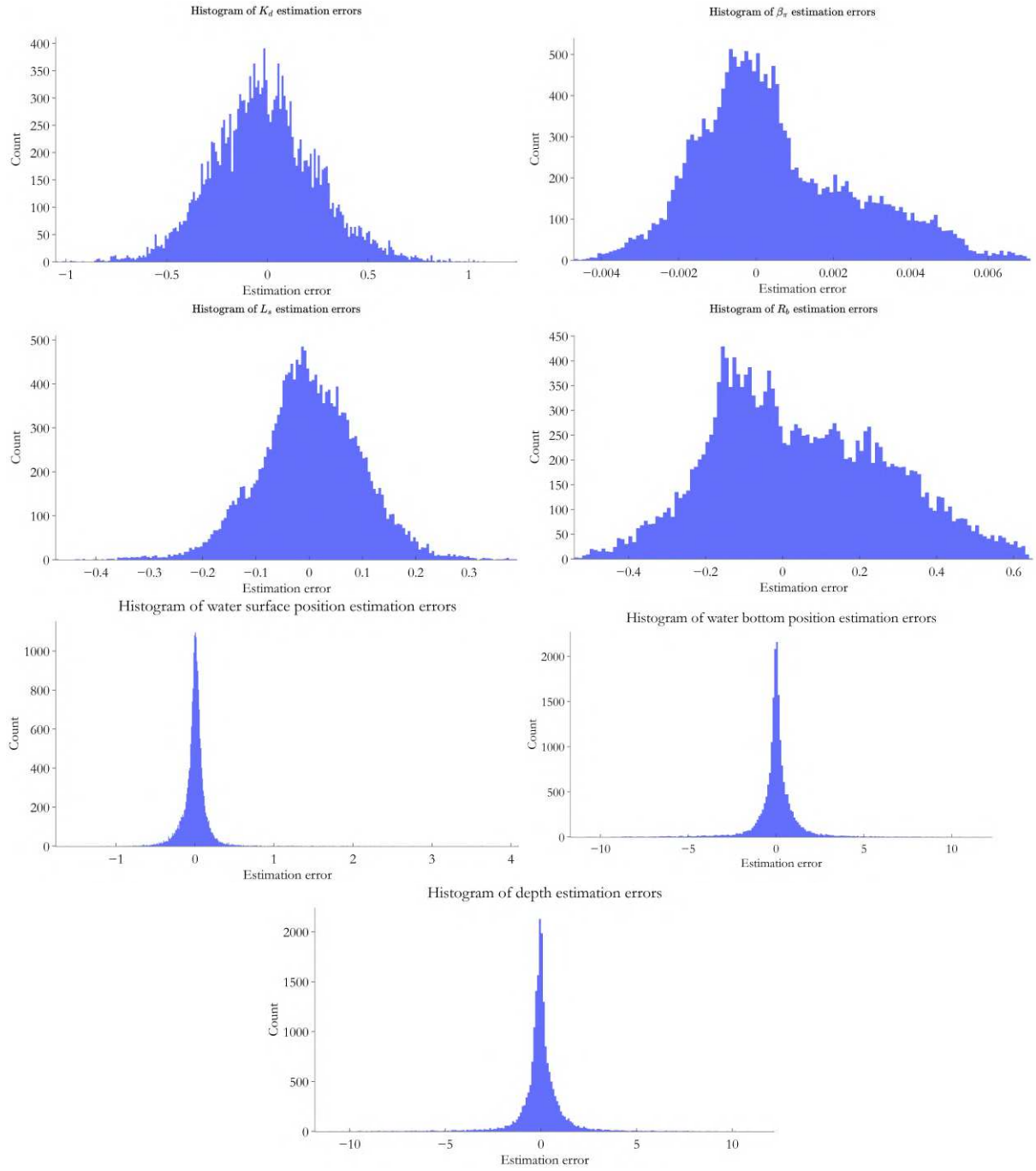


Figure 5.18: Distribution of the estimation error for each parameter.

Label	Mean error						
	K_d	β_π	L_s	R_b	S	B	D
3	0.17 ±0.23 m^{-1}	0.0027 ±0.0018	0.003 ±0.053	0.18 ±0.23	0.008 ±0.078 m	0.034 ±0.378 m	-0.011 ±0.38 m
4	0.25 ±0.26 m^{-1}	0.0022 ±0.0019	0.017 ±0.083	0.15 ±0.25	0.016 ±0.052 m	-0.115 ±0.597 m	-0.146 ±0.563 m
5	-0.05 ±0.23 m^{-1}	0.008 ±0.0017	-0.036 ±0.111	-0.02 ±0.23	-0.048 ±0.171 m	0.174 ±0.660 m	0.140 ±0.645 m
7	-0.04 ±0.20 m^{-1}	-0.0009 ±0.0009	-0.007 ±0.112	0.11 ±0.21	0.026 ±0.310 m	-0.009 ±0.453 m	-0.068 ±0.466 m
8	-0.20 ±0.24 m^{-1}	0.0014 ±0.0012	0.038 ±0.109	0.08 ±0.22	0.013 ±0.107 m	0.286 ±0.604 m	0.289 ±0.596 m
9	-0.002 ±0.19 m^{-1}	-0.0003 ±0.0009	-0.007 ±0.095	-0.03 ±0.24	-0.013 ±0.323 m	-0.164 ±1.673 m	0.008 ±1.53 m
10	-0.015 ±0.09 m^{-1}	-0.0001 ±0.0006	0.003 ±0.095	0.03 ±0.11	-0.001 ±0.121 m	-0.131 ±3.035 m	-0.277 ±3.11 m
11	-0.22 ±0.19 m^{-1}	0.0010 ±0.0023	0.005 ±0.069	-0.14 ±0.06	0.003 ±0.068 m	-0.019 ±0.307 m	-0.088 ±0.28 m

Table 5.12: Estimation error for each parameter depending on the type of bathymetric environment simulated.

Detailed error analysis

It is difficult to depict the processes behind estimation errors considering the interconnection of the estimated parameters. Consequently, to better assess the strengths and weaknesses of the network, we evaluate the estimation error for each variable depending on the parameters used for simulation. To this end, we analysed error distributions for each predicted variable depending on the value of the parameters used for simulation. We also compare prediction errors depending on the type of waveform simulated. Labels 3, 4, and 5 correspond to turbid waters. Labels 7, 8, and 9 represent clear to moderately turbid waters. Label 10 contains waveforms simulated for great depths and clear waters, and label 11 represents shallow waters with low albedo seabeds. Table 5.12 presents the metrics obtained for each parameter depending on the waveform type (label). Lastly, we also explore the impact of the impulse function used and the type and level of noise added. Table 5.13 summarizes the error for each parameter depending on the impulse function used to generate the waveform. The following paragraphs concentrate on the observations made after analysing all these elements.

Globally, the estimations seem robust to variations of bottom geometry, emitted power, emitted pulse

Impulsion	Mean error						
	K_d	β_π	L_s	R_b	S	B	D
Gaussian	-0.02 ±0.26 m^{-1}	0.0003 ±0.0020	-0.002 ±0.089	0.04 ±0.23	0.0003 ±0.183 m	0.065 ±1.387 m	0.026 ±1.40 m
EV	-0.01 ±0.25 m^{-1}	0.0003 ±0.0020	0.01 ±0.095	0.04 ±0.23	-0.004 ±0.174 m	0.004 ±1.288 m	-0.020 ±1.251 m
GEV	-0.003 ±0.25 m^{-1}	0.0004 ±0.0020	0.01 ±0.094	0.04 ±0.23	-0.014 ±0.161 m	-0.017 ±1.289 m	-0.035 ±1.277 m
Log-norm.	-0.01 ±0.26 m^{-1}	-0.0009 ±0.0022	-0.01 ±0.10	0.06 ±0.23	0.020 ±0.214 m	-0.023 ±1.279 m	-0.047 ±1.309 m

Table 5.13: Estimation error for each parameter depending on the impulse function used to simulate the waveform.

duration and β_π variations. Indeed, the mean error and the standard deviations of the errors did not vary significantly with increasing or decreasing emitted power, pulse duration or bottom component width.

We also observe that the type of pulse used for simulation does not impact significantly the estimation performance. The error is slightly higher for log-normal impulses, which are more asymmetric. Similarly, the network seems to be robust to different noise types and levels. The error distributions for both types of noise and the three different levels used are very similar. We observe a slightly higher error for high noise, for which the error distribution spreads on lightly wider values.

In the following paragraphs, we recap the main error tendencies for each estimated parameter.

Estimation of β_π Globally, the network is not very good at predicting β_π. Low values are over-estimated while the larger ones are under-estimated. The error decreases for medium values, typically between 0.002 and 0.004, and for deep clear waters (label 10), for which the error standard deviation is also lower. However, prediction error is its highest in turbid waters.

Estimation of K_d K_d is predicted more accurately, but some trends can still be observed. For this parameter too, it seems that the network is worse at predicting extreme values - very low and very high K_d. In shallow waters with dark seabeds, K_d is systematically overestimated. On the contrary, it is underestimated in turbid waters. In moderately turbid waters, when L_s is over $\simeq$0.4, K_d is increasingly overestimated. Similarly, in shallow waters with dark seabeds and moderately turbid waters, lower reflectance values are associated with a higher overestimation of K_d. Globally, error standard deviations are larger for moderately to very turbid waters. There also seems to be an impact of depth: the error on K_d is lower past $\simeq$ 4 m of depth. However, it is difficult to explain such dynamics, as they can be linked to other

parameters affecting the waveforms simulated at lower depths.

Estimation of L_s Contrary to K_d, there is no clear impact of depth on the estimation of L_s, except for an increase in error standard deviation for very deep clear waters. However, when the incidence angle on the water surface increases, L_s estimation errors tend to increase for all types of water. L_s is then mostly over-estimated. While no significant variations are observed with increasing K_d, in very deep clear waters, the L_s error significantly increases with increasing L_s.

Globally, for other types of waters, low L_s factors are overestimated and high L_s values are underestimated - similarly to K_d, the network does not handle extremes very well.

Estimation of R_b Reflectance is the second parameter - with β_π - for which estimation error is high. Reflectance is systematically overestimated in shallow waters with dark seabeds and tends to be underestimated for other types of waters. However, globally, the error is lower for clear and moderately turbid waters than for turbid ones. Very turbid waters are characterized by high error variations, demonstrating unstable performances. The error standard deviation is much higher for such waters than for less turbid ones. For these types of waveforms, the underestimation of R_b increases with the value of K_d, illustrating more difficulty in estimating reflectance when turbidity is high.

It is interesting to note that in shallow waters with dark seabeds, the overestimation of R_b decreases with L_s up to $\simeq 0.4$. This suggests that water surfaces letting most of the energy through are not characterized correctly, which results in an error in the estimation of R_b.

Lastly, the estimation error is more significant for higher reflectances. It tends to stabilize for reflectances between 0.2 and 0.5, but features the same increasing error among extremes than observed for the other parameters.

Estimation of the position of the water surface There are no particular correlations between β_π, K_d or depth values and errors in water surface positioning. Similarly, the incidence angle on the water surface does not impact the ability of the network to position it. However, the standard deviation of surface positioning errors is much higher for moderately turbid waters and increases with increasing reflectances. However, in moderately and very turbid waters, this standard deviation decreases with increasing L_s.

Estimation of the position of the water bottom As expected, bottom positioning errors are highest for deep moderately turbid water and for very deep waters (label 10, where depths reach 20 m). For the latter category, the error increases with water turbidity. However, the network seems robust to varying L_s, reflectances, or water bottom geometries.

Estimation of the depth Similarly to what we observed on water bottom positioning, the highest errors and the highest error variability in the depth estimations are reached for very deep clear waters and more globally for weak seabed returns. The variability of the depth estimation error increases in waters with higher K_d. For every category except shallow waters with dark seabeds, depth tends to be over-estimated - for very deep clear and very turbid waters - or under-estimated - moderately turbid waters - with increasing K_d.

However, for depth estimation, too, the network seems robust to varying reflectances and water bottom geometries.

5.4.3 Illustration of typical results

Tables 5.14, 5.15, and 5.16 provide visual overviews of the results obtained. For each type of setting, illustrations of typical results for different waveforms are presented. They show the high quality of the water surface and water bottom positioning, including in challenging settings. They also illustrate the ability of the network to capture physical relationships between parameters in some cases, while struggling more with others. For example, type 8 waveforms, representing waters with a high loss of energy at the air/water interface, feature a much higher estimation error for R_b and β_π than type 7 waveforms (see Table 5.15).

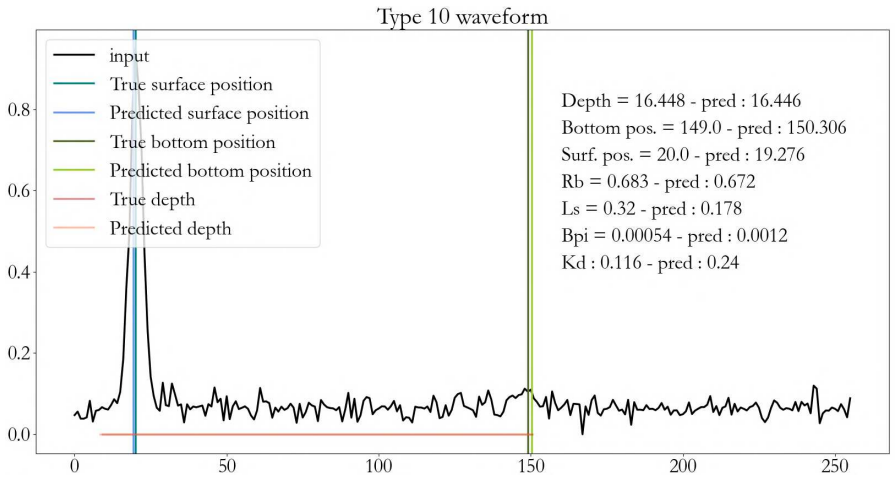
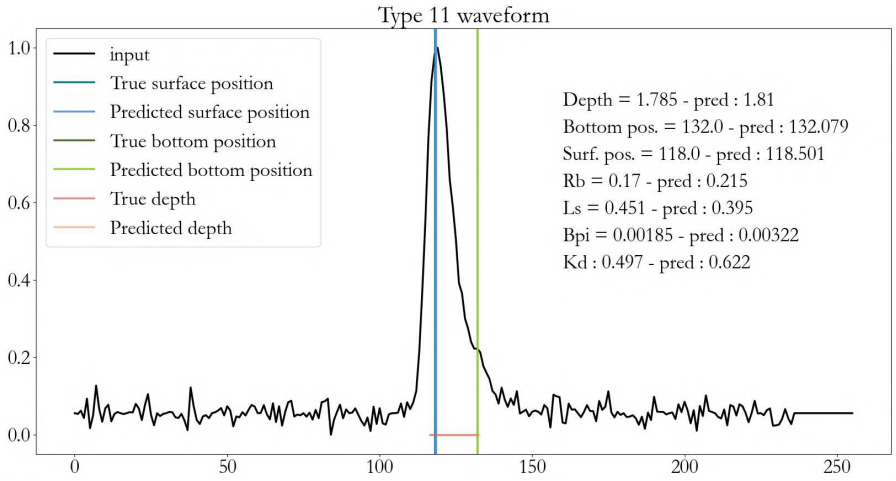
Very deep clear waters

Shallow waters with dark seabeds


Table 5.14: Results of the estimation network in very deep clear waters and shallow waters with low albedo seabeds.

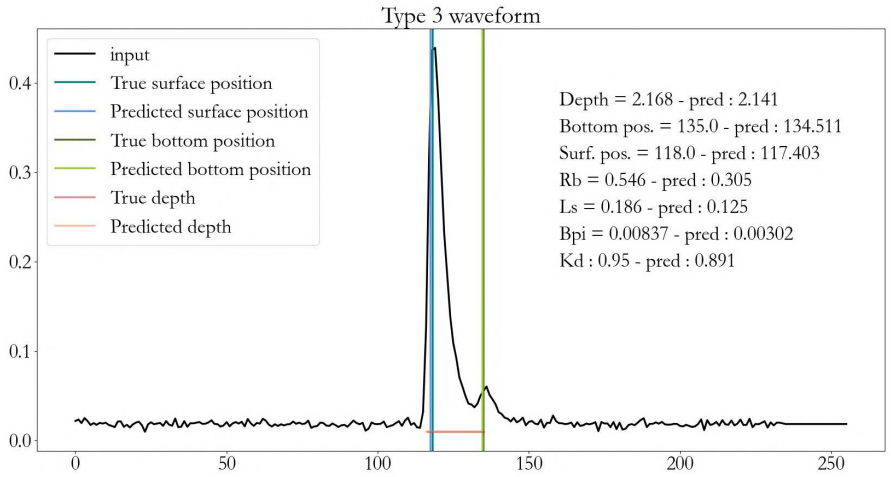
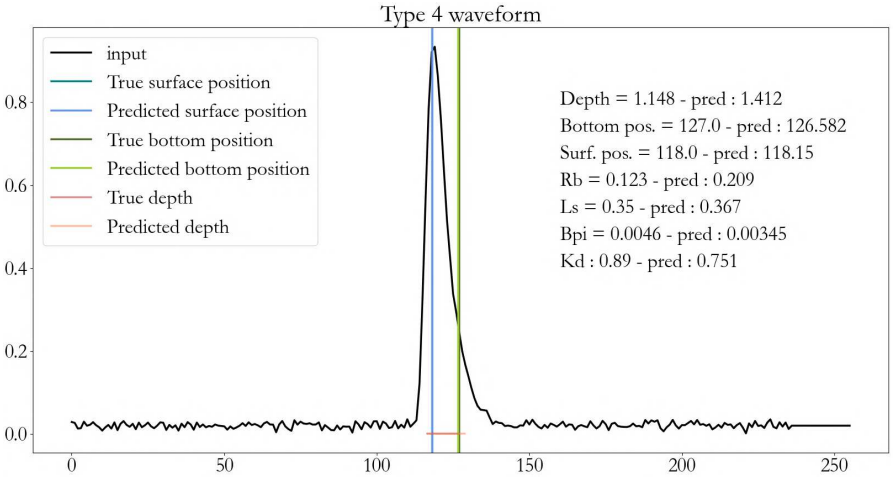
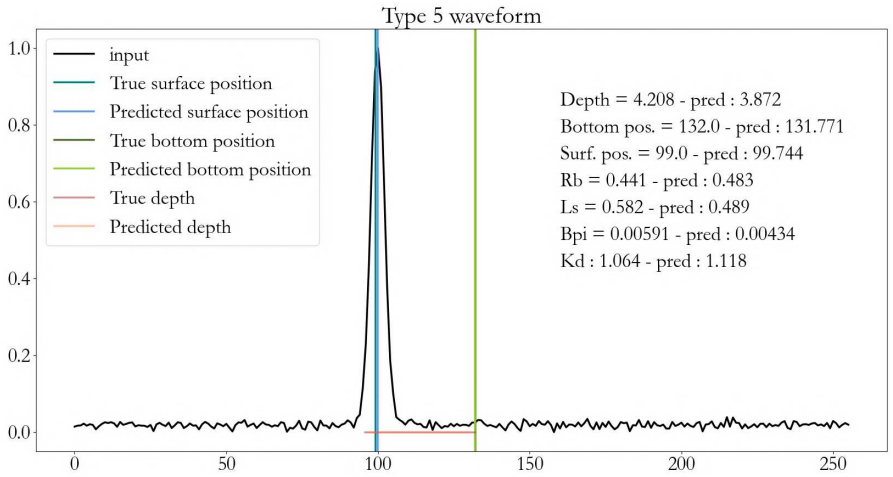
Very turbid waters




Table 5.15: Results of the estimation network in moderately turbid waters.

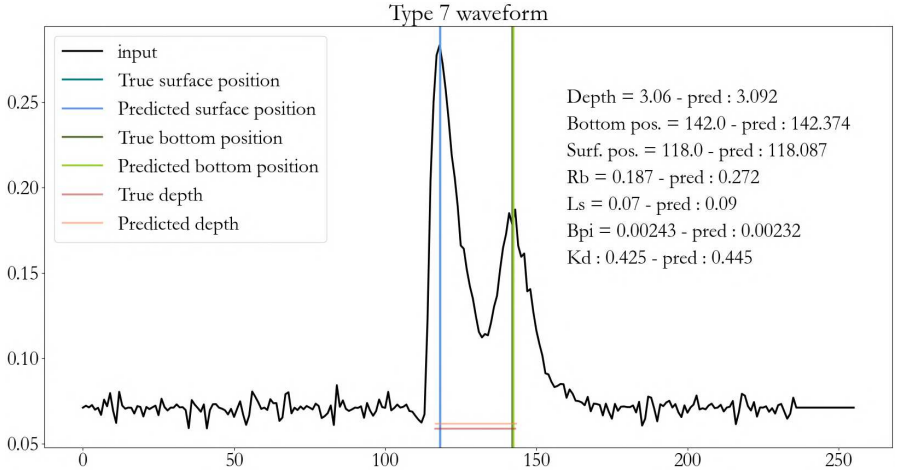
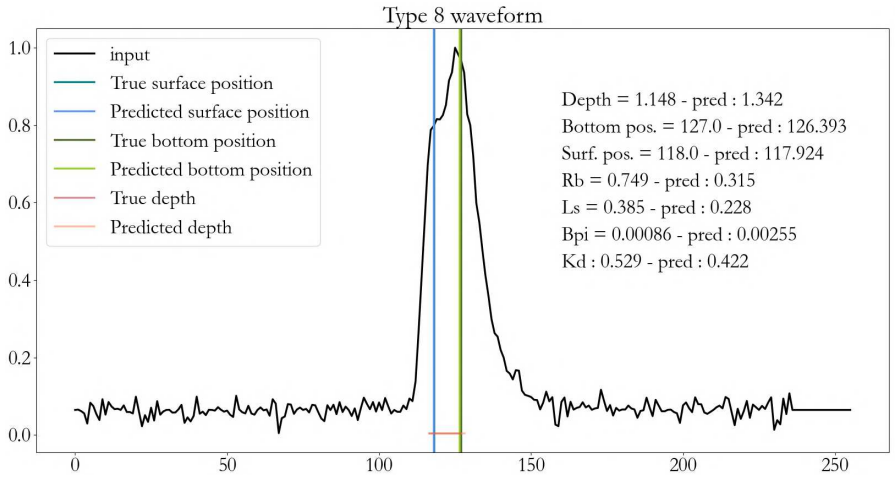
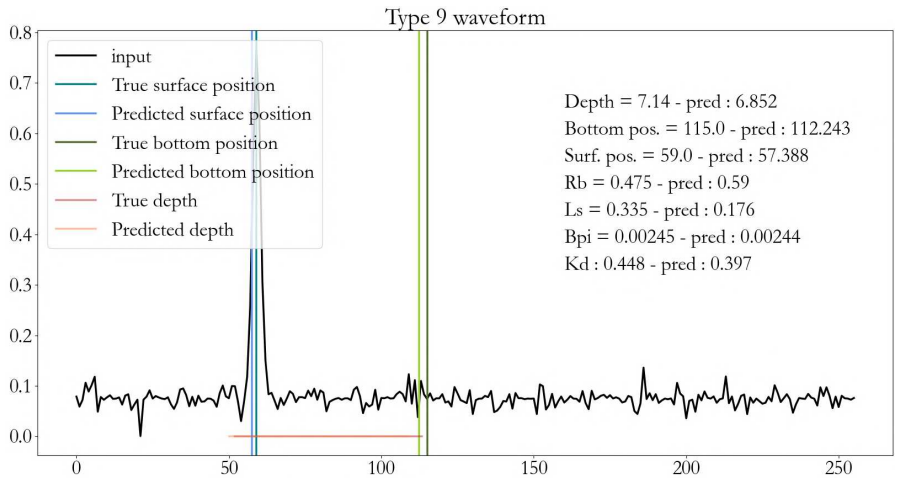
Clear to moderately turbid waters
Type 7 waveform 
Type 8 waveform 
Type 9 waveform 

Table 5.16: Results of the estimation network in very turbid waters.

5.4.4 Discussion

In Section 5.4, we have proposed a convolutional network dedicated to the estimation of water optical properties, and of water surface and bottom positions. The results appear promising, namely for water surface and bottom positioning, depth estimation, and K_d evaluations. Indeed, as illustrated in Tables 5.14, 5.15, and 5.16, our network predicts very precisely the position of the water surface and water bottom in challenging cases such as overlapping echoes of very shallow waters, and weak returns in deeper or turbid environments. These results are particularly interesting relative to the literature on this subject. Indeed, often, the range of parameters estimated in one shot is much lower, mostly including turbidity through K_d and sometimes depth (Richter et al., 2022; Lu et al., 2023; Montes-Hugo et al., 2016). Furthermore, we have seen in Chapter 2 that most of the approximations of K_d are based on the possibility to locate the water column, and thus the seabed (Lu et al., 2023; Feygels et al., 2003; Richter et al., 2021). When this return is mixed with the water surface or too weak, such methods can thus not perform. In most cases, the estimated K_d is also closer to a sensor-related parameter, linked to the system's FOV (Guenther, 1985; Lu et al., 2023; Feygels et al., 2003). However, in this study, by having access to the value of K_d, we train the network to estimate K_d and not K_{lidar} as in Lu et al., 2023 and incorporate the loss linked to narrow FOVs in the L_s parameter. However, our results illustrate the complexity of handling the interconnections between the physical parameters shaping the bathymetric waveform. For example, the network struggles with the relationship between loss of energy at the air/water interface and seabed reflectance. It could indicate that some combinations are not represented enough, such as low K_d with very low seabed reflectances. However, it also indicates that our network potentially does not learn enough representations to be able to capture such fine interconnections.

Besides complex physical interconnections, our network also struggles to estimate the extremes of the parameter distributions. For example, it overestimates low values of K_d and underestimates high values of K_d. The precision is also different depending on the parameters. For example, β_π, which is characterized by particularly small values, is estimated with greater error and instability than surface position, seabed position, L_s, K_d or depth.

Its ability to locate the position of the water bottom is however quite surprising. Indeed, although uncertainty and error increase in very deep waters, the network manages to locate bathymetric components relatively well. For example, in very turbid waters with depths up to 10 m, the mean error in the bottom position is $\simeq$ 0.17 m, with a standard deviation of $\simeq$ 0.66 m.

The low error on water surface positioning (see Table 5.12 and the histograms in Figure 5.18) is also promising. Usually, accurately positioning the air/water interface is a challenge due to complex interactions between the green light and the upper layer of the water column (Guenther, 1985; Mandlburger et al., 2013). Consequently, positioning precisely the water surface requires an additional NIR wavelength

in order to derive accurate depths with green laser soundings. However, our results suggest that it is possible to derive air/water interface position without relying on a second wavelength, which is promising for the development of lighter sensors, that could be embedded in uninhabited vehicles.

Lastly, the robustness of the network to different sensor configurations and seabed geometries illustrates the capacity of convolutional networks to solve complex problems. Indeed, it would have been expected to have a larger rate of error in parameter estimation depending on noise or pulse stretching (Bouhdaoui et al., 2014), namely for components positioning, noise being a limiting factor of decomposition and deconvolution approaches.

To improve these preliminary results, several strategies are possible. The first is related to the network architecture. Struggles to learn interconnections between some parameters may be a symptom of a too-simple architecture. It is possible that learning more features would help deal with some of the limitations observed. The depth of the convolutional encoder could thus be increased. It could also be modified with the principles of Inception networks (Szegedy et al., 2015), which alternate different convolution kernel sizes. It has been shown in Pelletier et al., 2019 that the size of the convolution kernel has an impact on the information extracted in applications to sequential data. Thus, relying on different receptive fields could be a strategy.

We also suggest training the network in several steps. We made the choice to exploit 4 different components after realizing it helps the network perform better overall. With this setting, training the different decoders separately successively could lead to increased performances of each prediction layer. Indeed, with a pre-trained encoder, four successive optimization phases concentrating on one decoder at a time could favour the learning of features and prediction rules more adapted to each task.

The use of the attention mechanism could also be helpful. In preliminary works about waveform decomposition with autoencoders (Letard et al., 2023), we have observed that attentive convolutional encoders work well for bathymetric waveform processing. Using multi-head attention could maybe help capture relations between surface loss and seabed reflectance estimation, for example. Moreover, when experimenting with Transformers for waveform decomposition, we got positive results in terms of optimization and results (these experiments require further work to derive more rigorous results). Changing the structure of the encoder for an attention-based one is thus another possibility.

Histograms also suggest that parameter estimation could be improved by considering neighbouring waveforms simultaneously. Indeed, parameters presenting error distributions following Normal laws centred on 0 could be predicted with reduced error by averaging information spatially. Furthermore, similarly to the approaches relying on orthowaveforms (Richter et al., 2021; Richter et al., 2017; Richter et al., 2022), feature learning may be easier on increased SNR. We thus suggest, as for the YOLO-based detection of components, to implement an encoder taking into account adjacent waveforms as contextual

information.

Exploiting spatial context could also be useful to obtain spatially consistent estimations of WOPs and IOPs. In the near future, we wish to further explore the application of the network to real data. Namely, we would like to compare estimated values of K_d to other estimations obtained with regressions between depth and received power.

Preliminary experiments of the application of the network to real data show similar results to those obtained for the YOLO-based network. Concretely, in direct applications, the network performs poorly in the detection of the water surface and the water bottom. They are often missed or located in noise portions of the waveform. Application to real data should be further explored, as the domain adaptation employed does not seem to be optimal. Indeed, the performance of the network in terms of water surface and bottom positioning did not transfer to real data: when projecting the detections into a point cloud, all detected components were systematically located above the water surface.

Considering our network's ability to detect water surface and water bottom positions, we are exploring the extension of the approach to global waveform inversion. We are currently investigating the possibility of adding other decoders to estimate the incident pulse and the water bottom geometry. In our preliminary experiments, we observe promising results for pulse and seabed geometry estimation when they are performed separately from physical parameters estimation. Combining both tasks results in a global performance decrease for all predictions, confirming our intuition that the network should be optimized differently to better face the complexity of the task.

5.5 Conclusion

In this Chapter, we have experimented with physical models for waveform simulation to build advanced processing tools. In section 5.3, we proposed to use an object detection and identification framework to locate and classify bathymetric waveform components simultaneously. Our results show the possibility of relying on this method to separate water surface from water bottom backscatters in extremely shallow waters with high precision and recall ($\simeq$ 95%). This approach also detects seabed returns up to 19 m depending on the local conditions. In practice, the network detects 70% of water bottom components with signal-to-noise ratios as low as 15 dB. Through simulation of various bathymetric environments and acquisition settings, we obtain a method that is robust to changes in seabed geometry and is not limited by noise. Relatively to existing methods, the network has the advantage of performing in very challenging settings without relying on multiple iterations or assumptions on the nature of the signal.

We also explored the transfer of knowledge learned from simulated data to real waveforms. Using optimal transport-based domain adaptation, we show that the network successfully applies to shallow

coastal waters. However, in deeper waters, our domain adaptation approach seems to reach its limits, as the seabed is not detected, while it was identified in similar simulated environments.

In the second half of the chapter (section 5.4), we investigated water optical parameters and depth estimation using a convolutional network with four separate decoders. Our method is able to predict seven parameters simultaneously in one single step. The results are promising, but also show the need to further improve the architecture. Indeed, some physical interconnections between parameters are not captured by the network, resulting in difficulties in estimating some parameters. However, the network predicts the position of the water surface with high precision and manages to locate the water bottom in challenging conditions (overlapping returns and weak echoes except for very deep waters) in one shot. Although its applicability to real data should be further evaluated, the network could be useful to position the water surface without relying on a second wavelength in the NIR spectrum.

In the end, physical models enabled to develop advanced processing methods, handling challenging settings such as highly overlapping very weak echoes. Without data simulation, it would have been impossible to label data to train our neural networks. Preliminary results suggest the transferability of the approach to real waveforms. This is of great potential as neural networks have the advantage of being more straightforward than classical methods, as they do not rely on iterative procedures and they can handle noise or mixed components.

GENERAL CONCLUSION & PERSPECTIVES

Contributions

In this thesis, we proposed **a set of methods dedicated to the extraction of environmental knowledge from topo-bathymetric lidar surveys**. This work is motivated by the **growing need for land-water interface observation methods**, as these ecologically rich areas are increasingly threatened by natural and anthropic pressure. Our main assumption is that topo-bathymetric lidar remote sensing has the potential to address this need, by delivering adapted measurements of the characteristics of land-water interfaces. By leveraging dual-wavelength laser telemetry, this sensor captures topo-bathymetric areas without interruption at the waterline, in the form of 3D point clouds and waveforms. The spatial patterns in the 3D point clouds provide rich information about the geometrical features of the surveyed areas. Waveforms provide complementary spectral information, that can be used to derive more backscatters from the environment, or physical traits of the intercepted objects. However, **adapted tools to exploit these data at their full potential are still expected**. Consequently, we investigated the solutions offered by processing methods that emerged since the advent of topo-bathymetric lidars.

Our main contributions to topo-bathymetric lidar remote sensing can be summarized as follows:

- **An open-source framework for bi-spectral point clouds classification accessible to thematic specialists with no expert knowledge of data processing and machine learning:** in Chapter 2, we highlighted the lack of solutions to classify bi-spectral point clouds. In Chapter 3, we thus introduced 3DMASC, a bi-spectral point cloud classification method that relies on tailored features and random forest models. Thanks to the associated feature and scale selection process, 3DMASC produces lightweight classifiers with high accuracy in topo-bathymetric environments such as rivers or coasts. The approach is fully explainable through feature importance assessments and SHAP analyses. It is also accessible to thematic specialists through an open-source plugin in the CloudCompare (Girardeau-Montaut, 2022) software. The classifiers require a limited amount

of labelled data and can be trained on common computers through a graphical interface or using the command line.

- **A bi-spectral lidar waveforms classification method for topo-bathymetric surface cover identification:** we also addressed the need for 3D classifications of land and seabed covers by developing an approach based on bi-spectral waveforms. Indeed, methods to identify surfaces from waveforms are scarce in topo-bathymetric contexts. However, waveforms contain rich spectral knowledge that allows the distinction of elements characterized by similar geometries. By extracting handcrafted features from green and near-infrared waveforms, we characterize surface covers of a coastal area at high spatial resolution. Our results highlight the contribution of bi-spectral information, and the ability to detect a wide range of classes with a single survey. However, as this method is based on classical machine learning, it requires a strenuous pre-processing step, that is not optimal in very shallow waters.

- **A waveform semantic segmentation approach for the classification of land and seabed covers:** to limit the need for pre-processing, which often implies site-specific decisions, we propose to use deep neural networks for semantic knowledge extraction from waveforms. To allow the direct identification of all covers present in lidar waveforms, we propose to perform semantic segmentation (i.e., 1 label per time step) instead of classification (1 label per waveform). Although we identify labelling-related limitations, our results overcome those we obtained with classical machine learning. The performances of the network suggest that learned features contain better information than handcrafted features: on green waveforms alone, the performances of our network are better than those we obtained with classical machine learning on bi-spectral waveform features.

- **A deep neural network for simultaneous bathymetric waveform component detection and classification:** with semantic segmentation, the labelling of overlapping backscatters is impossible. We thus go further in our exploitation of deep neural networks for waveform processing and propose to use an object detection framework to detect bathymetric lidar backscatters. To lift the limitations linked to data labelling, we train the network on a simulated dataset representing a large range of acquisition conditions. This allows us to have labelled returns in extremely shallow waters, where backscatters overlap and are sometimes impossible to spot with the naked eye. It also provides us with examples of very weak returns, another challenge of waveform processing difficult to tackle with manual labelling without introducing bias. This method produces promising results: it detects 95% of water bottom components in extremely shallow waters and 70% of components characterized by an SNR over 15 dB. Although some returns are missed in very deep waters, the application to the test dataset features many correct detections of very weak echoes. Lastly, the

transfer of the method to the detection of shallow coastal water components is successful. With this contribution, we also tackle the need for labelled data, which often constitutes a major drawback in knowledge extraction.

- **Preliminary elements on the estimation of water optical properties from lidar waveforms with no synchronous field measurements.** In light of our previous results with deep neural networks, we explore environmental parameter extraction with a convolutional neural network. We train it to predict the diffuse attenuation coefficient, the backscattering coefficient at 180°, the loss of energy at the air/water interface, and the seabed reflectance. We also estimate the water surface and the water bottom position, as well as the depth. Though the results are encouraging, they also feature estimation errors suggesting that this task requires a more complex network. Indeed, some interconnections between parameters, for example, loss at the water surface and prominence of the seabed return independently from its reflectance, are not captured. However, water surface and water bottom positions are particularly well predicted, except in greater depths. Several architecture and optimization changes could potentially improve these preliminary estimations.

- **To tackle the limitation of labelled data availability, we developed approaches using synthetic data and explored the possibility of applying them to real surveys.** To this end, we used physical models of radiation transfers in order to generate a set of labelled waveforms in various acquisition configurations. By doing so, we lift the need to manually label waveform components and have access to synchronous field measurements. Applying the obtained models to real waveforms requires domain adaptation, but preliminary experiments already make it feasible for water surface and water bottom detection in shallow coastal areas.

Last but not least, we also contribute to methodological improvements in the field of bi-spectral lidar remote sensing by providing the community with **open-source software and labelled datasets**[2] usable for point cloud classification.

Globally, with these methodological propositions, we highlight the **potential of deep neural networks to enhance information extraction from lidar waveforms**. We find their ability to model complex problems without requiring upstream feature engineering particularly suited to the challenges of waveform processing. It is also interesting to note that all three supervised networks proposed for green waveform processing seem to reduce significantly a major challenge of bathymetric lidar surveys: the water surface uncertainty. Indeed, all three approaches provided surface detections and positioning of extremely

2. 3DMASC CloudCompare plugin available at https://www.danielgm.net/cc/release/, Python scripts made available at https://github.com/p-leroy/lidar_platform, and datasets soon available on https://opentopography.org/

high accuracy, while it is commonly admitted that deriving the position of the air/water interface from bathymetric waveform components is a challenge (Guenther, 1985; Guenther et al., 2000; Lague et al., 2020; Mandlburger et al., 2013). In a context characterized by the increasing popularity of bathymetric lidar remote sensing, this may be an interesting avenue for the development of lightweight sensors unable to embed dual-wavelength systems.

Similarly, the possibility of classifying a large diversity of topo-bathymetric surfaces with a single green point cloud with 3DMASC is promising in light of the increasing availability of lidar surveys. Indeed, bathymetric lidar surveys of the French littoral are now all available in open access through the Shom[3]. The ongoing topographic lidar acquisitions over the complete metropolitan French territory also illustrate the increasing popularity of lidar remote sensing. **This thesis takes place at a time when lidar data are becoming easier to access for thematic specialists**, through open-access platforms such as OpenTopography (Krishnan et al., 2011) or governmental agencies[4]. Our methodological contributions thus resonate with current trends in the academic and industrial domains. In a time where academia has to rethink its practices relative to their carbon footprints, sharing environmentally costly data and providing tools to fully exploit it appears essential. By providing 3D data processing tools to environmental specialists, we also wish to make 3D data acquisitions more sensible and in line with current global issues, which force us to rethink the usefulness of operating airborne lidar sensors to derive 2D rasters. Finally, most of the methods we introduced have a wider scope of application than the sole processing of lidar waveforms, and may thus provide useful insights for other environmental assessments relying on 1D data.

Perspectives

Below, we list some perspectives opened by this work. They can be grouped in three main categories: methodological aspects, perspectives on the application to large and real datasets, and potential applications to other thematic fields.

Methodological perspectives: potential improvements in our neural networks can first be envisioned. In the medium term, our YOLO-based network could be adapted to handle component asymmetry. This would imply to estimate the x-coordinate of the bounding box, its left-side-width, and its right-side-width. The network could also be extended to topographic classes with the help of a topographic waveforms simulator (Gastellu-Etchegorry et al., 2016), and by simulating more types of underwater surfaces with our simulator. If it handles topographic and bathymetric classes,

3. https://diffusion.shom.fr/
4. For example, in France, https://geoservices.ign.fr/

the classification derived from waveforms using this approach could then be compared to the results that can be obtained with 3DMASC and point clouds. Depending on the use cases, YOLO could also be expanded with a regression module to predict parameters linked to the components, for example, vegetation height or ground slope. Our networks based on radiative transfer models could also directly integrate physical constraints.

Another perspective is to combine spatial and temporal encoding to incorporate spatial context into waveform processing neural networks. Concretely, it could take the form of two parallel encoders: one handling a point cloud portion, the other the associated waveform or ensemble of waveforms. This approach would exploit an increased amount of information while preserving high spatial density. It could help with the detection of faint ground returns (Magruder et al., 2010) and weak seabed backscatters, and globally allow the densification of point clouds and their simultaneous classification.

In the longer term, graph analyses could be added to 3DMASC to further enhance information extraction from environmental point clouds. Concretely, the spatial patterns in class repartition over the surveyed area could be studied using the classified point cloud. For example, a similar approach to that used in superpoint-based networks (Landrieu et al., 2018; Robert et al., 2023) could be envisioned. Adjacent points with the same label could thus be summarized as a superpoint, and the resulting ensemble of superpoints would be analyzed using graph theory. This would enable the retrieval of spatial repartition information such as connectivity, local diversity, or dispersion, among others.

Application to real and large datasets: our domain adaptation and prediction transport method should be refined to improve the application of networks developed on simulated data to real data. For example, transportation in the feature space instead of transportation in the input space could be used (Shi et al., 2022). Other methods to learn transformations using neural networks and adversarial training also exist (Shi et al., 2022), including some approaches specifically designed for cases such as ours, in which data from a domain are labelled and data from the other domain are not. Once an optimal strategy is defined, the estimation network should be finalized and assessed on inland waters to see if spatial variations of turbidity can be captured.

Applications in other contexts: collaborations with environmental specialists around time series processing could be developed. Indeed, the tools developed for waveform processing can be applied to other types of 1D signals, and many environmental studies are based on 1D data. Examples of other environmental information extraction with our networks could be bird identification and counting from birdsong recordings, watercourse parameters monitoring for hydrology, or land-use

change estimation from satellite time series.

To this end, a major long-term perspective would be to develop an unsupervised time series pattern detection network, inspired by our YOLO-based network, to extract typical pattern changes and anomalies from time series more globally. Again, it could have applications in a large range of environmental applications.

Finally, evaluating our methods on satellite lidar data would be particularly important, as there is also a global trend in satellite lidar development. Although satellite sensors have a much lower spatial resolution, they lift the need to plan multiple acquisitions for environmental monitoring and make global environmental assessments possible.

APPENDICES

List of publications

Several papers and conference proceedings derived from the work presented in this thesis were published. Some proceedings were not published in open access journals. We thus include the author version of these publications in the next pages.

Peer-reviewed journal papers

- **Letard, M.**, Collin, A., Corpetti, T., Lague, D., Pastol, Y., & Ekelund, A. (2022). *Classification of land-water continuum habitats using exclusively airborne topobathymetric lidar green waveforms and infrared intensity point clouds*. Remote Sensing, 14(2), 341.

- **Letard, M.**, Lague, D., Le Guennec, A., Lefevre, S., Feldmann, B., Leroy, P., Girardeau-Montaut, D. & Corpetti, T. (2024). *3DMASC: Accessible, explainable 3D point clouds classification. Application to bi-spectral topo-bathymetric lidar data*. ISPRS Journal of Photogrammetry and Remote Sensing, 207, 175-197.

Peer-reviewed conference proceedings

- **Letard, M.**, Corpetti, T., & Lague, D. (2023, July). *Bathymetric LiDAR Waveform Decomposition with Temporal Attentive Encoder-Decoders*. In IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium (pp. 4435-4438). IEEE.

- **Letard, M.**, Collin, A., Lague, D., Corpetti, T., Pastol, Y., & Ekelund, A. (2022). *Using Bispectral Full-Waveform LIDAR to Map Seamless Coastal Habitats in 3D*. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 463-470.

- **Letard, M.**, Collin, A., Corpetti, T., Lague, D., Pastol, Y., Gloria, H., et al. (2021, September). *Classification of coastal and estuarine ecosystems using full-waveform topo-bathymetric lidar data*

and artificial intelligence. In OCEANS 2021: San Diego–Porto (pp. 1-10). IEEE.

– **Letard, M.**, Collin, A., Lague, D., Corpetti, T., Pastol, Y., Ekelund, A., et al. (2021, July). *Towards 3D mapping of seagrass meadows with topo-bathymetric lidar full waveform processing*. In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS (pp. 8069-8072). IEEE.

TOWARDS 3D MAPPING OF SEAGRASS MEADOWS WITH TOPO-BATHYMETRIC LIDAR FULL WAVEFORM PROCESSING

Mathilde Letard[1], Antoine Collin[1], Dimitri Lague[2], Thomas Corpetti[3], Yves Pastol[4], Anders Ekelund[5], Gérard Pergent[6], Stéphane Costa[7]

1. EPHE, PSL Université Paris, CNRS UMR 6554 LETG, Dinard, France.
2. Univ Rennes, CNRS, Géosciences Rennes - UMR 6118, Rennes, France
3. CNRS UMR 6554 LETG, Rennes, France.
4. SHOM, Brest, France.
5. AHAB, Leica Geosystems, Hexagon, Jönköping, Sweden.
6. FRES 3041 – UMR 6134, University of Corsica, Corte, France.
7. Université Caen-Normandie, CNRS UMR 6554 LETG, Caen, France.

ABSTRACT

Topo-bathymetric lidar is a powerful tool to survey coastal ecosystems while ensuring data continuity between land and water regardless of the nature of the terrain, and allowing the collection of information up to several dozens of metres deep. This study analyzes the potential of full waveform lidar data to monitor key ecosystems for climate change mitigation: seagrasses. It proposes an original way of processing topo-bathymetric lidar waveforms to map their spatial repartition and extent in Corsica (France). Waveform statistical and shape parameters are computed and used to produce a map of seagrass meadows that reaches over 86% of overall accuracy. Seagrass height is also extracted, offering perspectives for structural complexity assessment and ecosystem services quantification.

Index Terms— Topo-bathymetric lidar, Lidar waveform, Seagrass, 3D mapping, Machine learning.

1. INTRODUCTION

Seagrasses grow in worldwide nearshore areas and occupy thousands of square kilometers [1]. Seagrass meadows provide food support and habitat to marine species and human communities, protect coastlines by ensuring wave attenuation and are critical blue carbon sinks: they are therefore of great ecological interest to many coastal areas [2]. Although their role in ocean-climate change mitigation is widely recognized, seagrasses are threatened: a third of the European seagrass area was lost between 1869 and 2016 [3] due to multiple anthropogenic and natural stressors [4].

Monitoring the extent, health and diversity of seagrass ecosystems is crucial to ensure efficient and sustainable management of coastal waters. Maps of seagrass meadows are currently made by processing observations acquired with various techniques [5], as for example satellites [5], [6], [7], airborne [7] or UAV [8] imagery. Bathymetry extracted from sonar [9], [10], [11] acquisitions or lidar surveys [12], [13] is also the base of many seagrass ecosystems maps. However, both passive imagery techniques and active sonar surveys have limitations: passive imagery is constrained by water turbidity [5], [14] while sonar is unusable in unsafe navigation areas. Collecting inaccessible information on seagrasses is therefore still necessary. There is also a strong need for mapping approaches that allow three-dimensional structural information on seagrasses.

The present study shows how topo-bathymetric lidar can be used to survey seagrass ecosystems and provide both two-dimensional – a map of the extent – and three-dimensional – structural complexity knowledge – information on their state. It features an original signal processing method that proves that not only bathymetry but also benthic return characteristics allow seagrass detection and 3D mapping. This method was assessed on the East coast of Corsica (France, Figure 1), where seagrasses play a key role in ecological equilibrium maintenance. It opens perspectives for the study of coastal ecosystems – topo-bathymetric lidar ensures data continuity between land and water – as well as possibilities to better understand seagrass structural complexity and ecology on extended areas, independently from the terrain's safety.

2. TOPOBATHYMETRIC LIDAR OPERATING PRINCIPLE

Lidar uses the backscatter of a laser pulse to collect data on encountered obstacles, resulting in a dense point cloud. Topo-bathymetric lidar emits laser pulses on two different wavelengths: green (515 nm) and infrared (1064 nm). While the infrared wavelength is fully reflected by water, the green wavelength penetrates the water column and allows sea floor

measurement in low turbidity situations. Airborne topo-bathymetric lidar is therefore ideal to collect continuous data on coastal areas.

Analysis of full waveform lidar data shows that the shape of the reflected signal provides information on the type of objects encountered. In marine areas, the backscattered signals contain three main parts: the sea surface return (a peak), the water column return and the benthic return (sea floor, a peak). Since light diffusion is impacted by water, the backscattered intensity of the benthic echo depends on the water column. Here, we focus on the benthic return's shape and statistical parameters and their use for seagrass detection and mapping.

3. MATERIAL AND METHOD

3.1. Study area

The study area is a 0.15 km² nearshore zone located near San Giuliano in Corsica, France. Seagrasses – *Posidonia oceanica* and *Cymodocea nodosa* – occupy around 42 000 ha of the East Coast of Corsica [15]. The selected area encompasses experimental shallow coastal waters including a part of the large *P. oceanica* meadow known as NATURA 2000 site *« Grand herbier de la côte Orientale »* and two small patches of *P. oceanica* surrounded by fine sediments. Depth ranges from 2 to 16.5 metres in this area.

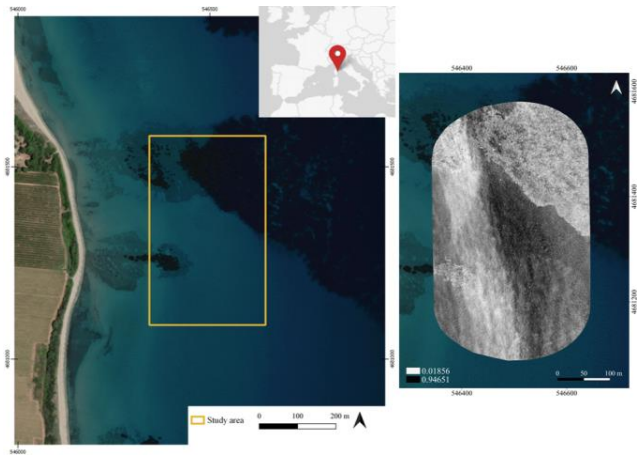


Figure 1. Location of the study area (WGS84 UTM32 N).

3.2. Data used

Topo-bathymetric lidar data acquired in October 2017 with a Leica HawkEye III system was used to conduct this study. The average point density across the zone is 0.99 point/m². Both the points and the full waveforms were used.

3.3. Data processing method

3.3.1. Signal processing

The waveforms (Figure 2) – i.e. the backscattered intensity signal – were first smoothed using a Savitzky Golay filter whose principle is to estimate piecewise polynomial functions to remove noise. The detection of benthic habitats implied to detect the peaks of the waveforms corresponding to the sea surface and to the sea floor. This was made by applying a threshold to the first derivative of the smoothed waveform. A two-step thresholding was adopted to detect both major and smaller peaks and avoid loss of information.

In order to remove the light attenuation component on the benthic echo, a decreasing exponential function is fitted to the water column component of the signal.

To remotely sense benthic habitats with these data, we converted the intensities into pseudo-reflectance values, dividing them by the intensity of the emitted laser pulse.

Various statistical parameters were then computed for each waveform, including benthic peak complexity, which was defined as the number of sign changes of the benthic return's first derivative. Four of these parameters describe the shape of the benthic return: skewness, kurtosis, time range and complexity. Eight other ones give statistics on the intensity variations of the benthic return: mean, median, maximum, variance, standard deviation, area under curve and amplitude. Topography information was also extracted: a bathymetry map, a digital surface model (including the vegetation above the seabed) and the height difference between them were computed. Finally, the pseudo-attenuation, the benthic return's maximum before signal attenuation correction and the sea floor return's maximum before correction were computed.

3.3.2. Image processing

The parameters extracted from the topo-bathymetric lidar waveforms were rasterized by linearly interpolating the values on a 1.5 m resolved grid and stacked together to create a 17-band raster.

3.3.3. Machine learning classification

To generate a map of the extent of seagrasses from the waveforms' parameters, we performed a supervised image classification with a maximum likelihood probabilistic algorithm. The algorithm was trained by 8400 pixels collected on seagrass and sediment areas on the stacked bands, with the help of terrain knowledge provided by aerial imagery. The same number of 8400 pixels were also selected for results' validation.

The whole set of bands was first classified into two classes: sediment and seagrass. Then, we evaluated the contribution of each attribute by computing the accuracy difference between a classification in which it was used and another in which it was not. The parameters that lowered the accuracy were removed from the set of bands systematically,

until the combination giving the best classification was found.

A contribution index of each band of the selected set was computed by dividing the accuracy difference between the classifications with and without the given attribute by the accuracy of the classification obtained with the whole set.

4. RESULTS

4.1. Lidar waveform processing

Figure 2 shows typical detection results for sediment and seagrass, and the output of the light attenuation correction.

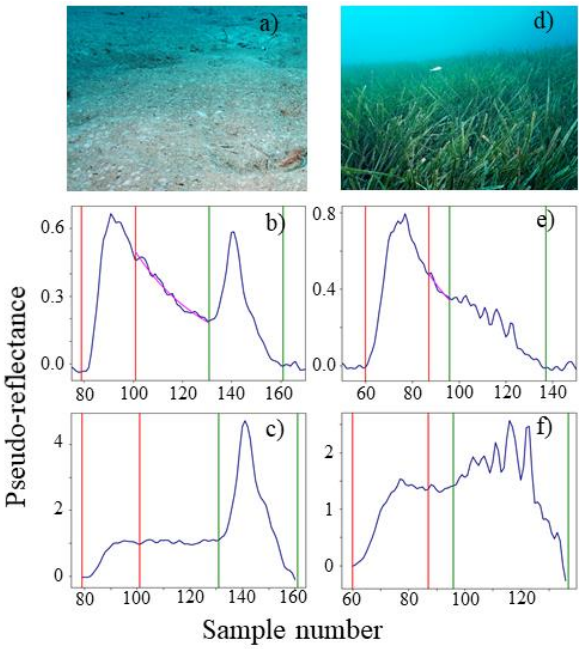


Figure 2. Extracts from the lidar dataset. a) sediment, b) uncorrected waveform, c) corrected waveform; d) seagrass, e) uncorrected waveform, f) corrected waveform. Red lines delimitate the sea surface return, green lines the benthic return; the pink curve is the fitted exponential.

4.2. Machine learning waveform statistics classification

The most efficient attributes' combination is formed by nine parameters: benthic return maximum before pseudo-attenuation correction, peak skewness, pseudo-attenuation index, bathymetry, peak complexity, peak intensity variance, peak median intensity, peak time range.

Calibration pixels have a separability score (Jeffries-Matusita distance) of 1.64 on these bands.

The maximum likelihood classification of the eight most accurate attributes reaches 86.4% of global accuracy and a Kappa coefficient of 0.73. The height difference between the DEM and DSM computed give an overview of seagrass height in the meadow areas. These are presented in Figure 3.

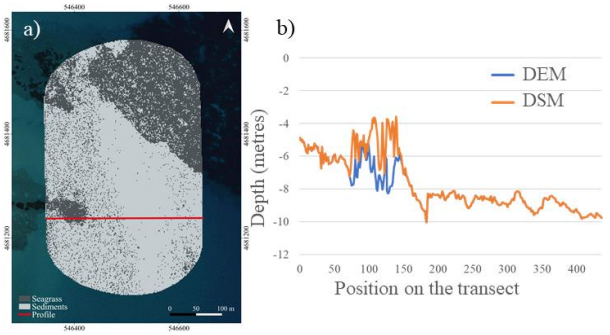


Figure 3. a) map of seabed types in the study area, b) top-of-vegetation and seabed depth profile (metres).

The contribution of each parameter of the selected set is presented in Figure 4. Contributions are sorted in descending order and range from 0.43 to 5.01 percent.

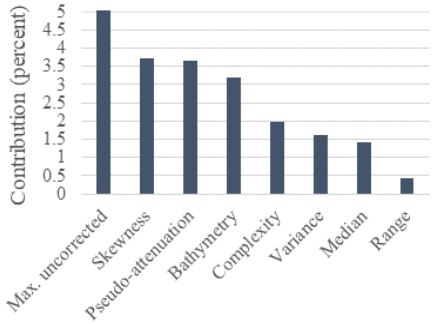


Figure 4: Contribution of the selected attributes to the global accuracy of the best classification (in percent).

5. DISCUSSION AND CONCLUSION

The method used to remove the light attenuation component efficiently corrects the effects of a sediment plume on the reflected intensities. It confirms the usefulness of full waveform data in moderately turbid waters, when intensity alone is not enough to detect bottom variations. However, the correction is perfectible: the exponential over-corrects the attenuation, resulting in very high reflectance measurements. This weakness explains the apparition of pseudo-attenuation as one of the main predictors and the absence of the corrected intensity in the selected attributes.

The analysis of the extracted parameters and their contribution to classification accuracy shows that the returned signal's shape and intensity as well as the bathymetry are all relevant to precisely describe the seascape and its habitats. Results also underpin that the benthic return's shape is determined by the nature of the sea floor and that waveform analysis is still underexploited.

The seabed map obtained reaches 86% of global accuracy and shows both seagrass patches and seagrass extended meadows. Some parts of the site are however misclassified: in very dense vegetation areas, the signal seems unable to penetrate the canopy and the echo has the

same shape as a bare ground echo, explaining the relatively low accuracy obtained. A correction that does not exponentially increase the intensity could compensate these effects, since the maximum of the benthic return has the potential to be more discriminant if well corrected.

Lidar avoids technical limitations encountered by satellite or aerial images-based approaches [5], [14], [9] and is therefore efficient to monitor coastal ecosystems [12], including seagrasses, which usually require landscape-specific mapping approaches to avoid mis-estimation of their extent due to their complex spatial organization [10]. It also offers the possibility to extract seagrass height (Figure 3), which provides insightful elements for seascape analysis and structural ecology assessment of seagrasses. This height estimation could however be improved by interpolating bathymetry under seagrass areas, since the vegetation absorbs all the signal at some point, preventing the extraction of a real bottom echo and biasing the bathymetry and the derived vegetation height.

The waveform decomposition [16], slope correction [13], use of bathymetry-derived parameters [12] or deep learning classification [17] might besides improve sea floor detection and 3D mapping. Although ground truth validation is expected to refine these results, our work already opens various perspectives for coastal habitats monitoring. Further investigation of full waveform data could allow seagrass species, phenological phase or health assessment and provide knowledge on their spatial dynamics depending on ocean currents, which can be studied with the same data.

6. AKNOWLEDGEMENTS

The authors would like to thank the Saur Group and the Région Bretagne for their support.

7. REFERENCES

[1] L.J. McKenzie, L.M. Nordlund, B.L. Jones, L.C. Cullen-Unsworth *et al.*, "The global distribution of seagrass meadows", *Environmental Research Letters 15*, 2020.

[2] E.B. Barbier, S.D. Hacker, C. Kennedy, E.W. Koch, A.C. Stier and C.R. Silliman, "The value of estuarine and coastal ecosystem services", *Ecological Monographs 81*, pp. 169-193, 2011.

[3] C. B. De los Santos, D. Krause-Jensen, T. Alcoverro, N. Marbà, C. M. Duarte *et al.*, "Recent trend reversal for declining European seagrass meadows", *Nature Communications 10*, 2019.

[4] R.J. Orth, T.JB. Carruthers, W.C. Dennison, C.M. Duarte, J.W. Fourqurean *et al,* "A global crisis for Seagrass Ecosystems", *BioScience 56*, pp. 987-996, 2006.

[5] L.J. McKenzie, M.A. Finkbeiner, H. Kirkman, *Global Seagrass Research Methods*, Elsevier Science, 2001.

[6] K. Topouzelis, D. Makri, N. Stoupas, A. Papakonstantinou and S. Katsanevakis, "Seagrass mapping in Greek territorial waters using Landsat-8 satellite images", *International Journal of Applied Earth Observation and Geoinformation 67,* pp.98-113, 2018.

[7] P.J. Mumby, E.P. Green, A.J. Edwards, and C.D. Clark, "Measurement of seagrass standing crop using satellite and digital airborne remote sensing", *Marine Ecology Progress Series 159,* pp. 51-60, 1997.

[8] D. James, A. Collin, T. Houet, A. Mury, H. Gloria, and N. Le Poulain, "Towards Better Mapping of Seagrass Meadows using UAV Multispectral and Topographic Data," *Journal of Coastal Research 95*, pp.1117-1121, 2020.

[9] J. Barrell, J. Grant, A. Hanson and M. Mahoney, "Evaluating the complementarity of acoustic and satellite remote sensing for seagrass landscape mapping", *International Journal of Remote Sensing 36,* pp. 4069-4094, 2015.

[10] V. Pasqualini, C. Pergent-Martini, P. Clabaut and G. Pergent, "Mapping of Posidonia oceanica using aerial photographs and side scan sonar: application off the island of Corsica (France)", *Estuarine, Coastal and Shelf Science 47*, pp. 359-367, 1998.

[11] T. Komatsu, C. Igarashi, K. Tatsukawa, S. Sultana, Y. Matsuoka and S. Harada, "Use of multi-beam sonar to map seagrass beds in Otsuchi Bay on the Sanriku Coast of Japan", *Aquatic Living Resources 16,* pp. 223-230, 2003.

[12] R. Zavalas, D. Ierodiaconou, D. Ryan, A. Rattray and J. Monk, "Habitat classification of temperate marine macroalgal communities using bathymetric LiDAR", *Remote sensing 6,* pp. 2154-2175, 2014.

[13] C.K. Wang and W.D. Philpot, "Using airborne bathymetric LiDAR to detect bottom type variation in shallow waters", *Remote sensing of Environment 106*, pp. 123-135, 2007.

[14] T. Kutser, E. Vahtmäe and G. Martin, "Assessing suitability of multispectral satellites for mapping benthic macroalgal cover in turbid coastal waters by means of model simulations", *Estuarine, Coastal and Shelf Science 67*, pp. 521-529, 2006.

[15] A. Valette-Sansevin, G. Pergent, K. Buron, E. Damier and C. Pergent-Martini, "Continuous mapping of benthic habitats along the coast of Corsica: a tool for the inventory and monitoring of blue carbon ecosystems", *Mediterranean Marine Science 20*, pp. 585-593, 2019.

[16] Y. Qin, T.T. Vu, and Y. Ban, "Toward and optimal algorithm for LiDAR waveform decomposition", *IEEE Geoscience and Remote Sensing Letters 9*, pp. 482-486, 2012.

[17] M. Letard, A. Collin, and Y. Pastol, "Improving mangrove and salt marsh mapping using superpixel-based deep classification of VHR superspectral satellite imagery and topobathymetric LiDAR data", *Sea Tech Week 2020 (12th edition),* Brest, 2020.

Classification of coastal and estuarine ecosystems using full-waveform topo-bathymetric lidar data and artificial intelligence

Mathilde Letard
Centre de géoécologie littorale, Ecole Pratique des Hautes Etudes (EPHE), Université Paris Sciences Lettres (PSL)
Dinard, France
mathilde.letard@ephe.psl.eu

Antoine Collin
Centre de géoécologie littorale, Ecole Pratique des Hautes Etudes (EPHE), Université Paris Sciences Lettres (PSL)
Dinard, France
antoine.collin@ephe.psl.eu

Thomas Corpetti
Littoral Environnement Télédétection Géomatique, UMR 6554 Centre national de la Recherche Scientifique
Rennes, France
thomas.corpetti@univ-rennes2.fr

Dimitri Lague
Géosciences Rennes, Université Rennes 1, UMR 6118 Centre National de la Recherche Scientifique
Rennes, France
dimitri.lague@univ-rennes1.fr

Yves Pastol
Service Hydrographique et Océanographique de la Marine (SHOM)
Brest, France
yves.pastol@shom.fr

Hélène Gloria
Centre de géoécologie littorale, Ecole Pratique des Hautes Etudes (EPHE), Université Paris Sciences Lettres (PSL)
Dinard, France
helene.gloria@ephe.psl.eu

Dorothée James
Centre de géoécologie littorale, Ecole Pratique des Hautes Etudes (EPHE), Université Paris Sciences Lettres (PSL)
Dinard, France
dorothee.james@ephe.psl.eu

Antoine Mury
Centre de géoécologie littorale, Ecole Pratique des Hautes Etudes (EPHE), Université Paris Sciences Lettres (PSL)
Dinard, France
antoine.mury@ephe.psl.eu

Abstract—Coastal and estuarine ecosystems are facing spatio-temporal changes and suffer from the effects of accelerated natural destructive processes due to climate change. Monitoring these areas is crucial to protect them and maintain the ecological balance of shorelines. In this context, full-waveform airborne topo-bathymetric lidar is a reliable tool to collect data seamlessly over land-water continuum zones, thanks to its dual wavelength configuration. It is therefore optimal for coastal habitats monitoring and mapping. However, lidar waveform processing often relies on peak detection and feature extraction that are difficult to configure and often sensitive to noise. In this article, we rather suggest not to rely on hand-crafted features by relying on U-time, a neural network inspired by the well-known UNet convolutional neural network, to identify peaks in waveforms and classify them to discriminate coastal ecosystems efficiently. The network is tested on green waveforms and we evaluate in addition the contribution of infrared intensities. Results show equivalent performances, and obtain over 92% of accuracy when accepting a 2 samples margin of error for peaks location, which does not impact heavily waveform analysis, considering usual peaks widths. Our study shows green waveforms alone allow habitats detection with a F-score of 94%, outperforming previous methods.

Keywords—full-waveform topo-bathymetric lidar, classification, coastal and estuarine ecosystems, temporal neural network

I. INTRODUCTION

Coastal and estuarine areas embody exceptionally diverse ecotones, sheltering many ecosystems and providing ever growing human communities with key services: food, cultural activities and protection from natural hazards emerging from rising oceans and meteorological events [1], [2]. Therefore, they constitute a basis on which entire societies can rely and are the scene of a number of social-ecological processes at different time scales. From semi-daily tides to the development of entire touristic zones over decades through seasonal activities, land-water continuum areas are constantly evolving [3]. Water-level changes bring new landscapes at various time scales (daily, seasonally, yearly); in addition, sedimentary processes change the terrain's morphology while sudden events like storms generate rapid changes on ecosystems [3]. On both sides of the shoreline, ecosystems are under the influence of marine hydrodynamics and terrestrial hydrological processes, making coastal fringes and estuaries a meeting point for spatio-temporal changes. Monitoring these changes is crucial to protect such environments from accelerated natural destructive processes [4] as well as to ensure that these attractive areas can still sustain anthropic pressure.

The first step to monitor the land-water continuum areas' dynamics consists in modelling their spatial structure seamlessly by gathering information both below and above the air-water and the land-water interfaces. Current researches on this topic exploit unmanned airborne vehicle (UAV) [5], manned aerial [6] or satellite imagery [7]. These approaches have proven to be effective over land, where infrared wavelengths bring high accuracy in vegetation detection, but are limited by the water's clarity in estuarine or marine environments. The underwater areas are however often surveyed with acoustic data [8], which meets the spatial resolution requirements but does not bridge the gap between submerged and emerged domains and leaves non-navigable zones unseen.

To face these limitations, airborne topo-bathymetric lidar constitutes a reliable alternative by ensuring continuous information between land and water [9], [10]. LiDAR principle consists in sending a laser pulse and the backscattered signal (also called discrete echoes) embeds many information related to the encountered objects. In addition airborne lidar covers large areas, penetrates depth of up to dozens of meters and has a higher spatial resolution than hyperspectral satellite imagery. Topo-bathymetric lidar is therefore ideal to capture continuous data on coastal or riverside areas. However, even if classifying discrete echoes in terrestrial lidar is an explored issue with interesting techniques [11], analyzing echoes both above and below water ground is still a challenging issue [12]. This indeed requires rigorous pre-processing to remove noise and water surface induced returns, and available approaches mostly rely on geometrical features and spatial context information [13,11, 12].

Full-waveform lidar data contain rich information related to the terrain, that have proven to be relevant to solve classification problems and enable the mapping of land- and water uses [9], [10]. Coupled with innovative processing algorithms, these data make high density mapping of the land-water continuum feasible. However, the exploitation of full-waveform data often relies on peaks detection [14, 15, 16, 17] followed by feature extraction and classification [14, 9, 10], that are often all application-dependent and require advanced settings [15, 16]. Both traditional peak detection and feature extraction methods rely on gradient computation that are very sensitive to noise [16] and therefore, this prevents from reliable results. In this study, we process bi-spectral lidar data acquired with a Leica HawkEye III 4X sensor that produces point clouds of unprecedented density in underwater areas, and propose an approach able to detect and classify waveform components without supervised peak detection and feature extraction.

More precisely, we focus on the ability of U-time [18] , a neural network inspired from the well know UNet convolutional neural network adapted to image analysis, to classify waveform samples. In practice, we aim at (1) identifying real peaks from noise and signal not corresponding to a backscatter and (2) classifying them to discriminate coastal ecosystems efficiently and automatically detect the presence of water or vegetation. We propose neural network adapted to lidar waveform processing and classification technique that allows the detection and characterization of various coastal ecosystems, and that correctly separates subtidal, intertidal and supratidal areas, enabling the seamless monitoring of the coastal fringe at high spatial resolution. This network is calibrated and validated using very high-resolution UAV optical imagery and lays the foundations for future integrated approaches of the land-water continuum mapping.

II. Study Area and Data Acquisition

A. Study area

The study area (Fig. 1) consists in various coastal habitats located in Sables d'Or les Pins in Brittany, France (48°38'27''N: 2°24'24''O). This seaside region is set in an ecologically rich environment encompassing: fine sand and pebble beaches, a sandy dune, rocky areas provided with seaweeds, seagrass meadows, wooded areas, crop fields and salt marshes through which a river flows towards the sea. These ecosystems host a great variety of species: shellfishes, endemic dune plant vegetation, green, red or brown seaweeds, *Zostera Marina* plants, evergreen and deciduous trees, crops and endemic salt marshes plants such as *Halimione portulacoides*, sea poa or purslane.

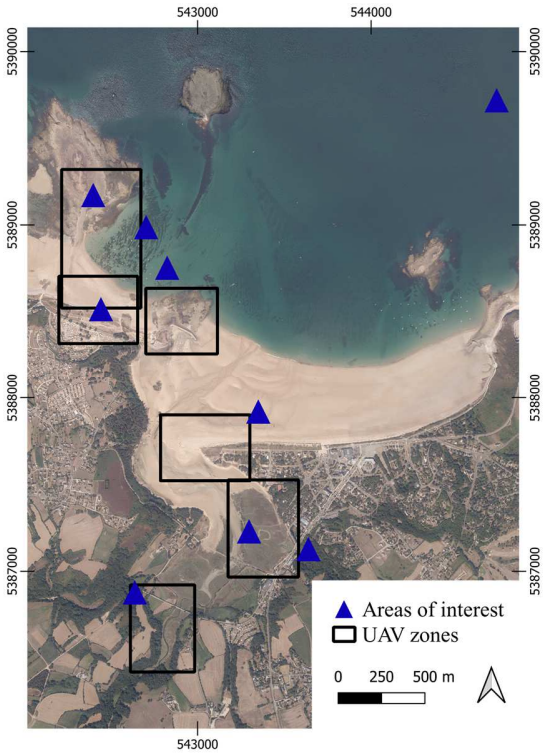


Fig. 1. Natural-coloured imagery of the study area: Sables d'Or les Pins, France (datum; UTM30N).

B. Full-waveform topo-bathymetric lidar

Lidar is an active sensor that emits laser pulses to perform range determination and collect data on encountered obstacles, resulting in backscattered echoes (whose maxima correspond to encountered objects). Topo-bathymetric lidar emits both a green laser pulse around 532 nm wavelength and an infrared laser pulse at 1064 nm. While the infrared wavelength is fully reflected by water, the green wavelength penetrates the water column and allows sea floor measurement in low turbidity situations. Airborne topo-bathymetric lidar is therefore ideal to collect continuous data on coastal areas.

Analysis of full waveform lidar data (i.e. all the backscattered signal and not only the localization of peaks) shows that the shape of the reflected signal provides information on the type of objects illuminated by the laser beam. In marine areas, the backscattered signals contain three main parts: the sea surface return (a peak), the water column return and the benthic return (corresponding to the sea floor, a peak) [12]. Since light diffusion is impacted by water, the backscattered intensity of the benthic echo depends on the water column's clarity and on the depth. It thus explains why similar benthic covers behave differently when hit by the bathymetric laser beam.

In this study dense topo-bathymetric lidar data were acquired over Sables d'Or les Pins in September 2019 with a Leica

This study was supported in part by a doctoral research grant from Région Bretagne and by the Saur Group patronage.

HawkEye III 4X sensor. This topobathymetric lidar system produces laser pulses at wavelengths of 513 nm and 1064 nm. The HawkEye III system is mostly used for hydrographic purposes ; it thus relies on two green channels to collect bathymetry information: a shallow channel that reaches depths up to 10 meters, and a deep channel for deeper waters. To reach dozens of meters deep seabeds, the system is set to emit more powerful beams in the green wavelength, compromising on point density to meet both accuracy and safety requirements. However, the HawkEye 4X relies on a technology that increases the density and the spatial resolution that can be met. The point density of the resulting dataset ranges from 10 points per square meter for the topographic wavelength, to (at least) one and five points per square meter for the deep and shallow bathymetric sensors, respectively. This corresponds approximatively four times the point density of previous Leica HawkEye topo-bathymetric systems. Full-waveform data are available for every bathymetric returns and for one topographic echo out of 32. Due to the green lasers' power and the receiver's limitations, the reflected signal is often saturated over emerged areas with high reflectivity. In these cases, the presence of the infrared channel provides additional information. In this study, only the green shallow channel full-waveforms are used, due to practical difficulties in extracting some infrared waveforms (the files delivered only empty waveforms). The infrared measured intensity was used instead.

C. *UAV surveys*

UAV imagery was acquired on five sub-sites in March and April 2021 to provide knowledge on sea- and land-cover over the areas of interest for this study. It was calibrated by an overall of 55 ground control points georeferenced using Topcon Hyper V and Trimble Geo7X D-GNSSs. Two cameras were used on the UAV: a RGB 5472 × 3648 DJI Phantom 4 Pro V2 producing 8-bit images, and a Parrot Sequoia+ including a 1280 × 960 NIR nadiral sensor (770 nm to 810 nm) with a zenithal irradiance sensor, both producing 16-bit images.

D. *Ground-truth photoquadrats*

Ground-truth data were collected under the form of geolocated photoquadrats, providing punctual knowledge of the land use - land cover / sea use - sea cover and helping in labelling future training data. 30 photoquadrats were collected on each sub-site overflown by the UAV and their positions were recorded with Topcon Hyper V and Trimble Geo7X D-GNSSs.

III. Methodology

A. *Data preprocessing*

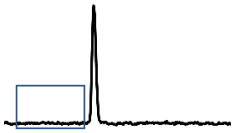
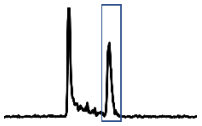
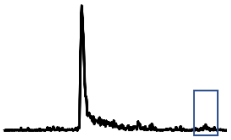
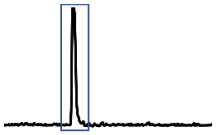
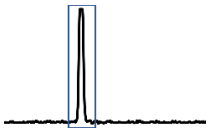
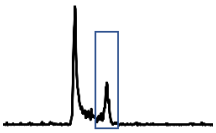
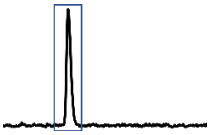
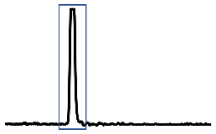
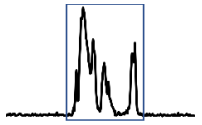
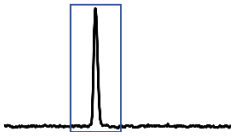
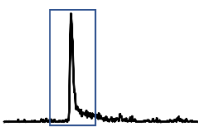
1) UAV imagery and ground-truh data post-processing: An array of five orthoimages and digital surface models were derived from the UAV imagery using the photogrammetric reconstruction procedure on the Pix4D software. Two separate processes were performed for each acquisition: a RGB and a Green-Red-NIR reconstruction. The positions of the ground control points were integrated to improve the quality of the derived products. The photoquadrats used as ground-truth were simply linked to their position after GNSS data post-processing.

2) Green lidar waveforms labeling: In order to train our waveform processing algorithm and assess its performances, 1500 green waveforms per habitat were selected and their echoes were located and labeled. Waveforms already contain returns positions obtained via the machine's constructor peak detection algorithm. We filtered these peaks locations and used them as ground truth after labeling them depending on the nature of the ground in the area (green point cloud files and green waveform files were linked using the gpstime parameter). Peak locations corresponding to noise points were removed using the noise classification flag present in the point clouds delivered by the Leica survey software. We then selected areas of interest for each type of ground to discriminate, using the UAV orthomosaics, an orthoimage from 2014 and the ground-truth data, and selected representative waveforms in these areas. The 2014 orthoimage was used as past ground-truth to confirm the ground type since there is 1.5 years gap between the lidar acquisition and the UAV and photoquadrat ones. For each representative area, the echoes were assigned to the label corresponding to the type of cover observable both in 2014 and in 2021. For underwater objects, several echoes corresponding to different cover types were present. The distinction between the seabed echoes and the sea surface return was made using the flag corresponding to the sea surface present in the point cloud. This detection is considered to be trustworthy as (1) the sea surface echo is easily distinguishable in shallow water areas and (2) this is the flag used by the French Marine hydrographic and oceanographic service to process their lidar point clouds and produce their underwater maps. No very shallow water areas were selected to create the labeled dataset in order to avoid return location uncertainties. Waveforms were individually visually checked before and after echoes selection and labeling to make sure no outliers were noticeable in the dataset. In practice nine covers were identified as well as a "water surface" class since it is relevant for underwater waveforms. Finally, a "non useful" class was created for all discrete values not assigned to a specific object. All the echoes classes are sumed up in Table I. A total of 13500 labels – 1500 per cover type – was generated for the corresponding waveforms, under the form of series of 512 values alternating between 0 (= non useful) and the label of the relevant class at the position of an echo. For each cover type, 1000 waveforms were dedicated to training and validating the classification and 500 were kept to test the trained algorithm. There are consequently two datasets: a 9000 items training dataset, and a 4500 items test dataset, one item consisting in a couple waveform-label.

*3) Green lidar waveforms preprocessing:*In order to achieve more generalizable results, we converted the backscattered intensities into pseudo-reflectance values. Indeed, lidar intensities depend on many acquisition related parameters and cannot be considered as an absolute measurement of a surface's albedo. Since the laser impulse is registered for each waveform, we divided the backscattered signal by the maximum of the impulse to obtain pseudo-reflectances, that would probably be less dependent on our dataset. To be passed as input of the U-time, the waveforms also were normalised between 0

This study was supported in part by a doctoral research grant from Région Bretagne and by the Saur Group patronage.

TABLE I. CLASSES USED TO LABEL INTENSITY SAMPLES IN THE LIDAR WAVEFORMS, INTEGER LABELS ASSOCIATED, PHOTOQUADRATS (NOT AVAILABLE FOR TERRAINS WITH LOW ACCESSIBILITY) AND TYPICAL SHAPE OF THE CORRESPONDING WAVEFORM COMPONENT.

Class	Label	Photoquadrat	Shape in waveform
Non useful	0		
Shallow sand	1		
Deeper sand	2		
Emerged rock	3		
Salt marsh	4		
Seagrass	5		
Pebble	6		
Field	7		
Tree	8		
Emerged sand	9		
Sea surface	10		

and 1, and only the 512 first samples were kept since the last samples of the waveforms usually only contain noise.

4) Infrared lidar data analysis: To construt our bispectral dataset, the infrared (IR) backscattered intensity was extracted from the IR point cloud at each position were a green waveform was selected for the train and test datasets. To get this parameter, the IR point cloud was first cleaned up by removing noisy points observable meters above the surface. The resulting point cloud was then rasterized at a 5 meters resolution, to obtain a digital intensity model. No interpolation was made in order to preserve the data, empty cells were consequently left empty. Finally, the value of the digital infrared intensity model was extracted for each position at which a green waveform was used. The 5 meters resolution was chosen to account for the impossibility to find a green point at the exact same position as an infrared point, due to the differences of both lasers and to their non-cofocality. Although they could not be converted into pseudo-reflectance values, the infrared intensities were also normalized between 0 and 1 before being passed to the U-time.

B. Deep learning for echoes detection and classification

1) Neural network architecture: Before entering into details, let us remind that the main idea behind deep neural networks is to model complex relations between inputs (waveforms here) and outputs (labels associated with each point) through the succession of a large number of elementary operations performed with neurons. The parameters associated with each neuron are learned during the training phase, aiming at finding the optimal parameters that minimize the loss function, i.e. the discrepancy between estimated labels and real ones, with ad-hoc optimization tools. When spatial/temporal relations exist between input data, as with waveforms, convolutional networks are often used to capture multi-scale relationships between inputs. In image processing, efficient networks performing convolution/deconvolution of the data have enabled impressive results for image understanding and classification. In this work, we rely on a 1-dimensional (1D) version of a convolutional/deconvolutional network, named U-time, to process the waveforms. The chosen architecture is a network inspired by the U-Net neural network, characterized by its so-called skip connections linking the encoder (convolution) and the decoder (deconvolution) parts of the model. To adapt the network to our time series, we only use 1D convolutions, contrary to the original architecture that was created to process 2-dimensional images. The encoder part consists in four convolution blocks of a fixed-size kernel, each encompassing two sequences of convolution, batch normalization and activation with a rectified linear activation unit (ReLU). The two last operations enable to keep only the main interesting part of the information. Max pooling, i.e.reducing by two the size of the filtered waveforms, is performed at the end of each of these blocks. By doing so, the convolution of the next level (with the same size kernel) embed large scale information. The decoder part consists of four sequences of upsampling, skip connection and convolution block. Both parts are linked by a convolution

block. Our architecture consequently has a depth of four max pooling layers, meaning that at the end of the encoding part, the waveform is synthesized into 32 samples, which seems adapted to the problem, considering there are between one and four to five echoes detected, which are often located in the second third of the signal, surrounded by noise at the beginning and the end. In practice the size of the convolution kernel is 3. More details about the architecture are presented in Fig. 2.

2) Training: Two different networks were trained: a monospectral model, processing green waveforms only, and a bispectral model that uses both the green waveforms and the infrared intensity values to perform echoes detection and classification.

Both networks were trained on batches of 30 samples using weighted categorical crossentropy as loss function and stochastic gradient descent as optimization process. The set of weights resulting in the lowest validation loss was kept for each configuration (mono- or bi-spectral).

3) Post-processing: Since the network tends to label several following samples as an echo while the labeled data locate an echo under the form of a single position in the waveform, the predictions were post-processed to keep only one sample as the location of the detected echo. Each time a continuous array of samples obtained a label different than zero, the sample having the highest probability of the array was kept as is, while the others were re-labeled with a zero (class "nothing").

4) Quantitative assessment of the results: Accuracy (ratio of good classification, best when its value is 1), precision (fraction of correct classes among each ground truth classes, , best when its value is 1), recall (fraction of correct classification for each estimated class, best when its value is 1), F-score (combination of precision and recall, best when its value is 1) and confusion matrixes were used to assess the quality of the predictions quantitatively. Different assessments were performed to identify the strengths and weaknesses of our approach.

The quality of the peak detection was assessed by considering this task as a binary problem (presence or absence of a return). By analyzing the predicted location of the samples having a label different than zero regardless of the accuracy of the label, we can assess the performance of our architecture for waveform components detection by computing accuracy, precision, recall and F-score. Considering the width of the peaks, the fact that the training labels correspond approximately to their maximum and the usual considerations for waveform processing, a margin of error can be accepted for the location of the peaks. The peak detection assessment was therefore also conducted with a tolerance of ±2 samples offset.

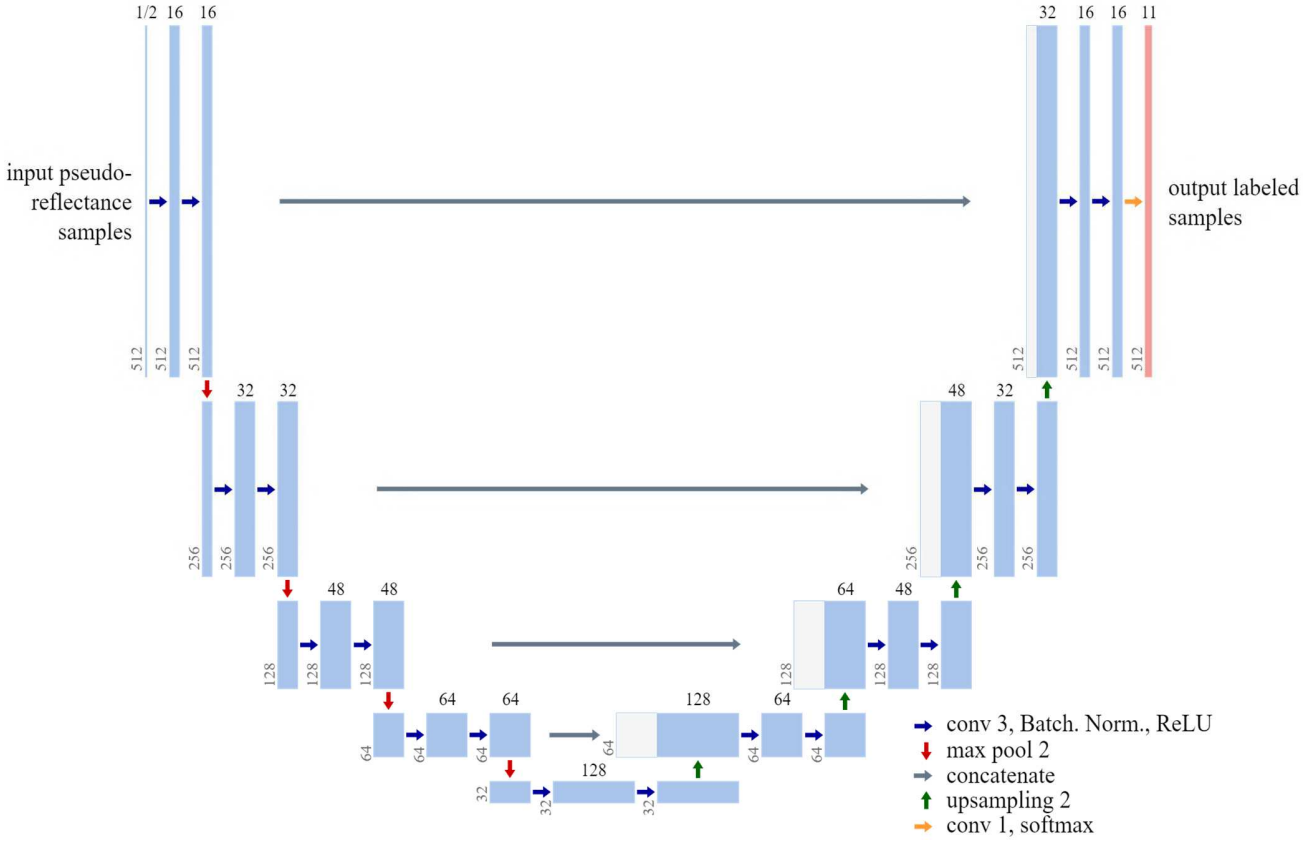


Fig. 2. Diagram of the 1-dimensional U-time neural network architecture implemented to classify full-waveform pseudo-reflectance samples and intensity values and detect and label returns corresponding to sea- or land-covers.

The second aim of our approach is to detect the land- or sea-covers in the waveforms. The different objects do not necessarily need to be placed correctly in the waveforms, since the simple information of their presence or absence will be used for mapping or monitoring purposes. Classes detection is therefore a relevant criterium to evaluate our approach. It was assessed by considering the classes predicted in each waveform, regardless of their occurrence or positions and aims at determining whether the neural network detects the different objects correctly. Accuracy, precision, recall and F-score were computed for this task too.

The last evaluation criterion is the quantification of the quality of the global classification, i.e. assessing the prediction made on each sample for each waveform. Confusion matrixes were computed to get detailed information on the behavior of the networks and the main confusions impacting both peak detection and classification.

IV. Results

A. *Monospectral network:*

The first network was trained only on the green waveforms. It reached its lowest validation loss (0.11) on the 106th training epoch. The weights obtained at this epoch were kept to test the model's performances. The following observations were made based on the testing dataset.

Qualitatively, the results show good identification of the different ecosystems. The parts of the waveform that are labeled as a return are consistent with the waveform components. As expected, before post-processing, continuous ranges of samples are often labeled as echoes of the same class, covering the whole peak. Since the ground-truth dataset is labeled with one sample indicating a peak, post-processing is useful if we want to assess quantitatively the quality of the detection and classification. Example of results obtained in the testing dataset are presented in Fig. 3.

1) Peaks detection: Although most of the returns seem to be correctly placed when visualizing the predictions, metrics presented in Table II show that less than half the returns predicted actually exist. However, more than half of the echoes are properly detected.

If we choose to tolerate an error of 2 in the predicted location of the echoes (i.e. consider that if an echo was predicted two samples further or earlier than the truth it is a correct prediction), the statistics improve significantly. Let us recall that this margin of error does not massively impact waveform analysis, since peaks are wider than 2 samples and the locations indicated in the training data correspond to the middle/maximum of the peaks.

TABLE II. Performance metrics of the peaks detection obtained with the green waveforms-based neural network

Metric	No location error tolerance	Location error tolerance of 2
Precision	0.42	0.63
Recall	0.62	0.94
Accuracy	0.99	0.99
F-score	0.50	0.76

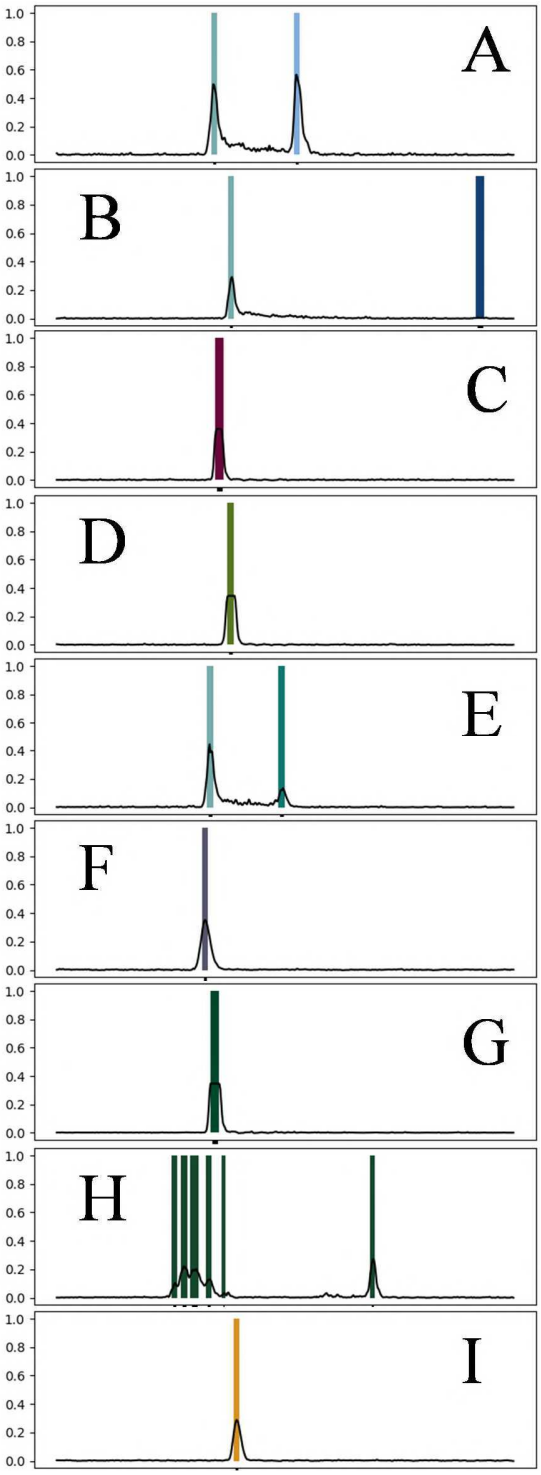


Fig. 3. Examples of predictions made by the green waveforms-based U-time. The black signal corresponds to the waveform. Each colored vertical line refers to a predicted peak. All nine cover types are pictured. A: shallow sandy seabed, B: deeper sandy seabed, C: rock, D: salt marsh, E: seagrasses, F: pebble, G: field, H: tree, I: emerged sand.

With this error margin acceptance, more than 60% of the echoes predicted are real, and 94% of the existing returns are identified by the model. There is still an overestimation of the presence of

returns, but few of the real echoes are missed. Furthermore, 98% of the echoes truly located 2 samples away from the detection were originally detected by the network but erased during post-processing.

2) Classes detection: In terms of land- or sea-cover identification, the monospectral network performs well. On average, 89% of the covers predicted at a given location are present in reality, and 99% of the objects hit by the laser beam are identified. Few covers are missed, but some are inaccurately predicted, as Table III reflects.

3) General classification: The average accuracy of the classification after post-processing is 0.61. The confusion matrix obtained (Fig. 4) gives further indication on the performances of the network for each type of sample. Considering the values present in the first column (predicted = 0), the major source of error is the mislocation or misdetection of the echoes. Indeed, the highest percentages of confusion are obtained when a label 0 is predicted instead of the actual label. Moreover, the percentage of misclassification often corresponds to the difference between 1 and the accuracy obtained for a class.

The few cover type errors can be summed up as follows. Shallow sandy seabed, trees and seagrasses are sometimes mixed up. Rock is also mistaken for other types of terrestrial covers having similar saturated echoes (pebble, tree, salt marsh vegetation or field), and inversely. Bare ground types can be misidentified too.

The confusion matrix obtained on the post processed results with a tolerance for a location error of 2 for the peaks (Fig. 5) confirms the observation made above. The major explanation for the low accuracies obtained for most of the classes is the improper location of the returns by a few samples. Cover types errors remain rare, but there are still missed returns, and false detections, mostly of deep sandy seabed and trees. The average accuracy reaches 92%.

TABLE III. PERFORMANCE METRICS OF THE CLASSES DETECTION OBTAINED WITH THE GREEN WAVEFORMS-BASED NEURAL NETWORK.

Metric	Average value
Precision	0.89
Recall	0.99
Accuracy	0.97
F-score	0.94

		Predictions										
	Label	0	1	2	3	4	5	6	7	8	9	10
Truth	0	**1,00**	0	0	0	0	0	0	0	0	0	0
	1	0,25	**0,74**	0	0	0	0,006	0	0	0,002	0	0
	2	0,53	0	**0,47**	0	0	0	0	0	0	0	0
	3	0,68	0	0	**0,29**	0,01	0	0,002	0,007	0,002	0,002	0
	4	0,55	0	0	0,002	**0,44**	0	0	0,004	0	0	0
	5	0,39	0,002	0	0	0	**0,60**	0	0	0,002	0	0
	6	0,31	0	0	0,002	0	0	**0,65**	0	0	0,03	0
	7	0,52	0	0	0	0	0	0	**0,48**	0	0	0
	8	0,33	0,001	0	0	0	0	0,001	0	**0,66**	0,001	0
	9	0,28	0	0	0	0	0	0,02	0	0	**0,69**	0
	10	0,26	0	0	0	0	0	0	0	0	0	**0,74**

Fig. 4. Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the green waveforms based neural network.

		Predictions										
	Label	0	1	2	3	4	5	6	7	8	9	10
Truth	0	**1,00**	0	0	0	0	0	0	0	0	0	0
	1	0	**0,98**	0	0	0	0,01	0	0	0,004	0	0
	2	0,06	0	**0,94**	0	0	0	0	0	0	0	0
	3	0,20	0	0	**0,75**	0,03	0	0,002	0,01	0,002	0,002	0
	4	0,07	0	0	0,05	**0,88**	0	0	0,004	0	0	0
	5	0,03	0,002	0	0	0	**0,96**	0	0	0,002	0	0
	6	0,02	0	0	0,01	0,002	0	**0,89**	0	0,002	0,07	0
	7	0,07	0	0	0,01	0	0	0	**0,92**	0	0	0
	8	0,09	0,001	0	0,003	0,001	0,002	0,001	0	**0,90**	0,001	0
	9	0	0	0	0,006	0	0	0,04	0	0	**0,95**	0
	10	0	0	0	0	0	0	0	0	0	0	**1,00**

Fig. 5. Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the green waveforms based neural network, with a tolerance of a location error of ± 2 samples for the returns.

B. *Bispectral network:*

The second network was trained on green waveforms and corresponding infrared intensity values. Its lowest validation loss (0.07) was observable on the 82nd epoch of the training process, providing the weights for the final model. The testing dataset was processed with this model and accounts for this configuration's performances for the detection and classification of waveform components.

Once again, visual appreciation of the predictions show proper identification of most of the ecosystems and the returns. Overestimation of the number of samples to be labeled for a given echo is still observable and justifies again the need for post-processing. Example of results are presented in Fig. 6.

1) Peak detections: The bispectral network shows slightly poorer precision, recall and F-score for peaks detection when no margin of error is tolerated than the monospectral model. 62% of the detected echoes do not exist in the truth dataset. And 2% less returns are properly detected. Table IV enumerates the performance statistics of this second network for peak detection:

If, as previsously, we set an error margin of ±2 for the location of the peaks, the statistics increase, but slighly less than it was the case for the first configuration. Almost 60% of the echoes are true positives, and the network suitably detects 94% of the true returns. As for the first network, 98% of the echoes actually located within 2 samples of the predictions were originally labeled as such, but erased during post-processing.

2) Classes detection: The bispectral network predicts land- and sea-covers somewhat better, as prove the metrics presented in Table V. Only 9% of the covers predicted are false positives, and 99% of the habitats are found. The average accuracy of the identification of habitats is 98%.

3) General classification: The global classification task was performed with an average accuracy of 0.61. More details about

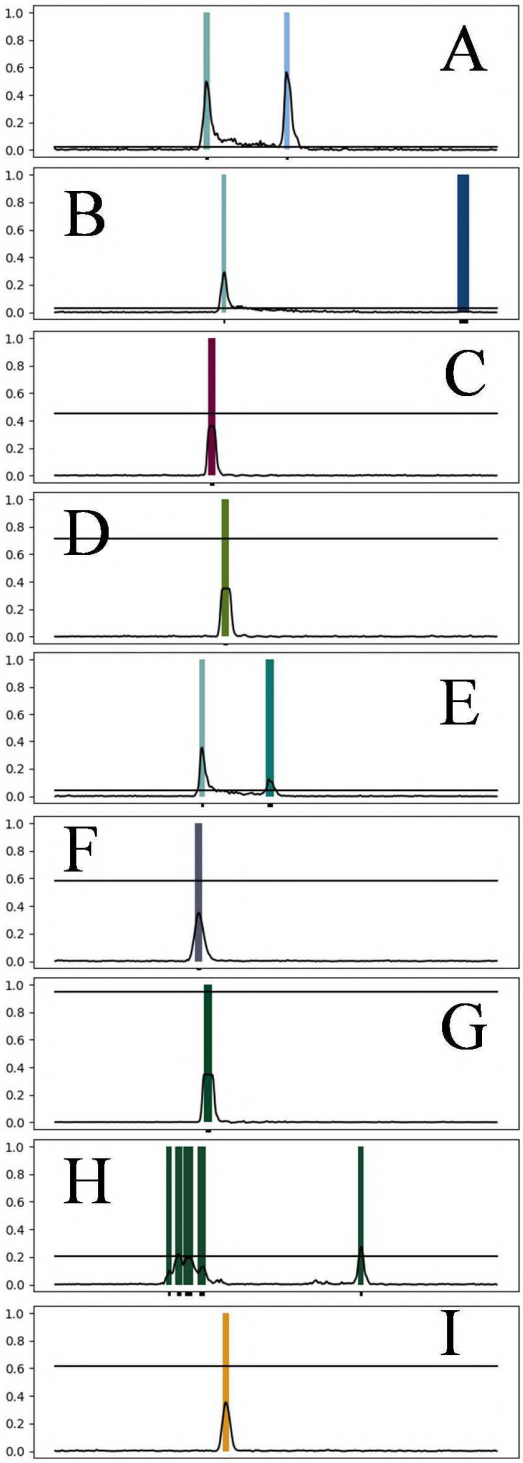


Fig. 6. Examples of predictions made by the IR intensities and green waveforms-based U-time. The black horizontal line represents the normalized IR intensity and the black signal corresponds to the green waveform. Each colored vertical line refers to a predicted peak. All nine cover types are pictured. A: shallow sandy seabed, B: deeper sandy seabed, C: rock, D: salt marsh, E: seagrasses, F: pebble, G: field, H: tree, I: emerged sand.

TABLE IV. PERFORMANCE METRICS OF THE PEAKS DETECTION OBTAINED WITH THE INFRARED INTENSITIES AND GREEN WAVEFORMS-BASED NEURAL NETWORK.

Metric	No location error tolerance	Location error tolerance of 2
Precision	0.38	0.59
Recall	0.60	0.94
Accuracy	0.99	0.99
F-score	0.46	0.73

TABLE V. PERFORMANCE METRICS OF THE CLASSES DETECTION OBTAINED WITH THE INFRARED INTENSITIES AND GREEN WAVEFORMS-BASED NEURAL NETWORK.

Metric	Average value
Precision	0.91
Recall	0.99
Accuracy	0.98
F-score	0.95

classification errors can be found in the confusion matrix presented in Fig. 7. Confusion between the label 0 and the others is agains the biggest source of errors and appears to be dragging down the accuracy scores of most of the classes.

Less confusion between classes other than “nothing” is observable. Underwater sand, seagrasses and trees are still mistaken for each other, as well as rock, trees and dry sand or pebble, fields and dry sand. Confusion between trees and sea surface also happen, which was not the case for the first network. However, globally, there seems to be less confusion between terrestrial covers. The separation between subtidal and supratidal areas is somewhat better, except for the false trees detections in marine environments.

Similarly to the green-based predictions, with tolerance for a margin error of 2 samples in the peaks locations, the average accuracy increases. It reaches 93%, a 1% improvement, with green and infrared based predictions. As shown in Fig. 8, it also

		Predictions										
	Label	0	1	2	3	4	5	6	7	8	9	10
Truth	0	**1,00**	0	0	0	0	0	0	0	0	0	0
	1	0,30	**0,69**	0	0	0	0,002	0	0	0,004	0	0
	2	0,53	0	**0,47**	0	0	0	0	0	0	0	0
	3	0,69	0	0	**0,30**	0	0	0	0	0,003	0,003	0
	4	0,42	0	0	0	**0,58**	0	0	0	0	0	0
	5	0,41	0	0	0	0	**0,59**	0	0	0	0	0
	6	0,40	0	0	0	0	0	**0,56**	0,002	0	0,03	0
	7	0,47	0	0	0	0	0	0	**0,53**	0	0	0
	8	0,41	0	0	0	0	0	0	0	**0,59**	0,001	0
	9	0,34	0	0	0	0	0	0,01	0	0	**0,65**	0
	10	0,27	0	0	0	0	0	0	0	0,0007	0	**0,73**

Fig. 7. Confusion matrix of the classification of waveforms’ pseudo-reflectance samples obtained with the infrared intensities and green waveforms based neural network.

		Predictions										
	Label	0	1	2	3	4	5	6	7	8	9	10
Truth	0	**1,00**	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	1	0,00	**0,99**	0	0	0	0,006	0	0	0,004	0	0
	2	0,04	0	**0,96**	0	0	0	0	0	0	0	0
	3	0,23	0	0	**0,76**	0,005	0,002	0	0	0,005	0,005	0
	4	0,06	0	0	0,015	**0,92**	0	0	0	0,006	0,002	0
	5	0,03	0	0	0	0	**0,97**	0	0	0	0	0
	6	0,01	0	0	0,006	0	0	**0,90**	0,004	0	0,075	0
	7	0,07	0	0	0,002	0,006	0	0,004	**0,92**	0	0,002	0
	8	0,12	0	0	0,001	0	0	0	0	**0,88**	0,002	0
	9	0,00	0	0	0	0	0	0,03	0	0	**0,97**	0
	10	0,00	0	0	0	0	0	0	0	0,001	0	**0,996**

Fig. 8. Confusion matrix of the classification of waveforms' pseudo-reflectance samples obtained with the infrared intensities and green waveforms based neural network, with a tolerance of a location error of ± 2 samples for the returns.

confirms that the slight offset in the location of some peaks accounts for most of the accuracy loss for ten out of eleven classes. Habitats confusions remain similar, and there are still inopportune detections of trees or deep sandy seabed, on top of missed echoes (for seven out of eleven classes).

V. Discussion

A. *Detection of returns*

The U-time detects almost entire peaks and not only their maximum of halfway location, as they are indicated in the training data. However, this peaks detection is consistent with typical waveform behavior: peaks constitute a significant pseudo-reflectance variation compared to their surroundings in the waveforms. Since noise has approximately always the same shape and pseudo-reflectance and is labeled as "non useful" in all the waveforms, the U-time attributing to the rest of the peaks labels different than 0 is expectable.

To assess the quality of our peak detection, we performed post-processing to obtain results comparable to the labeled data. The post-processed results showed poor precision and recall for peaks location. Since the peaks in our training data were located around their maximum, we know that an offset of about 2 samples in their predicted location will still correspond to the peak in the waveform data. Considering this, we obtain metrics that show the proposed FCN architecture is able to detect peaks in waveforms accurately and efficiently, since few preprocessing was needed, and no impulse function modeling or curve fitting was required. Moreover, the fact that 98% of the detected peaks located within the margin of error were labelled as peaks before pre-processing asserts the ability of the proposed approach to correctly detect peaks. However, it is clear that the main limitation of this approach is the waveform labelling procedure. Indeed, using discrete peaks locations is not adapted to convolutional neural networks, that are mostly based on patterns of variation, and will therefore perform better in identifying the whole range of sample corresponding to the peaks. Labelling the peaks continuously in the training data would avoid needing post-processing and allow the use of intersection of union as a relevant performance metric.

Furthermore, since the files storing the waveforms are structured to contain up to four returns, in some cases, all the actual returns might not be indicated. For example, our approach seems to overestimate the number of tree returns, but in some cases, the detected peaks seem to be relevant. This method could therefore be helpful to improve echoes extraction from waveform data as in [14], but the extra returns' accuracy has to be evaluated first.

For the peak detection task, the proposed approach showed satisfying performances with limited pre-processing, contrary to traditional full-waveform analyses. The green only based U-time showed somewhat better results, but the small difference makes both networks equivalent.

B. *Classification*

The detection of the different habitats, regardless of their positions in the waveforms, is more robust and accurate and shows little confusion. There are more false detections than missed detections, and further analysis and visualization of the predictions shows that most of the false detections could be avoided. Indeed, most of the false detection are trees returns in marine environments and deep sandy seabed returns before the water surface of after the seabed return. By adapting the network architecture, these issues could probably be solved. A recurrent neural network could be useful to make future predictions depend more on previous predictions along the pseudo-reflectance time series and avoid deep sandy seabed returns after another seabed return has already been identified. Region-based convolutional neural networks could also determine the useful range of the waveform and avoid untimely peak detection outside of it.

Another possible improvement is the addition of other types of data to further distinguish different environments. Geometrical features, traditionally used to classify point clouds [13, 11], could make better use of the rich spatial context inherent to lidar data. However, they imply pre-processing: it would therefore be interesting to see if the resulting improvements could advocate for a more complex processing chain. Using both green and infrared waveforms is another solution. Indeed, since green wavelengths penetrate water and infrared do not, infrared intensity is an efficient predictor for the presence or absence of water. The bispectral neural network already shows slight improvement in the separation of marine and terrestrial habitats, although tree returns are still detected in marine waveforms. Relying on more detailed infrared information would provide better distinction of land- and sea-covers.

Nonetheless, the results of our study show that there is not a massive difference in performance between the green waveforms-based network and the bispectral configuration, which is an interesting observation. General classification is performed with an average accuracy above 92% with an error tolerance, and habitats detection in waveforms show precision, recall, accuracy and F-scores of at least 0.9 for both networks. Our approach therefore allows classification of coastal and estuarine habitats based on lidar waveforms processing, even with only one wavelength. It also obtains higher performance metrics than existent methods for similar coastal environments mapping ([10] reported a 86% accuracy for seagrass and shallow

sand mapping, and [9] obtained 90% of accuracy on their classification of coastal ecosystems). The green waveforms-based network also identifies correctly almost 100% of the water surface returns for shallow and deep waters, a promising perspective for full-waveform lidar processing for coastal and estuarine areas management. Indeed, if the approach is valid on very shallow waters, it could make full-waveform lidar analysis more accessible and limit the need for a second wavelength dataset, considering topo-bathymetric lidar data acquisition are expensive and land-water separation is traditionally made by using both green and infrared point clouds [12].

VI. Conclusion

In this study, we aimed to propose a method for full-waveform lidar data analysis with limited need for pre-processing compared to traditional decomposition and feature extraction approaches. To achieve that, we put together a labelled dataset of green waveforms and intensities for 13500 locations. We then trained two temporal convolutional networks, inspired by the UNet architecture: one processing only green waveforms, the other processing green waveforms and infrared intensities. Our networks enable accurate sea- and land-cover classification through green lidar waveform processing. By tolerating an error of ±2 samples in the detected peaks' locations (this error being consistent with the way they are recorded and the usual peak width), the waveforms classification reaches an average accuracy of 92% and 93% for the monospectral and the bispectral networks, respectively. Water surface returns were correctly identified in almost 100% of the cases by the green waveforms-based network, which also shows the best peaks detection performances. The addition of infrared intensities does not show significant improvement of waveform samples classification. However, infrared waveforms could further improve habitats identification and thus coastal and estuarine ecosystems monitoring. Consequently, our further work will focus on tackling the main limitation of our approach – the waveform returns labeling procedure, that implies post-processing of the predictions –, the use of bispectral waveform data, and the application of the method to very shallow waters. Better leveraging the strengths of topo-bathymetric lidar data could avoid classes confusion, since the bispectral network has somewhat better performances for classes detection. Adapting the approach to improve peaks delimitation would also strengthen the results. We also plan to extend the method on wider datasets to assess its robustness when producing high resolution uninterrupted mapping of coastal and estuarine areas.

Acknowledgment

M. L. would like to thank all field-work team members for their help for ground-truth data acquisition and the University of Rennes 2 for lending field work equipment.

References

[1] E.B. Barbier, S.D. Hacker, C. Kennedy, E.W. Koch, A.C. Stier, and C.R. Silliman, "The value of estuarine and coastal ecosystem services," Ecological Monographs, vol. 81, pp. 169-193, May 2011.

[2] A. Mury, A. Collin, M. Jeanson, and S. Etienne, "Wave attenuation service by intertidal coastal ecogeosystems in the Bay of Mont-Saint-Michel, France: Review and meta-analysis," in Estuaries and Coastal Zones in Times of Global Change, K. Nguyen, S. Guillou, P. Gourbesville, J. Thiébot, Eds., Springer Water, 2020, pp. 555-572.

[3] J.P. Syvitski, N. Harvey, E. Wolanski, W.C. Burnet, G.M. Perillo, V. Gornitz et al., "Dynamics of the Coastal Zone," in Coastal Fluxes in the Anthropocene, C.J. Crossland, H.H. Kremer, H. Lindeboom, J.I. Marshall Crossland, M.D.A. Le Tissied, Eds. Springer-Verlag Berlin Heidelberg, 2005, pp. 39-94.

[4] S. P. Leatherman, K. Zhang, and B.C. Douglas, "Sea level rise shown to drive coastal erosion," Eos Transactions American Geophysical Union, vol. 81, pp. 55– 57, February 2000.

[5] A. Collin, S. Dubois, D. James, and T. Houet, "Improving Intertidal Reef Mapping Using UAV Surface, Red Edge, and Near-Infrared Data," Drones, vol. 3, 67, August 2019.

[6] A. Collin, S. Dubois, C. Ramambason, and S. Etienne, "Very high-resolution mapping of emerging biogenic reefs using airborne optical imagery and neural network: the honeycomb worm (Sabellaria alveolata) case study," International Journal of Remote Sensing, vol. 39, pp. 5660-5675, May 2018.

[7] A. Collin, N. Lambert, and S. Etienne, "Satellite-based salt marsh elevation, vegetation height, and species composition mapping using the superspectral WorldView-3 imagery," International Journal of Remote Sensing, vol. 39, pp. 5619-5637, April 2018.

[8] J. Barrell, J. Grant, A. Hanson, and M. Mahoney, "Evaluating the complementarity of acoustic and satellite remote sensing for seagrass landscape mapping," International Journal of Remote Sensing, vol. 36, pp. 4069-4094, July 2015.

[9] A. Collin, B. Long, and P. Archambault, "Merging land-marine realms: Spatial patterns of seamless coastal habitats using a multispectral LiDAR," Remote Sensing of Environment, vol. 123, pp. 390-399, March 2012.

[10] M. Letard, A. Collin, D. Lague, T. Corpetti, Y. Pastol, A. Ekelund et al., "Towards 3D mapping of seagrass meadows with topo-bathymetric lidar full waveform processing," International Geoscience and Remote Sensing Symposium (IGARSS), in press.

[11] W.Y. Yan, A. Shaker, and N. El-Ashmawy, "Urban land cover classification using airborne LiDAR data: a review", Remote Sensing of Environment, vol. 158, pp. 295-310, March 2015.

[12] D. Lague and B. Feldmann, "Chapter 2 - Topo-bathymetric airborne LiDAR for fluvial-geomorphology analysis," in Developments in Earth Surface Processes, vol. 23, P. Tarolli, S.M. Mudd, Eds. Elsevier, 2020, pp. 25-54.

[13] J. Zhang, X. Lin, and X. Ning, "SVM-Based classification of segmented airborne LIDAR point clouds in urban aeras," Remote Sensing, vol. 5, pp. 3749-3775, July 2013.

[14] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Pierrot-Deseilligny, and W. Puech, "Processing full-waveform lidar data: modelling raw signals," IAPRS, vol. 36, pp. 102-107, September 2007.

[15] C.E. Parrish, I. Jeong, R.D. Nowak, and R.B. Smith, "Empirical comparison of full-waveform lidar algorithms," Photogrammetric Engineering and Remote Sensing, vol. 8, pp. 825-838, August 2011.

[16] C. Wang, Q. Li, Y. Liu, G. Wu, P. Liu, and X. Ding, "A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry," ISPRS Journal of Photogrammetry and Remote Sensing, vol. 101, pp. 22-35, March 2015.

[17] C. Mallet, U. Soergel, and F. Bretar, "Analysis of full-waveform lidar data for classification of urban areas," The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 37, pp. 85-91, July 2008.

[18] M. Perslev, M. H. Jensen, S. Darkner, P.J. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," in Proc. NeurIPS, 2019, pp. 4417–4428.

USING BISPECTRAL FULL-WAVEFORM LIDAR TO MAP SEAMLESS COASTAL HABITATS IN 3D

M. Letard [1 *], A. Collin [1] , D. Lague [2], T. Corpetti [3], Y. Pastol [4], A. Ekelund [5]

[1] Ecole Pratique des Hautes Etudes - PSL Université, Coastal GeoEcological Lab, Dinard, France - (mathilde.letard@ephe.psl.eu
[2] Univ Rennes, CNRS, Géosciences Rennes - UMR 6118 - Rennes, France - dimitri.lague@univ-rennes1.fr
[3] Université Rennes 2, CNRS UMR 6554 LETG - Rennes, France - thomas.corpetti@univ-rennes2.fr
[4] Service Hydrographique et Océanographique de la Marine - Brest, France - yves.pastol@shom.fr
[5] Airborne Hydrography AB - Jönköping, Sweden - anders.ekelund@hexagon.com

KEY WORDS: Topobathymetric lidar, Full-waveform lidar, Classification, Coastal habitats, Habitat mapping.

ABSTRACT:

Mapping coastal habitats is essential to their preservation, but the presence of water hinders seamless data collection over land-water interfaces. Thanks to its dual-wavelength and optical properties, topo-bathymetric lidar can address this task efficiently. Topo-bathymetric lidar waveforms contain relevant information to classify land and water covers automatically but are rarely analysed for both infrared and green wavelengths. The present study introduces a point-based approach for the classification of coastal habitats using bispectral waveforms of topo-bathymetric lidar surveys and machine learning. Spectral features and differential elevations are fed to a random forest algorithm to produce three-dimensional classified point clouds of 17 land and sea covers. The resulting map reaches an overall accuracy of 86%, and 65% of the prediction probabilities are above 0.60. Using this prediction confidence, it is possible to map coastal habitats and eliminate the classification errors due to noise in the data, that generate a clear tendency of the algorithm to over-estimate some classes at the expense of some others. By filtering out points with a low prediction confidence (under 0.7), the classification can be highly improved and reach an overall accuracy of 97%.

1. INTRODUCTION

The global coastal population has been growing fast for a few decades. In 2003, 41% of the world's global population lived within 100 km of the coastline (UNPD, 2005), and 21 of the 33 world's megacities were located on coastal fringes. Moreover, Martínez et al. (2007) showed that the wide diversity of ecosystems located at the interface between the Earth's oceans and continents produced 77% of the estimated economic value of the services and goods provided by ecosystems around the world in 2007, based on the method in (Costanza et al, 1997). All of these observations converge to show how ecologically, socially and economically important coastal ecosystems are (Martínez et al., 2007, Costanza et al., 1997 and Barbier et al., 2011). However, marine and terrestrial habitats are threatened by climate change and anthropic pressure (Barbier et al., 2011), and numerous studies agree that their evolution is difficult to anticipate and must be monitored to ensure continual support to littoral communities (Barbier et al., 20011, Martínez et al., 2007, Costanza et al., 1997).

Currently, the observation of coastal ecosystems without interruption between marine and terrestrial domains remains a methodological challenge due to the presence of water, which complicates their exploration with passive imagery (Kutser et al., 2020), and due to their vast diversity. Topo-bathymetric lidar is particularly suited to the task, as an active sensor able to penetrate the water surface and collect information both on ground and on the sea- or river-bed (Philpot, 2019, Lague and Feldmann, 2020). Lidar surveys of coastal areas are mostly used as 2D rasters (Wedding et al., 2008), or 3D point clouds (Hansen et al., 2021, Tulldahl and Wikström, 2012), that are classified to study given land or sea covers. Nonetheless, the origin of lidar data, which lies in signal processing of the lidar waveforms - the complete laser signal backscattered by the environment - and its potential contribution to coastal habitats monitoring remain underexplored. Lidar waveform analysis has been developed for topographic lidar data classification (Mallet and Bretar, 2009, Mallet et al., 2011, Reitberger et al., 2009, Zorzi et al., 2019), but little methodological research on the exploitation of bathymetric or topo-bathymetric waveforms exists. Yet, these data include more detailed information on the physical properties of the environment surveyed than the elevation contained in rasters or point clouds, by registering also the way the Earth's covers interact with light. This information is particularly useful to classify covers that have distinct spectral signatures (e.g., Letard et al., 2022) or subtle geometric features at decimeter scale that discrete echoes cannot capture (e.g. Launeau et al., 2018).

Exploiting lidar data requires adapted processing methods, as this sensor produces rich but complex information on the environment, with sometimes dozens of point records and several waveforms of up to a thousand of samples per square metre. Although efficient tools exist to process lidar derived rasters or point clouds, methods to efficiently exploit the knowledge enclosed by the waveforms are still expected, as they remain mainly experimental.

In this study, we propose a topo-bathymetric lidar waveform-based coastal habitat classification and explore its abilities to map 17 different types of marine and terrestrial covers. This method relies on the use of a random forest algorithm to classify features extracted from green and infrared (IR) lidar waveforms. The article focuses on addressing three main questions: (1) are topo-bathymetric lidar waveforms usable for detailed coastal habitat mapping?, (2) what is the added value of bispectral lidar data compared to a simple bathymetric lidar survey? and (3) what prediction confidence can we expect of our algorithm?

* Corresponding author

2. MATERIALS AND STUDY AREA

In this section, we briefly recall the operating principles of full-waveform lidar and describe our datasets.

2.1 Full-waveform lidar

Airborne lidar sensors emit laser pulses towards the ground and record the backscattered signal, from which two types of data can be obtained: lidar waveforms and lidar point clouds. Waveforms are time series of the complete backscattered signal, while point clouds are obtained through processing of the waveforms: each major peak in the waveform corresponds to an object encountered by the laser beam. The coordinates of these obstacles can be computed using the time range for the laser signal to travel back and forth, which enables the reconstitution of the whole scene surveyed into 3D point clouds. Topographic lidars operate with an IR laser, which is unable to travel through water surfaces. Topo-bathymetric lidars use a green laser in addition to the IR one, green lasers being able to reach the ground below the water surface. They are thus able to collect data on marine and terrestrial environments, without interruption between submerged or emerged domains.

Lidar waveforms contain important information on the physical properties of the objects encountered by the laser emitted, namely through the way light is reflected by them. They can therefore be used to map the Earth's covers (Collin et al., 2012, Letard et al., 2022). Each peak in the waveform has a different shape depending on what it originated from on the ground. Typical bathymetric waveforms have the particularity of integrating three main components: a peak produced by the water surface and elongated by a water column component, and another peak corresponding to the bottom if it is reached. Two typical waveform examples are presented in Figure 1.

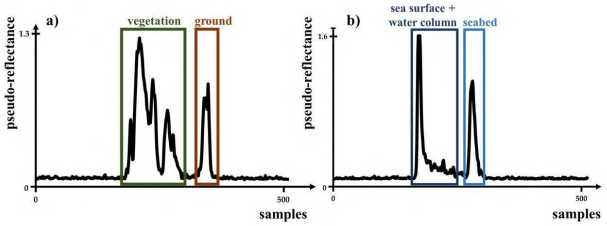


Figure 1. Typical example of a) a bathymetric waveform and b) a topographic waveform.

2.2 Lidar dataset

The lidar dataset was acquired in September 2019 for the Litto 3D® (Shom et al., 2021) project operated by the French Hydrographic Office (Shom). It was collected with a Leica HawkEye III full-waveform topo-bathymetric lidar using wavelengths at 1064 nm and 515 nm and collecting data with an elliptic pattern. Each green waveform and every 32 IR waveforms were recorded with a time frequency of 1.8 GHz, resulting in a backscattered intensity value every 556 picoseconds. The survey was led with a constant laser amplification, and the intensity of each emitted pulse linked to a waveform is available. There are on average 6.7 green waveforms and 0.5 IR waveforms per square metre. The green laser's spot size diameter is 1.8 m, while the IR laser's is 0.2 m.

2.3 Study area and ground-truth data acquisition

The study area is located on the coasts of Brittany, France in a town called Sables d'Or les Pins (48.6373, -2.4067). It features typical coastal habitats such as salt marshes, seagrasses, sandy dunes, sandy beaches, pebble beaches, pine trees or macroalgaes. It is presented in Figure 2.

A ground-truth data acquisition campaign took place in 2021 to gather knowledge on the land and sea covers in this area, through unmanned airborne vehicles (UAV) and unmanned surface vehicle (USV) RGB imagery. Figure 2 shows the areas covered by these acquisitions. The UAV used was a RGB DJI Phantom 4 Pro V2, and the USV was a PowerVision PowerDolphin. The UAV flights were calibrated with a total of 55 ground control points. 150 geolocated photoquadrats were also collected on site. Photogrammetric reconstructions of the UAV images, and an RGB orthoimage from 2014 were used to create the labelled lidar training and test datasets (necessary to perform habitat classification) via photointerpretation.

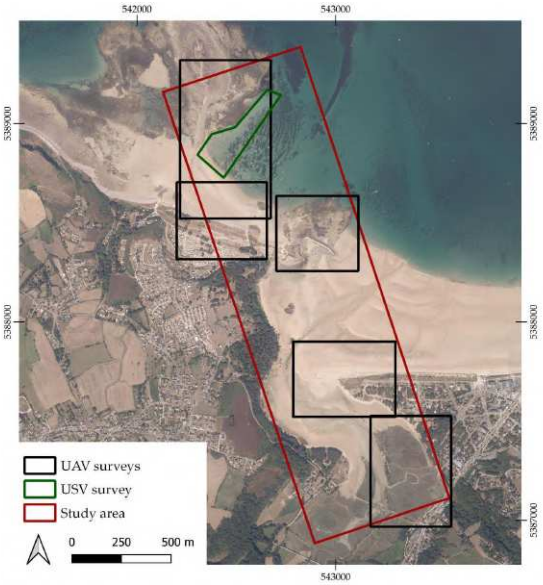


Figure 2. The area studied, and the location of the 6 ground-truth data acquisitions (WGS 84/UTM 30N).

3. METHOD

In this study, we evaluate a point-based classification method relying on the classification of features extracted from bispectral waveforms. Green and IR waveforms are first processed to extract features describing them. The attributes obtained with each wavelength are then matched in a single dataset and fed to a random forest algorithm to be classified into 16 classes. These classes are presented and illustrated in Table 1.

Class name	Illustration	Class name	Illustration
Algae		Seagrass	
Submerged sand		Submerged rock	
Rock		Pebble	

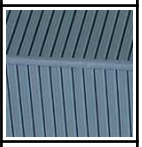
Wet sand		Surf zone	
Dry sand		Artificial ground	
Boat		Roof	
Car		Salt marsh	
Low vegetation (*lawn*)		Intermediate vegetation (*shrub*)	
Tree			

Table 1. Presentation of the land and sea covers classes studied.

3.1 Extraction of spectral features from the lidar waveforms and constitution of a bispectral dataset

To extract parameters describing the ground modelled by the waveforms, the peaks corresponding to the ground cover had to be isolated from the noise (topographic waveforms) and from the water components (bathymetric covers). All waveforms are first divided by the emitted intensity of the laser pulse they reflected. They are then smoothed using a Savitzky-Golay filter to attenuate the noise. Green and IR waveforms were not processed the same way since they do not have the same behaviour in the presence of water. In both cases, the first-derivative of the waveform was computed and thresholded in order to detect peaks in the signals and isolate the noise.
For green waveforms, if more than one peak were detected, features were computed on the group of peaks located after this first peak. The exponential decay of the lidar waveform under the influence of water was corrected as explained in (Letard et al., 2021). If only one peak was identified, features were computed on this peak.
For IR waveforms, features were computed on the group of peaks identified. The features are presented in Table 2.

Feature name	Definition
Complexity	Number of sign changes of the first derivative
Mean	Mean pseudo-reflectance
Median	Median pseudo-reflectance
Maximum	Maximum pseudo-reflectance
Standard deviation	Standard deviation of the pseudo-reflectance
Skewness	Skewness of the peak
Kurtosis	Kurtosis of the peak
Area under curve (AUC)	Area under the curve formed by the peak
Time range	Time duration of the peak
Height	Difference of elevation between the beginning and the end of the peak
Position of the maximum	Position of the maximum in the peak (in sample indices)
Difference of elevation (DZ)	Difference between the elevation of the IR return and the green return

Table 2. Definition of the features extracted in the waveforms.

These features were computed for each waveform, resulting in two sets of waveform features: a green waveform features dataset and an IR waveform features dataset. They were combined into a bispectral dataset using a nearest neighbour matching method: each green waveform features set was associated to its closest neighbour in the IR waveforms features dataset. This was made using the software "CloudCompare" (Girardeau-Montaut, 2016), which relies on the euclidean 3D distance of each point to the rest of the cloud's components to find its nearest neighbour.

3.2 Random forest classification

The features were computed as attributes of the points forming the two point clouds, in order to avoid any information loss that could occur when rasterizing the data. These point clouds were then directly classified into the 16 classes defined above.
A random forest classifier was chosen for its performance on multiclass problems implying dozens of features and its robustness to overfitting. The possibility to retrieve feature importance and prediction probability made it particularly suited to our needs. This algorithm has also been tested multiple times in 3D point clouds classification research, with consistent observations of high accuracy in land cover identification (Yan et al., 2015, Chehata et al., 2009). The random forest model employed was set to contain 150 trees and classical parameters. We used the implementation of the "scikit-learn" library (Pedregosa et al., 2011).
A set of 1000 samples of each class was used to train the model. Another set of 500 distinct samples of each class were then used to assess the quality of the model's predictions. These samples form the training and test datasets of 16000 and 8000 feature sets respectively, that are shown in Figure 3.

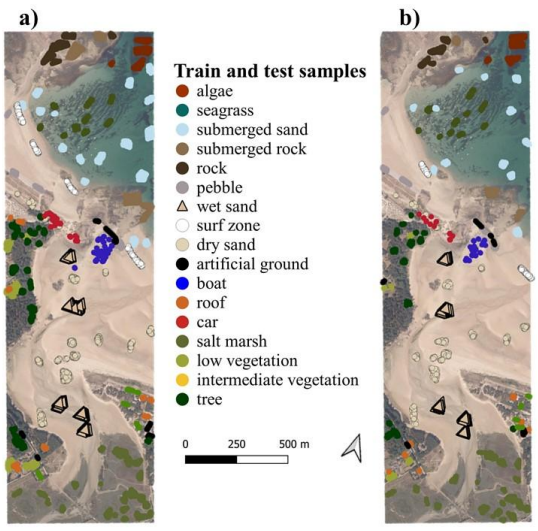


Figure 3. Distribution of the distinct a) train and b) test samples across the studied area (the natural colored imagery was acquired four years prior to the lidar survey).

3.3 Classification result assessment

To quantify the performances of our classifier, the overall accuracy (proportion of correct predictions, best when its value is 1), the average precision (proportion of correct detections of each label, best when its value is 1), the average recall (proportion of points of each label that are identified correctly, best when its value is 1) and the average F-score (combination of precision and recall, best when its value is 1) were computed for each classification on the test dataset (data that were never seen by the algorithm during training).
A class-wise analysis was also performed, by detailing the metrics obtained for each class for each experiment.

3.4 Feature selection

To avoid potential negative feedback on the classification accuracy due to information redundancy among the 24 predictors, an importance analysis was performed to select the most relevant attributes for the final habitat map. To evaluate the contribution of each predictor to the overall classification accuracy, they were each successively dropped to compare the overall accuracy obtained without them to the reference accuracy obtained on the complete set of features. The predictors that contributed negatively to the classification accuracy were removed from the classification attributes.

3.5 Production of a 3D habitat classification

The output of our method consisted in a set of labelled waveform features vectors. Using the coordinates of each waveform, this dataset was turned into a 3D point cloud of the terrestrial and marine habitats of Sables d'Or les Pins.

4. RESULTS

4.1 Results obtained with different sets of predictors

Classification experiments were led on five different sets of predictors in order to evaluate their relevance and added-value. These five sets are the following:

- all green waveform features (11 features)
- all IR waveform features (11 features)
- DZ (one feature only)
- green and IR waveform features (22 features)
- all waveform features, plus DZ (23 features)

The performances of the classifications obtained for these sets are presented in Table 3.

Model	OA	Precision	Recall	F-score
Green	0.823	0.825	0.823	0.821
IR	0.315	0.309	0.314	0.292
DZ	0.216	0.218	0.215	0.216
Green + IR	0.846	0.85	0.846	0.842
Green + IR +DZ	0.848	0.854	0.848	0.843

Table 3. Classification performances for different predictors.

Overall, the most relevant predictors for the classification of coastal land and sea covers are descriptors of the green lidar waveforms. IR data or differential elevation values appear to misclassify more than two thirds of the points.

4.2 Classification performances difference between both wavelengths

Since each wavelength of the topo-bathymetric lidar was designed to survey a specific type of environment (IR laser for topography and green laser for bathymetry), we performed an in-depth analysis of the classification results obtained using successively green data only and green + IR data. This analysis features the mean precision and recall obtained for each class. It allowed us to observe which type of habitat was best described by each wavelength. The mean prediction confidence obtained for each point is also taken into account, in order to better understand the potential classification errors and what they imply. This prediction confidence corresponds to the probability of membership of each point to the class it was assigned. The complete class-wise analysis conducted for the green model, and the Green + IR + DZ model are presented in Figure 4.

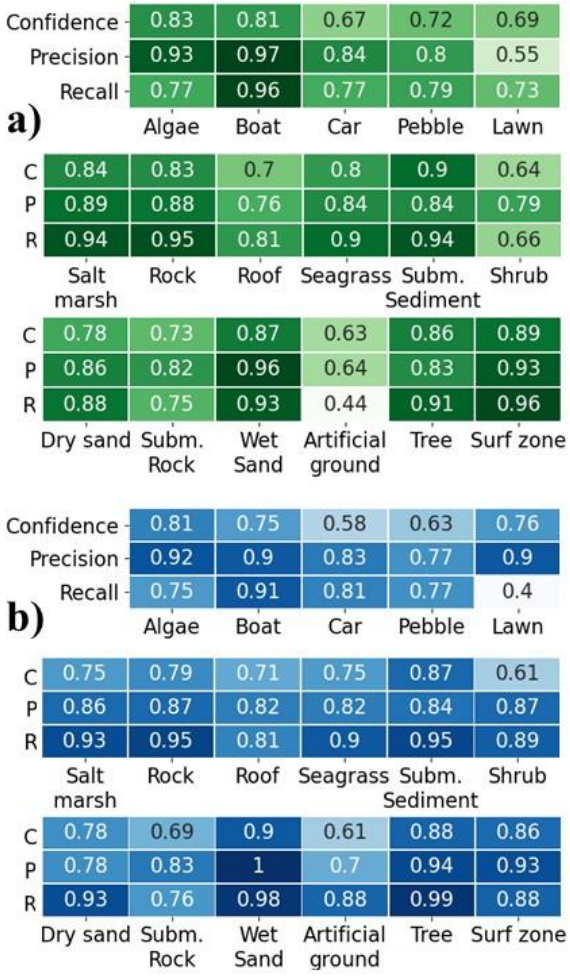


Figure 4. Class-wise classification metrics obtained when using a) green and b) bispectral waveform features as predictors.

The use of green waveform parameters produced an accurate labelling, though the algorithm showed weaker performances on topographic classes as *lawn* and *artificial ground*. The combination of both wavelengths and DZ produces a more accurate result and improves the recall of every class except low vegetation (*lawn*). Class-wise recall and precision values reveal that some classes were overestimated at the expense of others. This is the case of *submerged rock*, *algae* and *lawn*, which have lower recall values than precision.

4.3 Bispectral dataset classification

To generate the final coastal habitat classification, we analysed the contribution of each feature to the overall accuracy, and excluded from the predictors the features that impacted it negatively: green waveforms' skewness, IR waveforms' AUC, IR waveforms' skewness, IR waveforms' maximum and IR waveforms' mean. The metrics and the map obtained with the final set of features are presented in Table 5 and Figure 5.

OA	Precision	Recall	F-score
0.856	0.862	0.856	0.852

Table 5. Performance metrics obtained after selecting the features based on their contribution to the classification.

Selecting attributes based on their importance makes the overall accuracy reach 86%. Globally, the classifier's tendency to overestimate *submerged rock* or *algae* identified with the low recall in Figure 4 is observable in Figure 5, where these classes respectively invade the seagrass meadow and the surf zone.

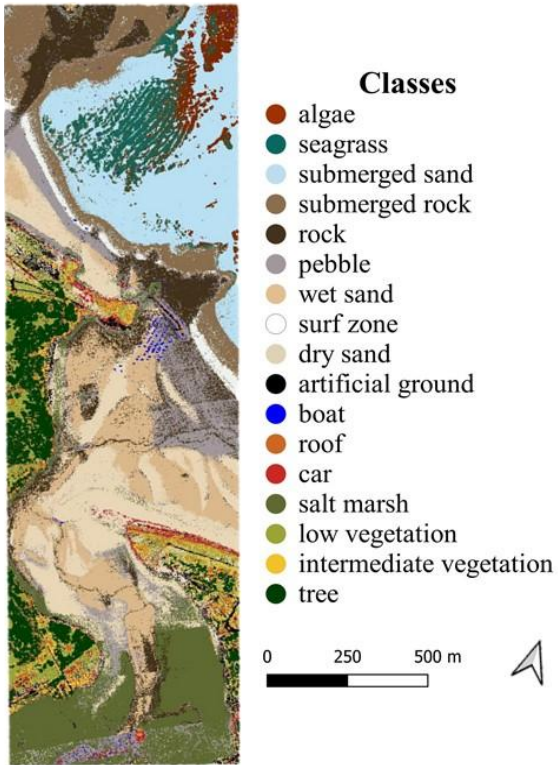


Figure 5. Coastal habitats map obtained when classifying a selected set of bispectral waveform features.

Other obvious confusions exist between *rock* and *dry sand*, or *submerged rock* and *surf zone* or *submerged sediment*. They are also revealed by the precision and recall values of these classes in Table 4. One of the main confusions is between *pebble* and sand; yet it is not as clearly quantified by the metrics.

4.4 Prediction confidence analysis

To further assess the abilities of our method to classify land and sea covers, we analysed the prediction confidence across the studied area. Figure 6 shows the maps obtained when setting a confidence threshold, below which the points are labelled as *unclassified*. Most points are kept with a threshold set at 70% (which means the probability that the point belongs to the class it was given is at least 70%). This is in line with the mean confidence of 77% obtained on the test dataset (see Figure 4). However, when the threshold is increased at 90%, more complex areas, mainly at the interface between different classes, disappear, as they are classified with a lower confidence.

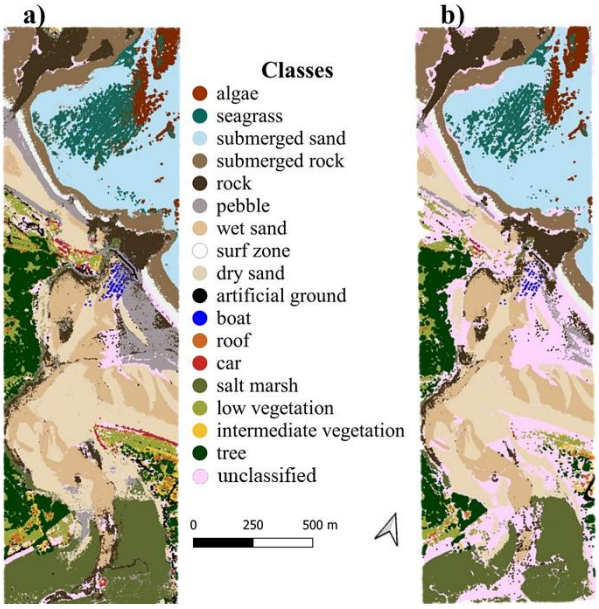


Figure 6. Land and sea covers map obtained at a) a 70% confidence level and b) a 90% confidence level.

A closer look at the misclassified samples shows that the confidence level is globally lower for them. Indeed, on the test dataset, samples that were wrongfully classified have a median confidence of 47% with a standard deviation of 17%. The overall accuracy values obtained when filtering the points based on their confidence predictions confirm that misclassified points can be discarded using this criteria: Table 6 presents the accuracies obtained for different confidence thresholds.

Threshold	0.6	0.7	0.8	0.9
OA	0.95	0.97	0.98	0.99

Table 5. Overall accuracy of the resulting classification depending on the prediction confidence threshold.

5. DISCUSSION

5.1 Usability of full-waveform lidar for coastal habitat mapping

The final result obtained confirms the observations of (Mallet et al., 2011, Letard et al., 2022, Collin et al., 2012) and the potential of lidar waveforms for classification tasks. Here, a single dataset made the classification of 17 different land and sea covers possible with high accuracy (86%, see Table 4). The resulting 3D map is presented in Figure 7. It is dense and has a high spatial resolution, suited to the realisation of ecological assessments such as ecosystem services evaluation, as performed in Martínez et al. (2007) and Costanza et al. (1997). Considering the thematic objectives of this research, the results are promising, since all habitats that provide goods and services are described with a relatively high precision (on average, 87%, see Figure 4) and in 3D, contrary to other methods developed in existing papers (Collin et al., 2012, Letard et al., 2021).

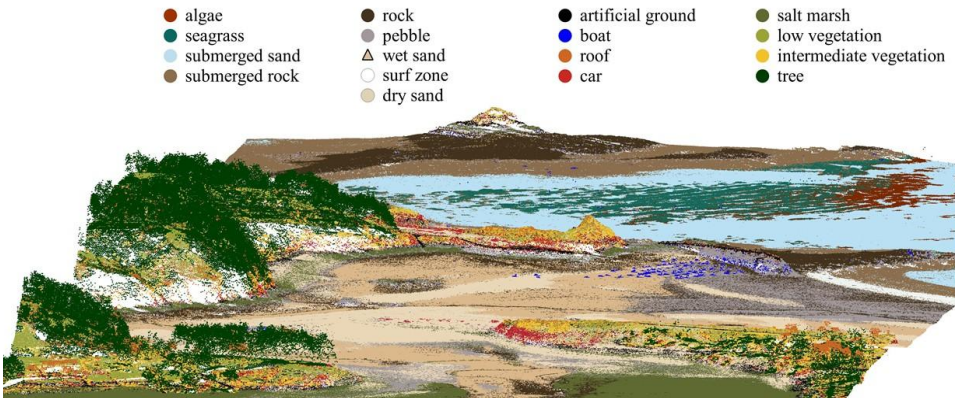


Figure 7. Resulting 3D point cloud of a coastal area and its habitats classified with full-waveform topo-bathymetric lidar

One of the main objectives of this study was to develop a method for seamless spatial modelisation of marine and terrestrial habitats. Here, the land-water continuum is classified without interruption, which is a key methodological aspect of coastal habitat monitoring (Collin et al., 2012). Figure 7 shows that the output of the classification produces an uninterrupted restitution of the land-water interface. However, our approach has limitations that necessitate further investigation: the classifier has a tendency to overestimate *algae* at the expense of *seagrass*, and *pebble* at the expense of sand, for example, as observed in previous research (Letard et al. 2022). Even though these classes are close semantically and can have similar waveform signatures due to their size, physical properties, and texture relative to the laser spot size, a better distinction is needed for ecological applications. *Boat* and *car* are also often falsely detected. Figures 5 and 7 illustrate well this issue, as a great number of points are classified as *car* on land, and part of the boulders of the dyke are labelled as *boat*. Figure 8 focuses on the sandy dune, which features many false detections of *car*.

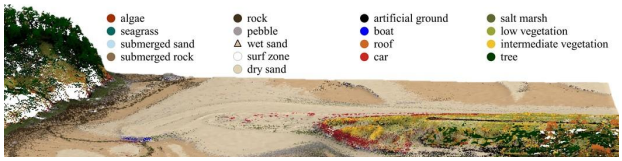


Figure 8. Extract of the resulting 3D classification: the sandy dune of Sables d'Or les Pins and its surroundings.

These errors may be explained by the albedo of cars and boats, which can vary considerably between two different types of vehicles. The wide colour spectrum they can have is difficult to model in a training dataset which can lead the classifier to learn confused information. The procedure to adopt with such classes in order not to compromise the wider objective - ecological monitoring - should be further discussed. These classes could be merged in a more global *vehicles* class, or an *unclassified* class could be added to handle unusual feature vectors and avoid their detection at the expense of natural habitats mapping, as in Figure 8 where the sandy dune - a key ecosystem - is mapped as *car*. Despite these mistakes, the possibility to have 3D assessments of the spatial repartition of different ecosystems already bears encouraging perspectives for coastal ecology, and could enhance results outlined by previous studies using rasterized topo-bathymetric lidar data (Wedding et al., 2008, Collin et al., 2012). The highly informative content of lidar waveforms, already stated in existing research on the topic (Mallet et al., 2011, Collin et al., 2012) is illustrated by their constant contribution to the resulting classification accuracy (Table 3). Lidar waveforms contain enough information to describe surface covers, despite their lack of information on their neighbourhood's geometry and spatial repartition (Table 3). This loss of spatial context information - compared to point cloud classification techniques involving neighbourhoods (Brodu and Lague, 2012) - is also one of the strengths of waveform-based processing. It avoids spatial averaging of information that can result in classification artefacts depending on the neighbourhood radius defined (Brodu and Lague, 2012). Classifications based on waveforms and not on spatial context may consequently gain in horizontal resolution, keeping in mind the influence of the laser's footprint diameter (Letard et al., 2022). However, both wavelengths do not perform equally, as expected since they were each specifically designed for different environments (Philpot, 2019). Infrared waveforms cannot be used alone to study both dry and wet environments: the metrics obtained for the IR model (Table 3) quantify the limitations of topographic lidar for the survey of highly diverse environments, and corroborates previous research results (Letard et al. 2022). Green lidar waveform features perform better: their classification reaches 82% of overall accuracy, and similar values of precision and recall. Their ability to label some classes of ground is limited, which is why dual-wavelength datasets are relevant.

5.2 Contribution of the bispectral information

The combination of infrared and green datasets produces the most accurate classification overall, as in (Letard et al., 2022). The addition of the infrared features to the green parameters improves the distinction of the different types of grounds, and results in an increase of 3% of the overall accuracy, precision and recall. This can be explained by the properties of the lidar system: the infrared laser modelises a more concise and precise surface, while the larger size of the green laser may sometimes result in the mixing of different land covers into one single return, but is able to penetrate through water (Philpot, 2019). We suggest that the combination of two types of spot sizes and wavelengths optimises the information collected on a given area in terms of albedo, water content and surface rugosity, which all impact the waveform and characterise natural surfaces.

Having two different lasers also provides a good understanding of the vertical complexity of the scene (Collin et al., 2012). Surfaces are sampled with varying sizes of laser beams thanks to the dual-laser system. A wider footprint hits a wider portion of surface at a time: it can mix information about several layers of covers and create intermediate points between the canopy and the ground. This phenomenon is documented in all laser scanning systems (Brodu and Lague, 2012). The smaller footprint of the infrared laser may not alway penetrate through dense covers, but creates less mixed points. The combination of

both sources of data results in local differences of elevation and a greater vertical density over more complex surface covers and non planar areas. Bispectral lidar thus gives a more thorough review of the vertical structure of the environment, which can explain the improvement of the classification accuracy (+0.2%) when adding DZ to the predictors. Using DZ is also a way of including spatial context data, which is particularly contributive in point clouds classification (Brodu and Lague, 2012), to the model, without making it too dependent on the training area or involving neighbours.

Finally, the few predictors discarded after the importance analysis - green waveforms' skewness, IR waveforms' AUC, skewness, maximum and mean - show that both wavelengths contain relevant information. They also confirm the theoretically more exhaustive nature of the green waveform. Indeed, infrared waveforms seem to contain less essential details on the surveyed scene, as most of the features dropped after this step concerned the characteristics of the infrared return, whereas nearly all descriptive parameters of the green returns were useful to the random forest model.

5.3 Classification algorithm and prediction confidence

Overall, our observations corroborate existing research on the classification of lidar data using random forest algorithms (Yan et al., 2015, Hansen et al., 2021, Letard et al., 2022). As documented in (Letard et al., 2022), the model was quickly applied to a great number of features and points, and shows a good ability to exploit information and detect the 17 classes. The average prediction probability is 77%. This criteria and the qualitative results in Figures 5 and 6 outside the test dataset show that the classifier is robust to overfitting and is generalisable to a wider scene. The average prediction confidence being high, it can be used to filter the results and preserve the overall quality of the map even if it means compromising on the 3D density in some areas, as introduced by Brodu and Lague (2012). Table 5 and Figure 6 show how points can be removed by applying a confidence criteria in order to prioritise solid predictions, and thus improve the overall accuracy, as in Brodu and Lague (2012). The points affected by this filtering step give indications on the strengths and weaknesses of the random forest model. Misclassified points with high prediction probability are evidence of training errors. This is the case of the false detections of *submerged rock* along the surf zone, which Figure 9 shows more clearly. In Figure 9, the false detections of *submerged rock* are located along the *surf zone* points. The submerged rocks detected next to the rocks in the foreground are true positives. These mislabelled points have a confidence value higher than 90%, as Figure 6 b) reveals, which means that their descriptive features correspond to the usual range of statistics describing *submerged rock*. In this case, further investigation on the most distinctive feature of *submerged rock* could help identify the origin of the issue. A hypothesis could be that the similar DZ (i.e. water depth) of all the samples of *submerged rock* in the scene introduces a bias that causes the surf zone to be confused for *submerged rock*.

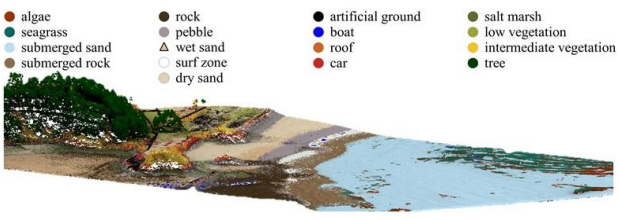


Figure 9. Extract of the resulting 3D classification: the land-water interface and its habitats.

The prediction confidence is also lower for areas at the interface between very distinct environments (Figure 6.b), where the waveforms might integrate mixed information due to the size of the laser spots. The same issue was documented in Brodu and Lague (2012) in 3D terrestrial laser scanning data classification. These areas were less documented in our training data, and the prediction probability underlines this lower confidence.

On the other hand, errors made with low confidence reveal poorly represented ranges of values among the training dataset, that are difficult to place in the possible labels. They reveal less about the training process than about the quality of the dataset. Brodu and Lague (2012) identify that some areas of 3D point clouds can have lower classification confidence due to a smaller point density compared to the rest of the scene. In their case, the classifier relies on geometrical features, and this difference in the dataset's constitution explains why their features are not as distinctive. In our case, outliers and erratic values of intensity or elevation in the data can be difficult to classify correctly, since they are not included in the training data, that are rigorously selected for their representativity of each class. In this respect, our results illustrate how important the quality of the lidar dataset is for classification tasks. Vertical or radiometric calibration issues can severely impact the possibilities to detect the nature of the surface, as we can see in Figure 5, and as suggested in Brodu and Lague (2012). Flight lines can be recognized through the classification obtained around the salt marsh and on the beach between the boat moorings and the dune. In our dataset, these lines had saturated intensities difficult to interpret, and the elevations at the interface between flight lines differed by several centimetres. The possibility to analyse the prediction confidence is thus a great indicator of the potential biases in the method. It also offers the possibility to improve the classification accuracy (Brodu and Lague, 2012): more than 10% of OA can be gained thanks to confidence-based filtering (see Table 5).

6. CONCLUSION AND OUTLOOK

In this article, we presented a method to map terrestrial and underwater habitats using a single source of data: full-waveform topobathymetric lidar. Our results show how fit this sensor is to survey diverse environments: using a random forest algorithm, we obtain classification accuracies above 85% with a mean prediction confidence of 77%. By filtering predictions depending on their classification confidence, the quality of the resulting map can be increased by up to 10%. Computing the prediction confidence also gives an interesting insight on the origin of the classification errors, which either reveal training issues, or erratic values in the initial dataset. These could be tackled with a deeper analysis of the most descriptive features for each class, and with particular attention to the calibration of the dataset and the definition of the classes during training. In the end, we obtain an interesting 3D map of 17 different land and sea covers that has potential for future ecological assessments. The presence of vertical structure information through the different spatial repartition of the points between the two wavelengths makes this result encouraging for the ecological monitoring of coastal areas. Indeed, combining vertical structure information to the knowledge waveforms provide on the physical properties of the environment could serve as ecosystem services proxy data. Lidar waveform processing also enables a finer horizontal resolution by avoiding spatial averaging of information. In future work, classifying each peak in the waveforms independently could improve the vertical density of the map to further help with ecosystem services valuation.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of the Saur Group, the Région Bretagne ARED program and the Nantes-Rennes topo-bathymetric lidar platform. They are also grateful to those who helped with data acquisitions.

REFERENCES

Barbier, E.B., Hacker S.D., Kennedy C., Koch E.W., Stier A.C., Silliman C.R., 2011: The value of estuarine and coastal ecosystem services. *Ecological Monographs*, 81, 169-193.

Brodu, N., Lague, D., 2012. 3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology. *ISPRS Journal of Photogrammetry and Remote Sensing* 68, 121–134. https://doi.org/10.1016/j.isprsjprs.2012.01.006

Chehata, N., Guo, L., Mallet, C., 2009: Airborne lidar feature selection for urban classification using random forests. In *Proceedings of the Laserscanning*, Paris, France, 1-2 September 2009.

Collin A., Long B., Archambault P., 2012: Merging land-marine realms: spatial patterns of seamless coastal habitats using a multispectral LiDAR. *Remote Sensing of Environment*, 123, 390-399.

Costanza, R., d'Arge, R., De Groot, R., Farber, S., Grasso, M., Hannon, B., Van Den Belt, M., 1997: The value of the world's ecosystem services and natural capital. *Nature*, 387(6630), 253-260.37.

Girardeau-Montaut, D., 2016: CloudCompare. Available online: https://www.danielgm.net/cc/ (accessed on 1 March 2022).

Hansen, S.S., Ernstsen, V.B., Andersen, M.S., Al-Hamdani, Z., Baran, R., Niederwieser, M., Steinbacher, F., Kroon, A., 2021: Classification of Boulders in Coastal Environments Using Random Forest Machine Learning on Topo-Bathymetric LiDAR Data. *Remote Sensing* 13, 4101, doi:10.3390/rs13204101.

Kutser T., Vahtmäe E., Martin G., 2006: Assessing suitability of multispectral satellites for mapping benthic macroalgal cover in turbid coastal waters by means of model simulations. *Estuarine, Coastal and Shelf Science*, 67, 521-529.

Lague, D., Feldmann, B., 2020: Chapter 2—Topo-bathymetric airborne LiDAR for fluvial-geomorphology analysis. In *Developments in Earth Surface Processes*, Tarolli, P., Mudd, S.M., Eds., Remote Sensing of Geomorphology, Elsevier: Amsterdam, The Netherlands, Volume 23, pp. 25–54.

Launeau, P., Giraud, M., Ba, A., Moussaoui, S., Robin, M., Debaine, F., Lague, D., Le Menn, E., 2018. Full-Waveform LiDAR Pixel Analysis for Low-Growing Vegetation Mapping of Coastal Foredunes in Western France. *Remote Sensing* 10, 669. https://doi.org/10.3390/rs10050669

Letard M., Collin A., Lague D., Corpetti T., Pastol Y., Ekelund A., Pergent G., Costa S., 2021: Towards 3D mapping of seagrass meadows with topo-bathymetric lidar full-waveform processing. *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, 8069-8072.

Letard, M., Collin, A., Corpetti, T., Lague, D., Pastol, Y., Ekelund, A., 2022: Classification of Land-Water Continuum Habitats Using Exclusively Airborne Topobathymetric Lidar Green Waveforms and Infrared Intensity Point Clouds. *Remote Sensing* 14, 341. https://doi.org/10.3390/rs14020341.

Mallet C., Bretar F., 2009: Full-waveform topographic lidar: State-of-the-art. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64, 1-16.

Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, Advances in LIDAR Data Processing and Applications 66, S71–S84. https://doi.org/10.1016/j.isprsjprs.2011.09.008.

Martínez, M.L., Intralawan, A., Vázquez, G., Pérez-Maqueo, O., Sutton, P., Landgrave, R., 2007: The coasts of our world: Ecological, economic and social importance. *Ecological Economics*, Ecological Economics of Coastal Disasters 63, 254–272. https://doi.org/10.1016/j.ecolecon.2006.10.022.

Philpot, W., 2019: *Airborne Laser Hydrography II*. doi.org/10.7298/jxm9-g971.

Reitberger, J., Schnörr, Cl., Krzystek, P., Stilla, U., 2009: 3D segmentation of single trees exploiting full waveform LIDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing* 64, 561–574. https://doi.org/10.1016/j.isprsjprs.2009.04.002.

Shom, IGN, Etat Français, Région Bretagne, DREAL Bretagne, FEDER, 2021: Litto3D® Bretagne. Soon available online: https://diffusion.shom.fr/pro/risques/altimetrie-littorale.html.

Tulldahl, H.M., Wikström, S.A., 2012: Classification of Aquatic Macrovegetation and Substrates with Airborne Lidar. *Remote Sens. Environ.*, 121, 347–357, doi:10.1016/j.rse.2012.02.004.

UNDP, 2005: Human Development Report 2005: International Cooperation at a Crossroads; Aid, Trade and Security in an Unequal World. New York, USA: UNDP.

Wedding, L.M., Friedlander, A.M., McGranaghan, M., Yost, R.S., Monaco, M.E., 2008: Using Bathymetric Lidar to Define Near-shore Benthic Habitat Complexity: Implications for Management of Reef Fish Assemblages in Hawaii. *Remote Sens. Environ.*, 112, 4159–4165, doi:10.1016/j.rse.2008.01.025.

Yan, W.Y., Shaker, A., El-Ashmawy, N., 2015: Urban Land Cover Classification Using Airborne LiDAR Data: A Review. *Remote Sens. Environ.*, 158, 295–310, doi:10.1016/j.rse.2014.11.001.

Zorzi, S., Maset, E., Fusiello, A., Crosilla, F., 2019: Full-Waveform Airborne LiDAR Data Classification Using Convolutional Neural Networks. *IEEE Transactions on Geoscience and Remote Sensing* 57, 8255–8261. https://doi.org/10.1109/TGRS.2019.2919472.

BATHYMETRIC LIDAR WAVEFORM DECOMPOSITION WITH TEMPORAL ATTENTIVE ENCODER-DECODERS

Mathilde Letard[1], Thomas Corpetti[2] and Dimitri Lague[1]

[1] CNRS, Univ Rennes, UMR 6118 Geosciences, Rennes, France
[2] CNRS, UMR 6554 LETG, F-35000 Rennes, France

ABSTRACT

This paper is concerned with the decomposition of bathymetric lidar waveforms. Because of the presence of water, processing such data remains a challenge since water impacts their shape and signal-to-noise ratio, depending in particular on the associated turbidity. In this paper, we explore the use of attentive autoencoders to decompose bathymetric waveforms and recover their air/water interface, water column, and water bottom components simultaneously, without relying on assumptions about the impulse or target surface nature. On simulated waveforms, the method achieves lower decomposition error than existing approaches, handling overlapping echoes of very shallow waters and weak returns in deeper water. This opens to attractive strategies to process real bathymetric waveforms.

Index Terms— full-waveform lidar, bathymetric lidar, signal decomposition, waveform decomposition, attention UTime

1. INTRODUCTION

Though airborne LiDAR data are mainly used as 2D products or 3D point clouds, their origin lies in the processing of backscattered laser signals and their conversion into spatial information. The precise extraction of the various signal components linked to the hit targets is thus crucial to interpreting with precision the observed scene. However, this extraction entirely relies on the ability both to process the noise efficiently and to remain sensitive enough to subtle variations denoting low-reflective or vertically close targets. Research on full-waveform lidar data processing revealed the potential of finer analysis to retrieve additional targets and increase the resulting point clouds' density [1]–[3]. However, handling large datasets while minimizing the level of false detections – which produce noise in the point clouds and complexify their interpretation – is still a challenge, namely for airborne bathymetric lidar (ALB). ALB uses green lasers to penetrate water and produce bathymetric data. The resulting waveforms have two main peaks: one generated by the water surface and another occurring when, and if, the water bottom is hit [4], [5]. ALB signals have to compose with additional sources of noise due to the optical scattering of the laser beams in the water column, which 1) exponentially attenuates the reflected intensity of the water bottom and 2) reflects a portion of the signal towards the receptor without signifying a solid target, resulting in a third waveform component that elongates the surface peak [4], [5]. In addition, extracting meaningful information from waveforms is not trivial because of the noise that disrupts the signal and gets mixed up with weak information, which is often lost during processing. In this paper, instead of a deconvolution by inversion of the pulse response function that can be challenging due to the noise in the signal, we explore the use of deep neural networks to tackle these problems and perform bathymetric waveform decomposition. We use a temporal attentive autoencoder to reconstruct the three main components of the bathymetric lidar signal simultaneously: the water surface, the water column, and the bottom. Our method targets three main limitations of existing waveform processing methods: (1) it does not rely on an iterative process, and (2) it simultaneously recovers the three main bathymetric waveform components.

2. WAVEFORM PROCESSING METHODS

The collected waveform is a sum of sub-signals that each result from a convolution between the emitted pulse and the target surface function [6]. Traditionally, waveform processing is thus made either with **deconvolution** methods [7], [8] or decomposition procedures relying on the **fitting of mathematical functions** to previously detected peaks in the signal [6], [9]. These approaches rely on various assumptions about the components' impulse function or nature. In [10] and [11], the authros show that waveforms can be considered as a sum of Gaussians corresponding to the objects in the illuminated cone. The emitted pulse and the target functions are thus often approximated with Gaussian functions of fixed width or more asymmetric shapes to fit the empirical sensor impulses [6], [12], [13]. When using the decomposition approach, Gaussian mixture models are often used to estimate iteratively the parameters of these Gaussian or other predefined functions, which are then used as proxies on the nature of the targets [14]. Deconvolution procedures include, among others, the Richardson-Lucy deconvolution and the Wiener filter deconvolution. Several comparative studies [8], [15], [16] show the superiority of Richardson-Lucy

deconvolution in retrieving weak returns and recovering the amplitude of the different components but underline its costly iterative process. Several limitations appear with these classical approaches:

(1) Decomposition methods relying on the fitting of mathematical functions are not always adapted to the presence of the water column component, which has an entirely different shape;
(2) Target surface functions may not always be similar/gaussian, and returns may be poorly approximated or mixed;
(3) The denoising methods or post-processing procedures used with these approaches to avoid false detections may lead to loss of information.

3. PROPOSED DECOMPOSITION NETWORK

We adapt U-Time, the 1D version of U-Net introduced in [17] and applied to bathymetric lidar waveforms in [18]. We use the resulting encoder-decoder structure to output three waveform components, as illustrated in Figure 1.

The encoder is a stack of 4 identical blocks comprising two series of 1D convolution, rectified linear unit activation, and batch normalization. Each block is followed by a max pooling operation of size 2, thereby reducing by half the length of the feature vector passed on to the next step. In addition, at each step, the number of convolution filters, initially set to 16, is doubled to increase the network's abstraction ability.

The decoder is a stack of four blocks with a different structure. Each block starts with a cross-attention mechanism, followed by up-sampling of size 2, a typical U-Net skip connection, and two series of 1D convolution, rectified linear unit activation, and batch normalization. Contrary to the encoder, the number of filters at each block is reduced by half, thus returning to 16 in the last convolution layers.

Additive attention, as introduced in [19], is used to better pass information about the waveform across the network and allow the model to focus on features from different representation spaces in the decoder, thereby reconstructing waveform components more effectively. Using this setup, each sample of the decoder layer attends to each sample of the corresponding encoder layer.

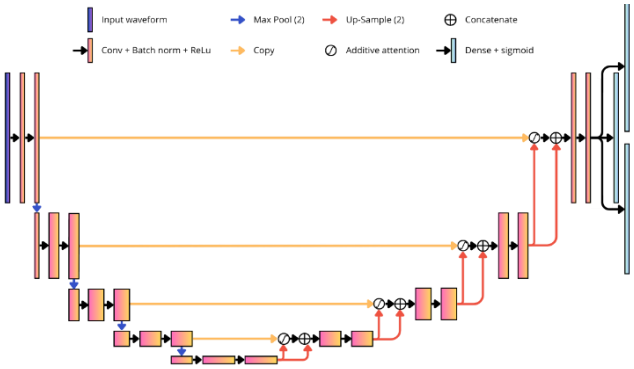


Figure 1. Illustration of the proposed waveform decomposition network architecture.

Between the encoder and the decoder, two series of 128 convolution filters with ReLu activation and batch normalization are applied to the 16 samples-long embedding of the input waveform.

At the end of the decoder, three dense layers with sigmoid activation allow the network to output three sequences corresponding to the three waveform components.

To constrain the network to learn both amplitude and shape relevance when reconstructing the waveform components, we designed a custom loss function, relying on the Mean Absolute Error (MAE) and the Bhattacharyya distance.

$$loss = \left(\frac{1}{N}\sum_{i=0}^{N} abs(pred - true) \times (1 + true)\right) + \left(1 - \sum\sqrt{\frac{pred}{\int pred} \times \frac{true}{\int true}}\right) \quad (1)$$

where N is the number of samples in the waveform, $pred$ is the predicted sequence, and $true$ is the true sequence. Roughly, the idea of the first term is to ensure global consistency in the estimations, this consistency being balanced by the waveform intensity (more weight is applied for locations with high intensities); the second term rather focuses on shapes of the decomposed waveforms. Training is performed on batches of 100 waveforms using an Adam optimizer with a constant learning rate of 0.0001. The model was trained for 500 epochs. One training epoch took, on average, 32 seconds on an Nvidia Quadro RTX 5000 GPU.

4. BATHYMETRIC WAVEFORMS SIMULATION

To simplify the training process (that requires pairs of input waveforms and associated decompositions), to evaluate the approach, and to fully control the physical conditions and their output, we use simulated data obtained with a bathymetric waveform simulator [20]. For training, validation and testing, we simulated three sets of 15000, 3700, and 10000 waveforms, respectively. In each dataset, five types of environments were simulated by varying the simulation parameters (the water body's depth and turbidity). Different impulse functions were used to generalize the method to different sensors and a systematic noise is added. The parameters used are summarized in Table 1.

Parameter	Value(s)	Parameter	Value(s)
Depth (m)	[0.15, 20]	Pulse amplitude	[0.05, 1]
Kd (m^{-1})	[0.08, 0.85]	Pulse width	[0.1, 1]
Bottom reflectance	[0.03, 0.85]	Pulse type	(Generalized) extreme value
Incidence on surface (rad)	[0, 0.8]	Incidence on bottom (rad)	[0, 1.35]
Vertical resolution (m)	0.0626		

Table 1. Parameters used to simulate lidar waveforms.

5. EVALUATION METRICS AND COMPARISON

To evaluate the performance of our method, we compare each retrieved component to its simulated equivalent.
For each of the three components, three metrics are computed. They are defined in equations (2) – (4), where N is the waveform length, $pred$ the predicted sequence, $true$ the true sequence, and KL the Kullback-Leibler divergence.

The weighted MAE, quantifies the mean element-wise amplitude error:

$$wMAE = \frac{1}{N}\sum_{i=0}^{N} abs(pred - true) \times (1 + true) \qquad (2)$$

The Kullback-Leibler divergence measures the relative entropy, or difference in information that two distributions contain. As it is not symmetric, we use a custom metric:

$$mKL = \frac{KL\left(\frac{pred}{\int pred}, \frac{true}{\int true}\right) + KL\left(\frac{true}{\int true}, \frac{pred}{\int pred}\right)}{2} \qquad (3)$$

The Bhattacharyya distance quantifies how far two samples of two distributions are from each other in time and amplitude. For consistency purposes, we used the metric B':

$$B' = 1 - \sum \sqrt{\frac{pred}{\int pred} \times \frac{true}{\int true}} \qquad (4)$$

These three metrics range between 0 and 1, where 0 indicates high similarity and 1 indicates high dissimilarity.

We compare our results to Gaussian Mixture Model (GMM) decompositions with two components – as the water column is not Gaussian, and as in [15] – and a Lucy-Richardson (RL) deconvolution. Other methods will be tested in the future.

6. RESULTS

Results obtained with the best model and with the classical approaches on the test data are presented in Tables 2 and 3.

Waveform component	Median wMAE	Median mKL	Median B'
Total	0.00333	0.00056	0.00028
Surface	0.00420	0.00040	0.00020
Column	0.00302	0.00056	0.00028
Bottom	0.00333	0.00074	0.00037

Table 2. Results obtained with the proposed approach.

Waveform component	GMM	RL
Total	0.235	0.122
Surface	0.147	/
Column	/	/
Bottom	0.098	/

Table 3. Results obtained with state-of-the-art approaches

Overall, the autoencoder reaches lower wMAEs than the GMM and RL methods, although it has higher amplitude errors than distribution distances, which shows that it tends to estimate component shapes better than their amplitude. Analyzing the metrics more deeply shows that there is no relationship between the value of wMAE, mKL and B' and the turbidity of the simulated water body. It also highlights the higher decomposition errors when processing waveforms at depth below 0.5 m. Qualitative results, presented in Figure 2 illustrate these observations: in Figure 2.a), which illustrates very shallow waters, the gap between the prediction and the simulation is broader than in Figures 2.b) and 2.c), and the amplitude error is noticeable, while the peaks appear close to their true locations. Figure 2.c) shows the ability of the network to recover very weak returns in deeper waters. Visual analysis of the corresponding GMM and RL results show that RL misses the bottom return in the cases presented in Figure 2.a) and mistakes noise for returns in the 2.c) one, while GMM with two components retrieves two returns in the first case, and falsely decomposes the surface return into two echoes in the second.

7. DISCUSSION AND CONCLUSION

Our results outline the potential of temporal attentive convolutional neural networks for lidar waveform decomposition. Using deep neural networks has the advantage of not necessitating setting fixed parameters. Compared to GMM and RL it is possible to reconstruct the three components of bathymetric waveforms simultaneously.

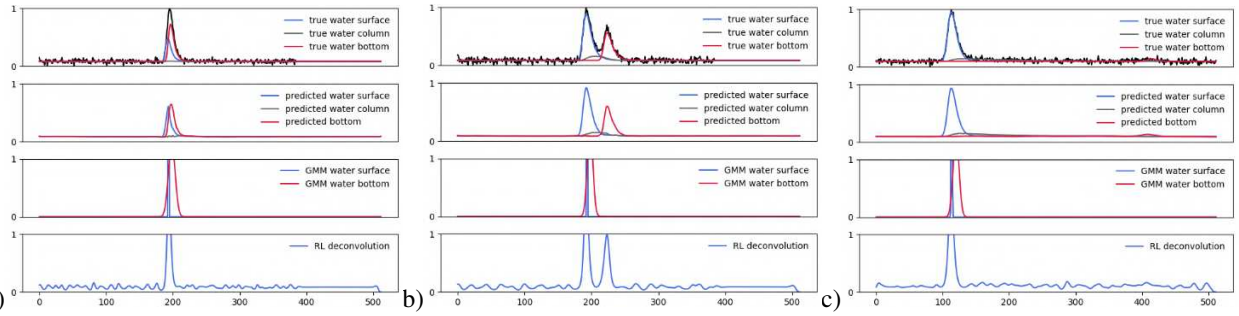


Figure 2. Examples of waveform decomposition results in a) very shallow, b) shallow and c) deep water.

Amplitude errors are lower than with these two traditional methods, and overlapping or weak echoes not captured with GMM and RL are detected.

Considering the correlation between the physical characteristics of the targets and their response functions, decomposing lidar waveforms into a sum of surface functions is of great potential to derive information on the nature or the structure of the objects detected. In the case of bathymetric lidar, it offers the potential to retrieve more information about the bathymetry. In addition, it may pave the way for water column turbidity assessments relying on remotely sensed rather than field measurements. Further developments to add physical constraints in the loss function and better estimate the maximum signal-to-noise ratio supported by the approach are being conducted. Application to real data should also give further information on the usability of attentive autoencoders to process large sets of bathymetric waveforms.

8. ACKNOWLEDGEMENTS

The authors are grateful to the Saur Group, the Région Bretagne and the Nantes-Rennes Topo-Bathymetric LiDAR platform for their financial support.

11. REFERENCES

[1] C. Mallet and F. Bretar, "Full-waveform topographic lidar: State-of-the-art," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 1–16, Aug. 2009, doi: 10.1016/j.isprsjprs.2008.09.007.

[2] C. Mallet, F. Bretar, M. Roux, U. Soergel, and C. Heipke, "Relevance assessment of full-waveform lidar data for urban area classification," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 6, Supplement, pp. S71–S84, Mar. 2011, doi: 10.1016/j.isprsjprs.2011.09.008.

[3] A. Chauve *et al.*, "Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest," *Int J Remote Sens*, vol. 30, no. 19, pp. 5211–5228, 2009, doi: 10.1080/01431160903023009.

[4] W. Philpot, *Airborne Laser Hydrography II*. 2019. [Online]. Available: https://ecommons.cornell.edu/handle/1813/66666

[5] D. Lague and B. Feldmann, "Topo-bathymetric airborne LiDAR for fluvial-geomorphology analysis," in *Developments in Earth Surface Processes*, P. Tarolli and S. M. Mudd, Eds., in Remote Sensing of Geomorphology, vol. 23. Elsevier, 2020, pp. 25–54. doi: 10.1016/B978-0-444-64177-9.00002-3.

[6] A. Chauve, C. Mallet, F. Bretar, S. Durrieu, M. Pierrot-Deseilligny, and W. Puech, "Processing Full-Waveform Lidar Data: Modelling Raw Signals," vol. 36 (Part 3/W52), pp. 102–107, Sep. 2007, Accessed: Jan. 03, 2023. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-00293129

[7] T. Zhou, S. C. Popescu, K. Krause, R. D. Sheridan, and E. Putman, "Gold – A novel deconvolution algorithm with optimization for waveform LiDAR processing," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 129, pp. 131–150, Jul. 2017, doi: 10.1016/j.isprsjprs.2017.04.021.

[8] J. Wu, J. A. N. Van Aardt, and G. P. Asner, "A comparison of signal deconvolution algorithms based on small-footprint LiDAR waveform simulation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6 PART 2, pp. 2402–2414, Jun. 2011, doi: 10.1109/TGRS.2010.2103080.

[9] G. Zhou *et al.*, "Comparison Analysis of Five Waveform Decomposition Algorithms for the Airborne LiDAR Echo Signal," *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 14, pp. 7869–7880, 2021, doi: 10.1109/JSTARS.2021.3096197.

[10] W. Wagner, A. Ullrich, V. Ducic, T. Melzer, and N. Studnicka, "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 60, no. 2, pp. 100–112, Apr. 2006, doi: 10.1016/J.ISPRSJPRS.2005.12.001.

[11] H. J. Zwally *et al.*, "ICESat's laser measurements of polar ice, atmosphere, ocean, and land," *J Geodyn*, vol. 34, no. 3–4, pp. 405–445, Oct. 2002, doi: 10.1016/S0264-3707(02)00042-X.

[12] M. A. Hofton, J. B. Minster, and J. B. Blair, "Decomposition of laser altimeter waveforms," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 4 II, pp. 1989–1996, Jul. 2000, doi: 10.1109/36.851780.

[13] B. Jutzi and U. Stilla, "Range determination with waveform recording laser systems using a Wiener Filter," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 61, no. 2, pp. 95–107, Nov. 2006, doi: 10.1016/J.ISPRSJPRS.2006.09.001.

[14] J. Niemeyer, J. D. Wegner, C. Mallet, F. Rottensteiner, and U. Soergel, "Conditional random fields for urban scene classification with full waveform LiDAR data," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 6952 LNCS, pp. 233–244, 2011, doi: 10.1007/978-3-642-24393-6_20/COVER.

[15] C. Wang, Q. Li, Y. Liu, G. Wu, P. Liu, and X. Ding, "A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 101, pp. 22–35, Mar. 2015, doi: 10.1016/J.ISPRSJPRS.2014.11.005.

[16] C. E. Parrish, I. Jeong, R. D. Nowak, and R. Brent Smith, "Empirical comparison of full-waveform lidar algorithms: Range extraction and discrimination performance," *Photogramm Eng Remote Sensing*, vol. 77, no. 8, pp. 825–838, 2011, doi: 10.14358/PERS.77.8.825.

[17] M. Perslev, M. Jensen, S. Darkner, P. J. Jennum, and C. Igel, "U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging," *Adv Neural Inf Process Syst*, vol. 32, 2019.

[18] M. Letard *et al.*, "Classification of coastal and estuarine ecosystems using full-waveform topo-bathymetric lidar data and artificial intelligence," *Oceans Conference Record (IEEE)*, vol. 2021-September, 2021, doi: 10.23919/OCEANS44145.2021.9705797.

[19] D. Bahdanau, K. Cho, and Y. Bengio, "NEURAL MACHINE TRANSLATION BY JOINTLY LEARNING TO ALIGN AND TRANSLATE".

[20] H. Abdallah, N. Baghdadi, J. S. Bailly, Y. Pastol, and F. Fabre, "Wa-LiD: A new LiDAR simulator for waters," *IEEE Geoscience and Remote Sensing Letters*, vol. 9, no. 4, pp. 744–748, 2012, doi: 10.1109/LGRS.2011.2180506.

Additional content to Chapter 3

Cloudcompare q3DMASC plugin implementation and operation

Using the q3DMASC plugin for classifier training or inference requires a labelled core point file and up to 3 accessory point clouds used to compute the features around each core point: PC1 (e.g., green channel), PC2 (e.g., NIR channel) and CTX (e.g., a point cloud with a populated classification field). For single point cloud classification, only one accessory point cloud is needed. A text file contains the description of point clouds, scales and features to be used for training. Upon training completion, a classifier file is saved and can be subsequently used with q3DMASC to apply the classifier to other point clouds.

Here are the main characteristics of the q3DMASC plugin implemented in the open source software CloudCompare (CC):

Accessibility: the q3DMASC plugin has been designed to be usable without programming language knowledge (e.g., Python) directly in the CC GUI. As such it makes a great introductory tool for non-specialists, for teaching and for quick tests without having to setup a complete programming environment. We have also modified the CC scissor tool to allow direct interactive labelling of 3D data, and introduced a tool to automatically split point clouds according to classes, and a new plugin for labelling data in 3D has just been released (QCloudLayers by Wiggins Tech). These simple tools associated with the neat 3D visualization of CC greatly facilitate the creation of labelled 3D data for training.

Speed: (CC) written in C++ has a well proven, fast and fully parallelized 3D neighbourhood search essential for fast computation of spherical neighbourhood or kNN search. While not critical during the training phase as a limited number of samples is necessary, this is essential during application and production phases to compute features on several millions of points without requiring a specific configuration (e.g. GPU).

Scalability: the q3DMASC plugin can be used in command line mode without GUI in order to apply the classifier in batch mode for large point cloud projects that would not fit in the computer memory. For instance, we have been able to use it routinely to process projects with more than 10 billions points using tiling strategies.

Non data source specific: while some features of 3DMASC are specific to Airborne lidar (e.g., multi-echo features), many geometric features can be used for any type of high resolution 3D point cloud created, for instance, from terrestrial lidar, Structure From Motion (SFM), Satellite Stereo Photogrammetry and multibeam sonar. There are in particular provision to use RGBNIR information that can be essential for SFM.

Flexibility in feature creation: to generate complex single or dual cloud features over several scales, the user has to create a text file containing the description of the various point clouds, the scales to be

used and the features to be computed. Complex single cloud features can be generated using the following formalism:

$$FEAT\_SC\#\_STAT\_PC\#$$

in which $FEAT$ corresponds to a predefined list of features (e.g., intensity, z, number of returns, sphericity, ….), $SC\#$ indicates the scale at which they will be calculated, $STAT$ is a statistical descriptor for point-based features sampled within the spherical neighbourhood (mean, mode, median, std, range, skew), $PC\#$ indicate the point cloud to be used for calculation around the core point.
Dual cloud features are generated with this formalism:

$$FEAT\_SC\#\_STAT\_PC\#\_PC\$\_MATH$$

In which $PC\$$ indicates the second cloud to be used and $MATH$ is an operator (minus, plus, divide, multiply). For instance, the Z mode difference (Figure 2) between the green channel (PC1) and the NIR channel (PC2) calculated at all possible scales is written $Z\_SCx\_MODE\_PC1\_PC2\_MINUS$.
Contextual features are constructed using the following formalism:

$$DZk\_SC0\_PC\#\_CTX\#$$

In which DZk (resp. DHk) indicates the vertical (resp. horizontal) distance to the k nearest neighbours, $PC\#$ indicates the PC considered and $CTX\#$ the number in the classification field to consider (e.g., 2 for ground, 5 for vegetation…). For instance, the average vertical distance to the 3 nearest ground points of the NIR channel (PC2) that holds a valid classification field is $DZ3\_SC0\_PC2\_CTX2$.

Explainability: We use a random forest algorithm that combines a good performance on many attributes, simplified feature selection, and robustness to overfitting. After training, the GUI version of 3DMASC outputs the overall accuracy, RF feature ranking and allows to manually remove features that are less contributing. After training completion, users can directly visualize feature values in 3D to understand why they contribute directly or not to classification success.

For training purposes, we chose the cross-platform OpenCV library (Bradski, 2000) implementation of Random Forests as it allows classifiers created in Cloudcompare to be used in Python and vice-versa. The downside of the C++ implementation of OpenCV is that the training is not parallelized, and is consequently much slower than the RF implementation, e.g., of scikit-learn (Pedregosa et al., 2011). RF training is thus the main bottleneck during classifier creation in the CC version. Classifier application is extremely fast, and feature calculation becomes the main bottleneck. Expert users can directly train their classifier in python with their favourite algorithm.

Point-based features and single/dual cloud features constructed from them in spherical neighbourhood:

Name	Single cloud feature stats.	Dual cloud features	
		Substraction	Division
Elevation	Std, Skew.	Mean,Median, Mode, Std, Skew.	-
Intensity	X	Std, Skew.	Mean, Median, Mode
Return Number	Mean	-	-
Numb. of returns	Mean	-	-
Echo Ratio	Mean	-	-
R, G, B	Mean, Mode, Median	-	-

Table 5.17: Point-based features available in 3DMASC.
*: not used as a point-based feature.

Dimensionality-based features computed in spherical neighbourhood:

Name	Formulation from eigenvalues	Dual cloud features
PCA1*	$\lambda 1/(\lambda 1 + \lambda 2 + \lambda 3)$	subtraction
PCA2*	$\lambda 2/(\lambda 1 + \lambda 2 + \lambda 3)$	subtraction
PCA3/Surf variation+	$\lambda 3/(\lambda 1 + \lambda 2 + \lambda 3)$	subtraction
Sphericity+	$\lambda 3/\lambda 1$	subtraction
Linearity+	$(\lambda 1 - \lambda 2)/\lambda 1$	subtraction
Planarity+	$(\lambda 2 - \lambda 3)/\lambda 1$	subtraction

Table 5.18: Dimensionality-based features available in 3DMASC.
*: Brodu et al., 2012; +: Weinmann et al., 2013

Complete list of features used in Chapter 3

Geometry-based features computed in spherical neighbourhood:

Name	Information	Dual cloud features
Verticality*	Varies between 0 (horizontal) and 1 (vertical)	subtraction
Detrended Roughness*	Std of distance between points and best fitting plane	subtraction
Curvature	Mean curvature in CC = average of principal curvatures	subtraction
Nb of points	-	subtraction
Anisotropy	Ratio of distance to center of mass and radius of sphere	subtraction
First Order Moment	Hackel et al., 2016	subtraction

Table 5.19: Geometry-based features available in 3DMASC.
*: Demantké et al., 2012

Height-based metrics computed in spherical neighbourhood:

Name	Formulation	Dual cloud features
Zrange	$Z_{max} - Z_{min}$	subtraction
Zmin	$Z - Z_{min}$	subtraction
Zmax	$Z_{max} - Z$	subtraction

Table 5.20: Height-based features available in 3DMASC.
Z is the core point's elevation, Z_{max} and Z_{min} are the maximum and minimum elevation in the spherical neighbourhood, respectively.

Contextual features in the NIR channel:

Name	Formulation	Target class
DZ to kNN	Mean vertical distance to k nearest neighbours	1064 nm ground
DH to kNN	Mean horizontal distance to k nearest neighbours	1064 nm ground

Table 5.21: Contextual features available in 3DMASC.

Supplementary materials to the chapter

Figure 5.19 presents the results of the experiments made to determine the scale to use for evaluation and the correlation threshold to apply during feature selection (OA=Overall Accuracy).

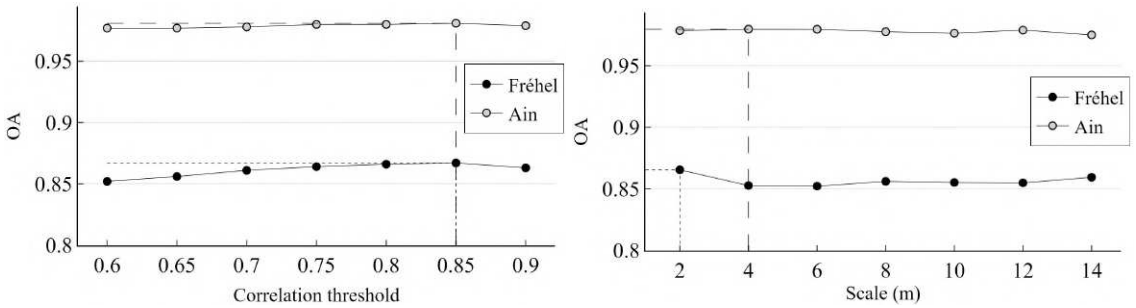


Figure 5.19: Results of the experiments performed to determine the correlation threshold to apply for feature selection, and the scale at which to evaluate each feature.

Figure 5.20 illustrates the variation of the OA depending on the number of predictors used. The predictors used at each iteration are identic to those involved in the realization of Figure 3.8 in the paper. This figure shows that OA and OOB display similar dynamics when pruning the predictors set.

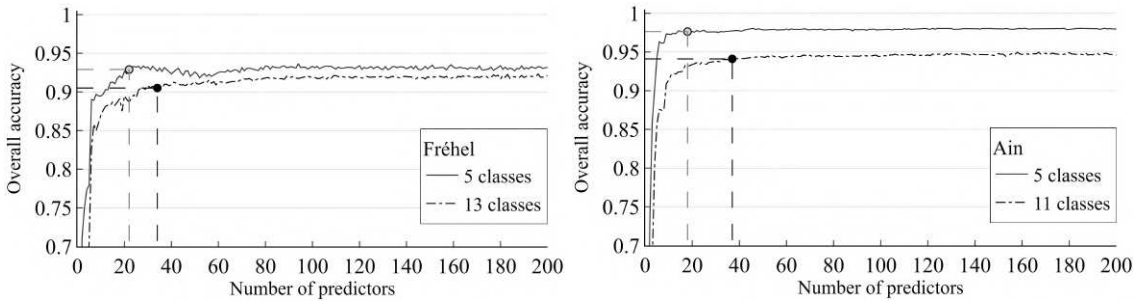


Figure 5.20: Overall accuracy depending on the number of predictors used for each experiment.

Figure 5.21 gives more detailed information on the correlation between features computed at different scales.

Figure 5.22 illustrates the impact of large scales on classification accuracy depending on the presence or absence of contextual features (vertical distances to a previously classified ground point cloud). Scales were removed iteratively per decreasing order.

Table 5.22 details the features used to compare 3DMASC to other approaches. The eigenvalues referred to are those obtained on the covariance matrix of spherical neighbourhoods. For detailed mathematical expressions of the different attributes, please consult the original papers (Chehata et al., 2009;

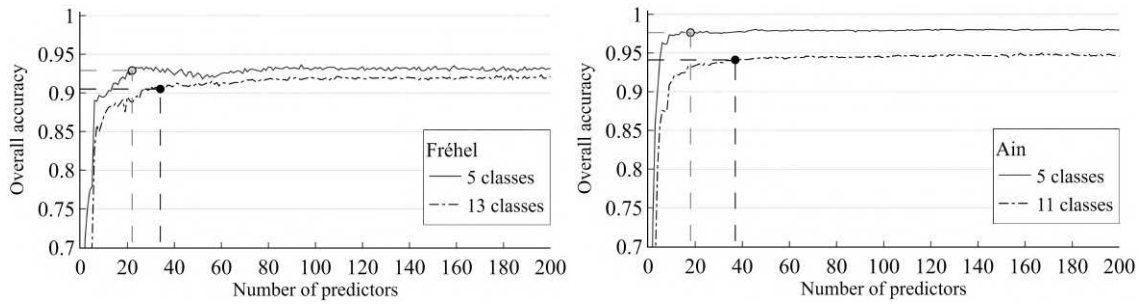


Figure 5.21: Linear correlation between features computed at scales separated by ds=1 m or ds=3 m for different families of features. SC = Single Cloud ; DC = Dual Cloud

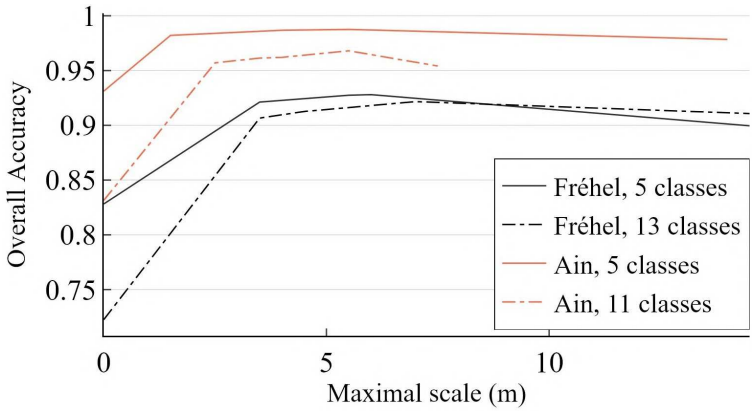


Figure 5.22: Classification performances depending on the maximal scale kept in the optimized predictor set, and in the predictor set augmented with contextual attributes.

Hackel et al., 2016; Thomas et al., 2018). The comparison to the fast point feature histogram introduced in Rusu et al., 2009 was made using the dedicated computation framework implemented in the Python Open3D library (Zhou et al., 2018).

Funding and data sources

This research was partially funded by the Saur Group and the Region Bretagne, who the authors thank for their support. The Titan DW sensor, operated by the Nantes-Rennes Lidar Platform has been funded by the Region Pays de la Loire with funding of the RS2E-OSUNA programs and the Region Bretagne with support from the European Regional Development Fund. Patrick Launeau is greatly acknowledged for his contribution in the acquisition of the Titan DW sensor. We thank Cyril Michon, Emmanuel Gouraud, William Gentile from Geofit-Expert company, and Laurence Hubert-Moy for their contribution in the overall operation of the Titan DW sensor. We thank Electricité De France (A. Barillier, A. Clutier) for commissioning the acquisition of the Ain River survey and providing access to the data.

Name	Thomas et al., 2018	Hackel et al., 2016	Chehata et al., 2009
Sum of eigenvalues	X	X	
Omnivariance	X	X	
Eigenentropy	X	X	
Anisotropy	X	X	
Linearity	X	X	X
Planarity	X	X	X
Sphericity	X	X	X
Curvature			X
Surface variation	X	X	
Verticality		X	
Verticality based on 1st eigenvector	X		
Verticality based on 3rd eigenvector	X		
Vertical moment (1st order)	X		
Vertical moment (2nd order)	X		
Number of points	X		
Statistical moments of eigenvectors (1st and 2nd order)	X	X	
Z range in neighbourhood		X	
Difference with minimal Z in neighbourhood		X	X
Difference with maximal Z in neighbourhood		X	
Standard deviation of Z in neighbourhood			X
Residuals of the fitting of a plane to the neighbourhood			X
Deviation angle of a fitted plan normal to the vertical			X
Variance of the deviation angles of the three dimensions of the neighbourhood			X
Distance to the fitted plan			X
Number of returns			X
Normalised return number			X

Table 5.22: Description of the features used in each approach compared to 3DMASC on the Ain dataset.

RÉSUMÉ EN FRANÇAIS

Cette thèse s'intéresse à la définition de méthodes permettant d'extraire de l'information sur l'état de l'environnement à l'aide du lidar topo-bathymétrique. Nous nous focalisons plus particulièrement sur le milieu côtier, en développant des méthodes d'apprentissage (classique et profond) sur les ondes complètes et les nuages de points.

L'importance des interfaces terre-eau

Les interfaces terre/eau, comprenant les rives des lacs, les berges des cours d'eau et les franges littorales, sont une partie cruciale de la Terre. Elles jouent des rôles clés dans les équilibres et cycles naturels maintenant les systèmes terrestres. En effet, les zones côtières, estuariennes et fluviales abritent de nombreux écosystèmes et fournissent aux communautés humaines en expansion constante des services essentiels tels que de la nourriture, des activités culturelles et une protection face aux dangers naturels liés à la montée du niveau de l'eau et aux événements météorologiques (Barbier et al., 2011). Par conséquent, elles constituent une base sur laquelle des sociétés entières peuvent s'appuyer et sont le lieu de nombreux processus socio-écologiques à différentes échelles temporelles. Des marées semi-quotidiennes attirant les promeneurs locaux ou les pêcheurs régionaux au développement de zones touristiques entières sur des décennies à travers des activités saisonnières, les zones de continuum terre-eau évoluent constamment (Syvitski et al., 2005). Les changements de niveau d'eau créent de nouveaux paysages à différentes échelles de temps (quotidiennes, saisonnières, annuelles), les processus sédimentaires modifient la morphologie du terrain, et des événements soudains tels que des tempêtes génèrent des changements rapides dans les écosystèmes (Syvitski et al., 2005).

De part et d'autre du rivage, les écosystèmes sont sous l'influence de l'hydrodynamique marine ou fluviale et des processus hydrologiques terrestres, faisant des franges côtières, des estuaires et des cours d'eau un point de rencontre pour les changements spatio-temporels. La surveillance de ces changements est cruciale pour protéger ces environnements contre les processus destructeurs naturels accélérés (Leatherman et al., 2000) et pour s'assurer que ces zones attractives peuvent toujours supporter la pression anthropique. En effet, en 2003, 41 % de la population mondiale vivait à moins de 100 mètres des côtes, qui

abritaient également 21 des 23 mégapoles mondiales. De plus, Martínez et al., 2007 ont montré que la grande diversité d'écosystèmes situés à l'interface des océans et des continents produisait 77 % de la valeur économique estimée des biens et services fournis par les écosystèmes dans le monde en 2007, selon la méthode de Costanza et al., 1997.

Les herbiers marins, les prés salés, les mangroves, les macroalgues, les dunes, les ripisylves et les plages sont des exemples d'habitats du continuum terre-eau. Ils interagissent continuellement avec le niveau de l'eau et fournissent des exemples clairs de la manière dont les écosystèmes soutiennent l'équilibre écologique de ces zones. Les herbiers marins garantissent la qualité de l'eau et sont d'importants puits de carbone, tout comme les prés salés et les mangroves (Barbier et al., 2011; Turner et al., 2015). La végétation côtière et riveraine offre également une protection contre les dangers hydrologiques pour les communautés locales et les infrastructures, et fournit de nombreuses activités récréatives telles que la plongée, la pêche, la baignade, le rafting et le char à voile (Barbier et al., 2011; Turner et al., 2015). Enfin, ces écosystèmes soutiennent une large gamme d'espèces endémiques en leur offrant des nurseries, de la nourriture et de l'oxygène (Barbier et al., 2011; Turner et al., 2015).

Toutes ces observations convergent pour montrer combien les écosystèmes des interfaces terre-eau sont importants d'un point de vue écologique, social et économique (Barbier et al., 2011; Costanza et al., 1997; Martínez et al., 2007). Cependant, les habitats fluviaux, marins et terrestres sont menacés par le changement climatique et la pression anthropique (Barbier et al., 2011), et de nombreuses études s'accordent à dire que leur évolution est difficile à anticiper et doit être surveillée pour garantir un soutien continu aux communautés littorales (Barbier et al., 2011; Costanza et al., 1997; Martínez et al., 2007).

La surveillance et la protection des interfaces terre-eau impliquent d'abord d'avoir accès à des outils fournissant des informations pertinentes pour les caractériser. De tels outils reposent nécessairement sur l'acquisition de données à une résolution temporelle et spatiale en phase avec les spécificités des lacs, des rivières et des côtes. Actuellement, l'observation ininterrompue des surfaces immergées et émergées le long de ces interfaces reste un défi méthodologique en raison de la présence d'eau (Gao, 2009; Kutser et al., 2020). La vaste diversité des types de surfaces présentes dans ces zones complique également leur exploration à une échelle temporelle pertinente avec leur évolution rapide, et sur des étendues larges et représentatives. La télédétection, détaillée ci-dessous, peut répondre efficacement à ce problème en fournissant un moyen de collecte de données sans nécessiter un accès direct aux zones fragiles ou éloignées et en permettant un relevé plus rapide de vastes portions de terrain.

La télédétection comme outil d'observation des interfaces terre-eau

La télédétection - ou *remote sensing* (RS) en anglais - désigne un ensemble de techniques utilisées pour étudier à distance les propriétés d'objets artificiels ou naturels en fonction de leurs interactions avec les ondes électromagnétiques (Rees, 2001). Typiquement, ces mesures sont réalisées à partir d'aéronefs, de satellites ou, plus récemment, de véhicules aériens inhabités (UAV).

En pratique, la télédétection consiste à surveiller à distance les radiations réfléchies ou émises par des objets, en utilisant une large gamme de capteurs dédiés (Rees, 2001). La télédétection est maintenant largement utilisée pour l'observation de la Terre (*Earth observation* - EO - en anglais), c'est-à-dire pour recueillir des informations sur les différents systèmes qui constituent la planète Terre. Les techniques d'imagerie, en particulier, constituent une grande partie des processus d'EO. Elles peuvent être séparées en deux catégories : l'imagerie passive, qui capture les radiations émises par les objets, et l'imagerie active, qui émet des radiations vers des objets d'intérêt et enregistre la manière dont elles les réfléchissent.

Les données d'imagerie par télédétection vont des photographies 2D aux modèles 3D de la topographie de la Terre grâce à la stéréoscopie et à la télémétrie laser (LiDAR, ou lidar). La stéréoscopie repose sur l'acquisition d'images avec un fort recouvrement pour déterminer les coordonnées 3D de leurs points communs, permettant ainsi d'obtenir des informations en 3D à partir de photographies 2D. D'autre part, le lidar est une technique d'imagerie active qui dérive une modélisation 3D de l'environnement en émettant des impulsions laser et en observant la manière dont elles sont réfléchies par les surfaces interceptées (Vosselman et al., 2010).

Traditionnellement, l'interface terre-eau est étudiée avec des méthodes d'imagerie passive ou des levés sonar (Gao, 2009; Kutser et al., 2020), ce qui soulève plusieurs enjeux:

- **La nécessité d'une approche intégrée :** comme expliqué précédemment, les écosystèmes et les hydrosystèmes peuplant l'interface terre-eau sont liés à la fois aux processus hydrologiques et terrestres. Étudier séparément les zones immergées et émergées le long de ces franges élimine ainsi une large part de contenu informatif. Cependant, les zones immergées et émergées ont traditionnellement été étudiées séparément. Typiquement, les zones immergées sont étudiées avec des techniques acoustiques - des systèmes sonar - (Barrell et al., 2015; Komatsu et al., 2003; Pasqualini et al., 1998). D'un autre côté, les zones terrestres bénéficient d'une large gamme de possibilités d'imagerie par télédétection, des photographies aériennes 2D aux modèles 3D de la topographie obtenus par imagerie active. Fusionner des campagnes distinctes sous-marines et terrestres - par exemple, coupler sonar et imagerie passive - pourrait être une solution. Cependant, les relevés séparés ont souvent une zone de recouvrement mince, qui peut être difficile à échantillonner avec des points de contrôle au sol pour trouver une transformation de coordonnées. En effet, les méthodes d'hydrographie ne

sont pas utilisables partout : les véhicules de surface inhabités ne peuvent pas être déployés trop loin de leur centre opérationnel, et les bateaux ne peuvent pas accéder aux zones dangereuses. L'échantillonage par sonar des eaux extrêmement peu profondes est donc impossible sans risquer d'échouer pour les deux outils. D'autre part, l'imagerie passive ne pénètre souvent pas suffisamment dans l'eau pour combler l'absence de données dans ces zones (Kutser et al., 2020). Le fait que les relevés côtiers soient encore séparés en campagnes bathymétriques et topographiques reposant sur de systèmes de coordonnées distincts témoigne par ailleurs d'une volonté encore récente de fusionner les deux domaines et de les étudier comme un système commun et complexe.

- **La difficulté de voir sous la surface d'eau :** les quelques approches offrant une observation continue du continuum terre-eau reposent généralement sur des images satellites ou aériennes qui couvrent les deux côtés de la surface d'eau (McKenzie et al., 2020; Mumby et al., 1997; Topouzelis et al., 2018). Ces données sont soit étudiées directement, soit utilisées pour estimer la bathymétrie et la topographie à l'aide de modèles d'inversion des transferts radiatifs. Les images satellites et aériennes ont l'avantage de couvrir de grandes zones avec des coûts de déploiement bien inférieurs aux relevés aquatiques ou aux expéditions sur le terrain à pied - en particulier dans le cas des capteurs spatiaux. Cependant, le principal problème avec l'imagerie passive est la gamme de profondeur dans laquelle elle est utilisable (Kutser et al., 2020). En raison de phénomènes optiques, au-delà d'un certain seuil de profondeur qui varie avec la clarté de l'eau, l'imagerie passive ne peut plus fournir d'informations sur ce qui se trouve sous la surface de l'eau en raison de l'atténuation de la lumière par l'eau. Dans les rivières ou lacs turbides, cette profondeur maximale peut être très faible, et peu voire aucune information ne peut être obtenue avec l'imagerie passive. Sauf dans des eaux très claires, cette option ne couvre donc pas vraiment l'interface terre-eau.

- **L'importance de l'accès à des informations sur la structure verticale du milieu :** l'accès à des approches fournissant des informations sur la structure tridimensionnelle des écosystèmes est essentielle pour permettre de études écologiques plus approndies comme par exemple l'évaluation des puits de carbone, de la production de biomasse ou encore des capacités d'atténuation des flux inhérents aux écosystèmes des interfaces terre-eau (Enríquez et al., 2019; Dubayah et al., 2020; Lu et al., 2023; Mury et al., 2020). Dans les eaux peu profondes, les études basées sur des capteurs passifs donnent accès à l'élévation de couverts sous-marins mais pas à l'élévation du fond marin lui-même lorsqu'il est recouvert par de la végétation (Stumpf et al., 2003; Lesser et al., 2007), ce qui fournirait des informations sur la structure verticale de celle-ci et permettrait des estimations de biomasse ou d'autres évaluations structurelles (Wedding et al., 2008; Lindberg et al., 2012).

- Enfin, des méthodes d'observation rapidement déployables en cas de catastrophes naturelles sont

nécessaires pour permettre une évaluation rapide des potentiels dommages.

Le lidar topo-bathymétrique est un outil d'imagerie active qui répond efficacement à ces besoins. Il repose sur la télémétrie laser, et exploite simultanément deux capteurs (Lague et al., 2020). L'un d'eux utilise la lumière proche-infrarouge (*near infrared* - NIR - en anglais), tandis que l'autre opère dans le spectre vert, pénétrant ainsi la surfaces d'eau (Guenther et al., 2000). Le lidar topo-bathymétrique est aéroporté, il fournit des données 3D à une densité verticale et horizontale élevée dans les eaux peu profondes avec une turbidité raisonnable et peut être rapidement utilisé sur de vastes étendues (Lague et al., 2020). Cependant, il reste largement sous-exploité pour l'extraction de connaissances environnementales sur les interfaces terre-eau.

Le lidar topo-bathymétrique a été introduit il y a quelques décennies, peu de temps après l'avènement du lidar bathymétrique, conçu spécifiquement pour les eaux côtières (Lague et al., 2020; Fernandez-Diaz et al., 2014). Ses objectifs sont de fournir des mesures mieux adaptées aux étendues d'eau douces, qui sont souvent plus étroites. Ce capteur produit un nuage de points 3D par longueur d'onde, générant ainsi deux échantillonnages distincts d'une même scène, notamment dans les zones végétalisées ou immergées. Il peut également enregistrer l'énergie complète reçue au fil du temps pour chaque tir laser : les formes d'onde complètes.

Cependant, les données lidar topo-bathymétriques sont très spécifiques et difficiles à traiter.
La présence d'eau influence fortement les nuages de points et les formes d'onde résultants (Guenther, 1985) et nécessite le développement de nouveaux outils. De plus, dans des contextes topo-bathymétriques, les longueurs d'onde NIR et verte sont toutes deux informatives. Parvenir à traiter de manière précise à la fois les zones émergées et immergées nécessite donc souvent l'exploitation simultanée de ces deux ensembles de données, ce qui constitue un défi de traitement conséquent.

Indépendamment de la présence d'eau, **les nuages de points 3D** posent de nombreux défis de traitement. Ils ont une structure irrégulière, présentent de nombreuses zones occultées et ont des densités irrégulières (Guo et al., 2021b). Ces caractéristiques rendent difficile l'identification automatique des objets qu'ils contiennent. De plus, en raison de leurs tridimensionalité, ils sont plus difficiles à traiter. Le traitement des nuages de points 3D a largement bénéficié des progrès globaux dans le traitement des données au cours de la dernière décennie (Guo et al., 2021b). Cependant, ces progrès ont principalement concerné le traitement des données topographiques (Morsy et al., 2022), et n'ont pas été pleinement transférés aux nuages de points bathymétriques. La première explication est l'avènement plus récent du lidar vert et son accessibilité réduite en raison de la puissance laser plus important qu'il requiert. Une autre explication majeure est la séparation persistante entre les sciences marines et terrestres. La plupart des approches développées pour le traitement des nuages de points se concentrent sur des zones terrestres telles que les villes ou les forêts (Mao et al., 2022b; Huang et al., 2021; Axelsson et al., 2023; Liu et al., 2021a), qui sont

plus accessibles à l'étude par télédétection. Cependant, les objets marins et terrestres présentent des caractéristiques distinctes en termes de forme, d'échelle et de matériau. Le transfert de méthodes développées pour le lidar topographique au lidar topo-bathymétrique n'est donc pas trivial. Par conséquent, très peu de méthodes publiées produisent des nuages de points 3D classifiés à partir du lidar topo-bathymétrique, et encore moins pour les surfaces topographiques et benthiques simultanément. Des outils de classification 3D des nuages de points bathymétriques, incluant la possibilité d'identifier avec précision une grande variété de classes à la fois immergées et émergées, sont donc toujours attendus.

Les formes d'onde sont également difficiles à traiter. Ce sont des séries complexes de puissance reçue et enregistrée par le capteur (Mallet et al., 2009) et, tout comme les nuages de points ont de nombreuses zones vides, elles contiennent souvent des informations sur moins d'un tiers de leur longueur, le reste étant du bruit. Leurs valeurs peuvent varier de plusieurs ordres de grandeur sur de courtes plages de temps en raison de la grande variété de conditions optiques rencontrées sur une zone topo-bathymétrique (Guenther et al., 2000). Enfin, les mesures d'intensité lidar dépendent fortement des conditions d'acquisition et nécessitent des connaissances expertes pour être traitées avec un biais limité (Kashani et al., 2015). La plupart du temps, les formes d'onde lidar ne sont donc exploitées que pour détecter les retours des surfaces immergées ou densément végétalisées, où les signaux deviennent trop faibles pour être détectés en vol. Elles sont ensuite rarement incluses par les constructeurs dans les jeux de données livrés. Les méthodes utilisées pour produire des nuages de points à partir de formes d'onde sont par ailleurs souvent couvertes par le secret industriel et sont principalement axées sur la précision de la bathymétrie plutôt que sur la maximisation de l'extraction de connaissances applicatives.

Ces facteurs rendent l'exploitation des formes d'onde lidar encore rare, d'autant plus pour la surveillance des écosystèmes topo-bathymétriques. Les approches existantes tentant de les exploiter les réduisent souvent à des rasters d'attributs 2D (Collin et al., 2012) ou ne ciblent qu'un côté de la surface d'eau (Hansen et al., 2021). Elles restent également très spécifiques aux zones ou aux conditions d'acquisition pour lesquelles elles ont été développées. Par exemple, le traitement des formes d'onde provenant de zones côtières peu profondes avec une faible turbidité est une tâche complètement différente de l'extraction d'informations à partir de levés en rivières turbides. Cependant, alors que les nuages de points sont une excellente source d'information géométrique (Hackel et al., 2016), les formes d'onde fournissent les connaissances spectrales nécessaires pour caractériser davantage les zones étudiées (Mallet et al., 2009). Elles pourraient donc largement bénéficier à l'observation des interfaces terre-eau si davantage de solutions de traitement étaient disponibles.

En fin de compte, des méthodes capables de gérer les spécificités des zones immergées et submergées séparément existent, mais peu parviennent à traiter les deux simultanément. Pour le traitement des formes d'onde bathymétriques, Xing et al., 2019 suggèrent même que trouver une méthode applicable à toutes

les conditions de profondeur et de turbidité reste hors de portée des possibilités actuelles de traitement. De manière générale, à l'heure actuelle, il est toujours difficile de trouver des approches de traitement des données lidar topo-bathymétrique généralisables à l'ensemble du continuum terre-eau.

L'hypothèse principale de cette thèse est que la télédétection lidar topo-bathymétrique pourrait considérablement améliorer notre compréhension des interfaces terre-eau. Cependant, des méthodes permettant d'exploiter pleinement les données qu'elle fournit doivent encore être développées. Ce travail vise donc à améliorer l'extraction de connaissances à partir de données existantes afin d'améliorer la modélisation des interfaces terre-eau. À cette fin, nous explorons comment l'apprentissage automatique peut s'appliquer aux traitement des données lidar topo-bathymétriques.

Les méthodes actuelles de traitement de données de télédétection

Globalement, le traitement des données vise à extraire des informations à partir de données brutes en les transformant à l'aide de différentes opérations. En pratique, cela consiste principalement à trouver des fonctions mathématiques permettant de prédire la une variable en fonction d'observations d'une autre. L'apprentissage automatique rassemble des méthodes utilisant des données empiriques pour résoudre ces problèmes mathématiques en optimisant les paramètres de fonctions modèle (Géron, 2022). Ce type de méthodes a généré des progrès consequents dans le traitement des nuages de points lidar, notamment dans l'identification des objets présents dans les données (Guo et al., 2021b). L'un des objectifs de ce travail est d'explorer comment les avancées existantes peuvent être étendues aux jeux de données bi-spectraux modélisant les zones terre-eau sans interruption.

Pour les formes d'onde, nous souhaitons évaluer l'applicabilité d'autres méthodes de traitement qui se sont révélées plus adaptées aux problèmes complexes : les réseaux neuronaux profonds.

Le potentiel des réseaux de neurones profonds

En parallèle du développement des lidars topo-bathymétriques, des méthodes de traitement capables de résoudre des problèmes plus complexes ont émergé et ont engendré des avancées significatives dans le traitement des données de télédétection. En particulier, l'apprentissage automatique et les réseaux de neurones profonds ont fait des progrès considérables grâce à l'augmentation de la puissance de calcul qui a caractérisé les dernières décennies (Goodfellow et al., 2016).

Les réseaux de neurones profonds sont constitués d'interconnexions de neurones organisés en couches (Goodfellow et al., 2016; Lecun et al., 2015). Chaque neurone effectue une combinaison linéaire de ses entrées associée à une fonction d'activation non linéaire. La connexion d'un nombre potentiellement élevé de neurones, organisés différemment en fonction de l'application (c'est ce qu'on appelle l'architecture), permet de modéliser des fonctions très complexes (Cybenko, 1989). La phase d'entraînement, réalisée par rétropropagation de l'erreur (Lecun et al., 1989; Rumelhart et al., 1986), consiste à estimer les paramètres de chaque neurone - les poids de l'opération linéaire.

Les réseaux de neurones peuvent adapter et généraliser leur apprentissage à de nouvelles entrées, ce qui en fait des outils puissants pour la recherche scientifique et l'analyse de données. Grâce à ce processus, les motifs caractéristiques des données se dégagent progressivement et sont utilisés pour construire des règles de prédiction adaptées à la tâche (Rumelhart et al., 1986). Alors que les méthodes de traitement traditionnelles reposant sur l'apprentissage automatique impliquent de définir en amont des transformation à appliquer aux données, les réseaux de neurones apprennent les opérations pertinentes pour les données et le problème directement sur les entrées (Goodfellow et al., 2016). Ils éliminent par conséquent le besoin de définir en amont des descripteurs, contrairement à l'apprentissage automatique classique qui repose sur une pré-transformation des observations.

Les réseaux de neurones profonds ont apporté de nouvelles perspectives dans les domaines du traitement d'images (Zhao et al., 2019; Khan et al., 2021), de la prédiction de l'évolution de séries temporelles (Lim et al., 2020; Wen et al., 2022), ou encore du traitement du langage (Bahdanau et al., 2014; Vaswani et al., 2017). Cependant, ils n'ont presque jamais été explorés pour améliorer le traitement des formes d'onde bathymétriques. Il est donc encore compliqué d'évaluer s'ils pourraient changer notre capacité à extraire des informations à partir des interactions du laser vert avec l'environnement via l'analyse des formes d'onde du lidar topo-bathymétrique.

Objectifs de cette thèse

L'objectif de cette thèse est de fournir de nouvelles perspectives sur l'extraction de connaissances à partir des levés lidar topo-bathymétriques pour l'observation des rivages. Nous souhaitons évaluer dans quelle mesure les nouvelles méthodes de traitement de données peuvent contribuer à cette tâche. Notre objectif principal est donc de combiner un contexte instrumental favorable - avec la popularité et la disponibilité croissantes des données lidar - et des circonstances méthodologiques positives - avec une puissance de calcul et une variété de solutions de traitement sans précédent.

Par conséquent, cette thèse aborde les question suivantes:

- **Comment répondre au manque d'outils accessibles pour l'extraction d'informations sé-**

mantiques à partir de nuages de points lidar bi-spectraux dans des zones naturelles complexes ? Peut-on remédier à l'indisponibilité de méthodes de classification de nuages de points adaptées aux levés topo-bathymétriques ? Est-il possible de combler cette lacune avec une méthode pouvant être utilisée sans expertise en informatique ?

- **Lorsque les formes d'onde sont disponibles, peut-on les exploiter pour distinguer un plus grand nombre de surfaces terrestres et marines ?** Combien de classes pouvons-nous distinguer en utilisant uniquement des formes d'onde ? Le manque d'information spatiale a-t-il un impact négatif sur les résultats ? Les formes d'onde suffisent-elles pour séparer des surfaces avec des géométries similaires ? Est-il possible de préserver la généralisabilité sans sacrifier la précision ? Peut-on gérer les biais liés à l'acquisition sans stratégies de prétraitement dépendantes du site ?

- **Dans quelle mesure pouvons-nous déployer le traitement des formes d'onde pour l'extraction de connaissances environnementales dans des environnements (topo)-bathymétriques ?** Peuvent-elles être utilisées pour estimer des paramètres environnementaux même en l'absence de mesures in-situ synchrones ? Est-il possible d'utiliser une approche similaire pour différents types d'eaux (côtières/intérieures, troubles/claires, très peu profondes/profondes) ?

- **Comment faire face à la question de la disponibilité de données labélisées compte tenu de la complexité de l'interprétation des formes d'onde ?** Il est impossible de labéliser manuellement des formes d'onde sans introduire de biais significatifs pour certaines tâches (principalement la régression et la détection d'objets). Pourtant, nous utilisons des approches supervisées pour évaluer le potentiel des réseaux neuronaux profonds. Dans ce contexte, comment pouvons-nous proposer des méthodes de traitement robustes si aucune campagne de mesure in-situ n'a été organisée pendant l'acquisition lidar ?

Contributions

Pour aborder chacune de ces questions, la thèse s'organise autour de cinq chapitres.

Le chapitre 1 offre une **présentation détaillée de la télédétection lidar**. Il introduit les capteurs lidar topographiques, bathymétriques et topo-bathymétriques et rappelle les principes physiques qui sous-tendent leur fonctionnement.

Le chapitre 2 présente les **stratégies existantes pour le traitement des données lidar**. Il rappelle d'abord le fonctionnement du traitement de données basé sur l'apprentissage et ses enjeux actuels.

Ensuite, il détaille les processus existants pour la classification des nuages de points et comment ils répondent à nos besoins. Enfin, il explore également les méthodes de traitement des formes d'onde et les défis restants dans le contexte de l'observation de l'interface terre-eau. En particulier, ce chapitre souligne le manque de solutions pour classifier les nuages de points bi-spectraux et l'absence d'approches exploitant les formes d'ondes bi-spectrales.

Après avoir introduit de manière approfondie les défis et les spécificités de l'extraction de connaissances à partir des données lidar topo-bathymétriques, nous proposons plusieurs contributions méthodologiques à ce sujet.

Dans **le chapitre 3**, nous présentons **3DMASC, une nouvelle méthode de classification de nuages de points bi-spectraux, et son application aux environnements topo-bathymétriques**. 3DMASC est le résultat de notre travail pour répondre au besoin d'outils accessibles pour le traitement de nuages de points environnementaux complexes, notamment en présence de deux nuages de points. Avec 3DMASC, nous introduisons également de nouveaux descripteurs des nuages de points qui pourraient être utilisées à d'autres fins que la classification dans le contexte de l'extraction de connaissances environnementales. Grâce à un processus de sélection de descripteurs et d'échelles automatique, 3DMASC produit des classifieurs légers aux résultats précis dans les environnements topo-bathymétriques tels que les rivières ou les côtes. L'approche est **entièrement explicable** grâce à l'évaluation de l'importance des descripteurs et à l'analyse SHAP. Elle est également **accessible aux spécialistes thématiques via un plugin open-source** dans le logiciel CloudCompare (Girardeau-Montaut, 2022). Les classifieurs s'appuient sur une quantité limitée de données labélisées et peuvent être entraînés sur des ordinateurs classiques via une interface graphique ou via le terminal.

Dans **le chapitre 4**, nous évaluons de **nouvelles possibilités pour l'extraction d'information sémantique à partir des formes d'onde lidar**. Nous abordons le besoin de classifications 3D des milieux terrestres et sous-marins en développant des approches basée sur les formes d'ondes bi-spectrales. Nous explorons leur capacité à caractériser des surfaces ayant une géométrie similaire, dans un premier temps en utilisant des approches classiques d'apprentissage automatique s'appuyant sur des descripteurs sur mesure. En extrayant des **attributs des formes d'ondes vertes et proches infrarouge**, nous obtenons une classification 3D d'une zone côtière à haute résolution spatiale. Nos résultats mettent en évidence la contribution de l'information bi-spectrale et la capacité à détecter un large éventail de classes avec un seul levé lidar. Cependant, comme cette méthode est basée sur de l'apprentissage automatique classique, elle nécessite une étape de prétraitement laborieuse, qui n'est pas optimale dans des eaux très peu profondes.

Dans un second temps, pour limiter les prétraitements, qui sont souvent spécifiques au jeu de données, nous proposons donc d'utiliser des **réseaux de neurones profond**s. Pour permettre l'identification directe de tous les objets présents dans les formes d'ondes lidar, nous proposons d'effectuer une **segmentation sémantique** (c'est-à-dire 1 label par pas de temps) au lieu d'une classification (1 label par forme d'onde). Malgré quelques limitations liées à la méthode de labélisation, nos résultats dépassent ceux obtenus avec l'apprentissage automatique classique. Les performances du réseau suggèrent que les attributs appris contiennent une information plus précise que ceux pré-définis : sur les seules formes d'ondes vertes, les performances de notre réseau sont meilleures que celles obtenues avec l'apprentissage automatique classique sur les formes d'ondes bi-spectrales.

Enfin, dans **le chapitre 5**, nous explorons l'extraction de connaissances plus fines à partir des formes d'onde lidar bathymétriques. Nous proposons notamment une nouvelle méthode pour **améliorer la détection des retours lidar dans les eaux très peu profondes et dans les eaux profondes ou troubles**. Avec l'approche de segmentation sémantique proposée dans le chapitre 4, l'identification d'échos superposés est impossible. Nous allons donc plus loin dans notre exploitation des réseaux de neurones profonds pour le traitement des formes d'ondes et proposons d'utiliser une **approche de détection d'objets pour identifier les échos dans les formes d'onde bathymétriques**. Pour lever les limitations liées à la labélisation des données, nous entraînons le réseau sur un jeu de données simulé représentant une large gamme de conditions d'acquisition. Cela nous permet d'avoir des retours labélisés dans des eaux extrêmement peu profondes, où les composants se chevauchent et sont parfois impossibles à repérer à l'œil nu. Cela nous fournit également des exemples de retours très faibles, un autre défi du traitement des formes d'ondes difficile à aborder avec une labélisation manuelle sans introduire de biais. Cette méthode produit des résultats prometteurs : elle détecte 95 % des composants benthiques dans des eaux extrêmement peu profondes et 70 % des composants caractérisés par un rapport signal/bruit supérieur à 15 dB. Bien que certains retours soient manqués dans les eaux très profondes, l'application au jeu de données test comporte de nombreuses détections correctes d'échos très faibles. Enfin, le transfert de la méthode à la détection d'échos en zones côtières peu profondes est réussi.

Nous étudions également la possibilité d'**estimer les paramètres optiques de l'eau, sans prétraitement et sans besoin de mesures in situ, en utilisant un réseau de neurones convolutif**. Ce réseau est conçu pour prédire le coefficient d'atténuation diffuse, le coefficient de rétrodiffusion à 180°, la perte d'énergie à l'interface air/eau et la réflectance du fond à partir de formes d'ondes bathymétriques. Il estime également la position de la surface de l'eau et du fond, ainsi que la profondeur. Bien que les résultats soient encourageants, ils présentent également des erreurs suggérant le besoin d'un réseau plus complexe. En effet, certaines interactions entre paramètres, par exemple, le lien entre perte d'énergie à l'interface terre/eau

et proéminence du retour du fond indépendamment de sa réflectance, ne sont pas capturées. Cependant, les positions de la surface de l'eau et du fond sont particulièrement bien prédites, sauf à des profondeurs plus importantes. Plusieurs modifications d'architecture et d'optimisation pourraient potentiellement améliorer ces estimations préliminaires.

En entraînant ces réseaux sur des **jeux de données synthétiques** et en explorant leur application à des données réelles par le biais de l'**adaptation de domaine**, nous abordons également le besoin de données labélisées, qui constitue souvent un inconvénient majeur pour l'exploitation des données. **En utilisant des modèles physiques de transfert radiatif, nous éliminons le besoin de labéliser manuellement les composantes des formes d'ondes et d'avoir accès à des mesures in-situ synchrones**. L'application des modèles obtenus aux formes d'ondes réelles nécessite une adaptation de domaine, mais des expérimentations préliminaires rendent déjà cela réalisable pour la détection de la surface de l'eau et du fond marin dans des zones côtières peu profondes.

Enfin, nous contribuons également aux améliorations méthodologiques dans le domaine de la télédétection lidar bi-spectrale en mettant à la disposition de la communauté **des outils open-source et des jeux de données labelisés**[5] utilisables pour la classification de nuages de points.

Dans l'ensemble, avec ces contributions méthodologiques, nous mettons en avant **le potentiel des réseaux de neurones profonds pour améliorer l'extraction d'information à partir des formes d'ondes lidar**. Nous constatons que leur capacité à modéliser des problèmes complexes sans nécessiter de travail préalable sur les descripteurs les rend particulièrement adaptés aux défis du traitement des formes d'ondes. Il est également intéressant de noter que les trois réseaux supervisés proposés pour le traitement des formes d'ondes vertes semblent réduire de manière significative un obstacle majeur des levés lidar bathymétriques : l'incertitude de la position de la surface de l'eau estimée. En effet, les trois approches fournissent des détections de surface d'une précision très élevée, alors qu'il est généralement admis que la détermination de la position de l'interface air/eau à partir des composantes bathymétriques des formes d'ondes constitue un défi (Guenther, 1985; Guenther et al., 2000; Lague et al., 2020; Mandlburger et al., 2013). Dans le contexte actuel caractérisé par la popularité croissante de la télédétection lidar bathymétrique, cela pourrait être une piste intéressante pour le développement de capteurs légers ne reposant pas sur des systèmes à double longueur d'onde.

De plus, la possibilité d'identifier un grand nombre de surfaces topo-bathymétriques avec un seul nuage de points vert grâce à 3DMASC est prometteuse, d'autant plus au vu de la disponibilité croissante

5. Plugin 3DMASC dans CloudCompare disponible sur `https://www.danielgm.net/cc/release/`, scripts Python accessibles sur `https://github.com/p-leroy/lidar_platform`, et jeux de données bientôt disponibles sur `https://opentopography.org/`

de levés lidar. En effet, les levés lidar bathymétriques du littoral français sont désormais tous disponibles en libre accès via le Shom[6]. Les acquisitions lidar topographiques en cours sur l'ensemble du territoire métropolitain français illustrent également la popularité croissante de la télédétection lidar. **Cette thèse intervient à un moment où les données lidar deviennent plus facilement accessibles pour les spécialistes thématiques**, via des plateformes en libre accès telles qu'OpenTopography (Krishnan et al., 2011) ou des agences gouvernementales[7]. Nos contributions méthodologiques résonnent donc avec les tendances actuelles dans les domaines académique et industriel. À une époque où le monde académique doit repenser ses pratiques en matière d'empreinte carbone, le partage de données coûteuses sur le plan environnemental et la fourniture d'outils pour les exploiter pleinement semblent essentiels. En fournissant des outils de traitement de données 3D aux spécialistes de l'environnement, nous souhaitons également rendre les acquisitions de données 3D plus judicieuses et conformes aux enjeux mondiaux actuels, qui nous obligent à repenser l'utilité de l'exploitation des capteurs lidar aéroportés pour dériver des rasters 2D. Enfin, la plupart des méthodes que nous avons introduites ont une portée d'application plus large que le simple traitement des formes d'ondes lidar et peuvent ainsi fournir des perspectives utiles pour d'autres études environnementales reposant sur des données 1D.

6. https://diffusion.shom.fr/
7. Par exemple, en France, https://geoservices.ign.fr/

Bibliography

Abady, Lydia et al. (Apr. 2014), "Assessment of quadrilateral fitting of the water column contribution in lidar waveforms on bathymetry estimates", *in*: *IEEE Geoscience and Remote Sensing Letters* 11.4, pp. 813–817, ISSN: 1545598X, DOI: 10.1109/LGRS.2013.2279271.

Abdallah, Hani et al. (2012), "Wa-LiD: A new LiDAR simulator for waters", *in*: *IEEE Geoscience and Remote Sensing Letters* 9.4, pp. 744–748, ISSN: 1545598X, DOI: 10.1109/LGRS.2011.2180506.

Abdallah, Hani et al. (2013), "Potential of Space-Borne LiDAR Sensors for Global Bathymetry in Coastal and Inland Waters", *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 6.1, pp. 202–216, DOI: 10.1109/JSTARS.2012.2209864.

Adams, Thomas et al. (Mar. 2012), "Extracting more data from LiDAR in forested areas by Analyzing waveform shape", *in*: *Remote Sensing* 4.3, pp. 682–702, ISSN: 20724292, DOI: 10.3390/rs4030682.

Adler-Golden, S M et al. (2005), "Remote bathymetry of the littoral zone from AVIRIS, LASH, and QuickBird imagery", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 43.2, pp. 337–347, ISSN: 1558-0644, DOI: 10.1109/TGRS.2004.841246, URL: https://ieeexplore.ieee.org/document/1386505.

Anders, K. et al. (May 2019), "High-frequency 3D geomorphic observation using hourly terrestrial laser scanning data of a sandy beach", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, 2/W5, Copernicus GmbH, pp. 317–324, DOI: 10.5194/isprs-annals-IV-2-W5-317-2019.

Anders, Katharina et al. (Jan. 2020), "4D objects-by-change: Spatiotemporal segmentation of geomorphic surface change from LiDAR time series", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 159, pp. 352–363, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2019.11.025.

Antonarakis, A. S. et al. (June 2008), "Object-based land cover classification using airborne LiDAR", *in*: *Remote Sensing of Environment* 112.6, pp. 2988–2998, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2008.02.004.

Arbeiter, Georg et al. (2012), "Evaluation of 3D Feature Descriptors for Classification of Surface Geometries in Point Clouds", *in*: *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1644–1650, ISBN: 9781467317368.

Aßmann, Andreas et al. (Jan. 2021), "Deep learning for LiDAR waveforms with multiple returns", *in*: *European Signal Processing Conference* 2021-January, pp. 1571–1575, ISSN: 22195491, DOI: 10.23919/EUSIPCO47968.2020.9287545.

Atzmon, Matan et al. (Mar. 2018), "Point Convolutional Neural Networks by Extension Operators", *in*: *ACM Transactions on Graphics* 37.4, pp. 1–12, ISSN: 15577368, DOI: 10.48550/arxiv.1803.10091, URL: https://arxiv.org/abs/1803.10091v1.

Axelsson, Christoffer R. et al. (Apr. 2023), "The use of dual-wavelength airborne laser scanning for estimating tree species composition and species-specific stem volumes in a boreal forest", *in*: *International Journal of Applied Earth Observation and Geoinformation* 118, p. 103251, ISSN: 1569-8432, DOI: 10.1016/J.JAG.2023.103251.

Babushka, Andriy et al. (Oct. 2014), "The Calculation of Received Optical Power During Airborne Laser Scanning of Water Object", *in*: *Geomatics, Landmanagement and Landscape* 3, pp. 7–15, ISSN: 23001496, DOI: 10.15576/GLL/2014.3.7, URL: http://gll.ur.krakow.pl/zasoby/74/1-3-2014.pdf.

Bahdanau, Dzmitry et al. (Sept. 2014), "Neural Machine Translation by Jointly Learning to Align and Translate", *in*: URL: http://arxiv.org/abs/1409.0473.

Baltsavias, E P (1999), "Airborne laser scanning: basic relations and formulas", *in*: *ISPRS Journal of Photogrammetry & Remote Sensing* 54, pp. 199–214.

Barbier, Edward B et al. (Nov. 2011), "The value of estuarine and coastal ecosystem services", en, *in*: *Ecological Monographs* 81.2, pp. 169–193, ISSN: 1557-7015, DOI: 10.1890/10-1510.1, URL: https://esajournals.onlinelibrary.wiley.com/doi/full/10.1890/10-1510.1.

Barlow, H.B. (Sept. 1989), "Unsupervised Learning", *in*: *Neural Computation* 1.3, pp. 295–311, ISSN: 0899-7667, DOI: 10.1162/NECO.1989.1.3.295, URL: https://dx.doi.org/10.1162/neco.1989.1.3.295.

Barrell, Jeffrey et al. (Dec. 2015), "Evaluating the complementarity of acoustic and satellite remote sensing for seagrass landscape mapping", *in*: *International Journal of Remote Sensing* 36.16, pp. 4069–4094, ISSN: 0143-1161, DOI: 10.1080/01431161.2015.1076208, URL: https://doi.org/10.1080/01431161.2015.1076208.

Bernard, T G et al. (2021), "Beyond 2D landslide inventories and their rollover: synoptic 3D inventories and volume from repeat lidar data", *in*: *Earth Surface Dynamics* 9.4, pp. 1013–1044, DOI: 10.5194/esurf-9-1013-2021, URL: https://esurf.copernicus.org/articles/9/1013/2021/.

Blomley, R et al. (2016), "3D Semantic Labeling Of ALS Point Clouds by Exploiting Multi-scale, Multi-type Neighborhoods for Feature Extraction", *in*: *6th International Conference on Geographic Object-Based Image Analysis, GEOBIA 2016: Solutions & Synergies*, pp. 1–8.

Blomley, R. et al. (Sept. 2017), "Using Multi-scale Features for the 3D Semantic Labeling of Airborne Laser Scanning Data", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, 2W4, Copernicus GmbH, pp. 43–50, DOI: 10.5194/isprs-annals-IV-2-W4-43-2017.

Bouhdaoui, Anis et al. (2014), "The Water Bottom Geometry Effect on Peak Time Shifting in LiDAR Bathymetric Waveforms", *in*: *IEEE Geoscience and Remote Sensing Letters* 11.7, pp. 1285–1289, DOI: 10.1109/LGRS.2013.2292814{\"{i}}, URL: https://hal.archives-ouvertes.fr/hal-01522009.

Boulch, Alexandre et al. (Apr. 2018), "SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks", *in*: *Computers & Graphics* 71, pp. 189–198, ISSN: 0097-8493, DOI: 10.1016/J.CAG.2017.11.010.

Bradski, G (2000), "The OpenCV Library", *in*: *Dr. Dobb's Journal of Software Tools*.

Breiman, Leo (Oct. 2001), "Random Forests", *in*: *Machine Learning* 45.1, pp. 5–32, ISSN: 1573-0565, DOI: 10.1023/A:1010933404324, URL: https://link.springer.com/article/10.1023/A:1010933404324.

Breiman, Leo et al. (Jan. 1984), *Classification and regression trees*, CRC Press, ISBN: 9781351460491, DOI: 10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN-JEROME-FRIEDMAN-OLSHEN-CHARLES-STONE, URL: https://www.taylorfrancis.com/books/mono/10.1201/9781315139470/classification-regression-trees-leo-breiman-jerome-friedman-olshen-charles-stone.

Brodu, Nicolas et al. (2012), "3D terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: Applications in geomorphology", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 68.1, pp. 121–134, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2012.01.006, URL: http://dx.doi.org/10.1016/j.isprsjprs.2012.01.006.

Bromley, Jane et al. (1993), "Signature Verification using a "Siamese" Time Delay Neural Network", *in*: *Advances in Neural Information Processing Systems* 6.

Bruggisser, Moritz et al. (July 2017), "Retrieval of higher order statistical moments from full-waveform LiDAR data for tree species classification", *in*: *Remote Sensing of Environment* 196, pp. 28–41, ISSN: 00344257, DOI: 10.1016/j.rse.2017.04.025.

Butler, Josh D et al. (Dec. 2020), "A high-resolution remotely sensed benthic habitat map of the Qatari coastal zone", en, *in*: *Marine Pollution Bulletin* 160, p. 111634, ISSN: 0025-326X, DOI: 10.1016/j.

marpolbul.2020.111634, URL: https://www.sciencedirect.com/science/article/pii/S0025326X20307529.

Bye, I.J. et al. (Jan. 2017), "Estimating forest canopy parameters from satellite waveform LiDAR by inversion of the FLIGHT three-dimensional radiative transfer model", *in*: *Remote Sensing of Environment* 188, pp. 177–189, ISSN: 00344257, DOI: 10.1016/j.rse.2016.10.048, URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425716304278.

Cao, Lin et al. (July 2016), "Tree species classification in subtropical forests using small-footprint full-waveform LiDAR data", *in*: *International Journal of Applied Earth Observation and Geoinformation* 49, pp. 39–51, ISSN: 1569-8432, DOI: 10.1016/J.JAG.2016.01.007.

Caye Daudt, Rodrigo et al. (Aug. 2018), "Fully convolutional siamese networks for change detection", *in*: *Proceedings - International Conference on Image Processing, ICIP*, pp. 4063–4067, ISSN: 15224880, DOI: 10.1109/ICIP.2018.8451652.

Cazorla, Romain et al. (2022), "Reducing domain shift in synthetic data augmentation for semantic segmentation of 3D point clouds", *in*: *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2022-October, Institute of Electrical and Electronics Engineers Inc., pp. 1198–1205, ISBN: 9781665452588, DOI: 10.1109/SMC53654.2022.9945480.

Chauve, A. et al. (2009), "Advanced full-waveform lidar data echo detection: Assessing quality of derived terrain and tree height models in an alpine coniferous forest", *in*: *International Journal of Remote Sensing* 30.*19*, pp. 5211–5228, ISSN: 13665901, DOI: 10.1080/01431160903023009.

Chauve, Adrien et al. (Sept. 2007), "Processing Full-Waveform Lidar Data: Modelling Raw Signals", *in*: 36 (Part 3/W52), pp. 102–107, URL: https://hal-lirmm.ccsd.cnrs.fr/lirmm-00293129https://hal-lirmm.ccsd.cnrs.fr/lirmm-00293129/document.

Chehata, Nesrine et al. (Dec. 2009), "Airborne LiDAR Feature Selection for Urban Classification Using Random Forests", *in*: . *Laser scanning 2009*, pp. 207–212, URL: https://hal.archives-ouvertes.fr/hal-02384719https://hal.archives-ouvertes.fr/hal-02384719/document.

Chen, Biwu et al. (Apr. 2017), "Multispectral LiDAR Point Cloud Classification: A Two-Step Approach", *in*: *Remote Sensing* 9.*4*, pp. 373–390, ISSN: 2072-4292, DOI: 10.3390/RS9040373, URL: https://www.mdpi.com/2072-4292/9/4/373/htmhttps://www.mdpi.com/2072-4292/9/4/373.

Chen, Tianqi et al. (Aug. 2016), "XGBoost: A scalable tree boosting system", *in*: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-August-2016, pp. 785–794, DOI: 10.1145/2939672.2939785.

Cheng, Hui-Xian et al. (June 2023), "TransRVNet: LiDAR Semantic Segmentation With Transformer", *in*: *IEEE Transactions on Intelligent Transportation Systems* 24.*6*, pp. 5895–5907, ISSN: 1524-9050,

DOI: 10.1109/TITS.2023.3248117, URL: https://ieeexplore.ieee.org/document/10056404/.

Chung, Junyoung et al. (2014), "Empirical evaluation of gated recurrent neural networks on sequence modeling", *in*: *arXiv preprint arXiv:1412.3555*.

Chust, Guillem et al. (Dec. 2010), "Capabilities of the bathymetric Hawk Eye LiDAR for coastal habitat mapping: A case study within a Basque estuary", en, *in*: *Estuarine, Coastal and Shelf Science* 89.3, pp. 200–213, ISSN: 0272-7714, DOI: 10.1016/j.ecss.2010.07.002, URL: https://www.sciencedirect.com/science/article/pii/S0272771410002477.

Collin, Antoine et al. (Nov. 2012), "Merging land-marine realms: Spatial patterns of seamless coastal habitats using a multispectral LiDAR", en, *in*: *Remote Sensing of Environment* 123, pp. 390–399, ISSN: 0034-4257, DOI: 10.1016/j.rse.2012.03.015, URL: https://www.sciencedirect.com/science/article/pii/S0034425712001344.

Corcoran, Forrest et al. (2021), "Diffuse Attenuation Coefficient (Kd) from ICESat-2 ATLAS Spaceborne Lidar Using Random-Forest Regression", *in*: *Photogrammetric Engineering & Remote Sensing* 87.11, pp. 831–840, DOI: doi:10.14358/PERS.21-00013R2.

Cortes, Corinna et al. (Sept. 1995), "Support-vector networks", *in*: *Machine Learning 1995 20:3* 20.3, pp. 273–297, ISSN: 1573-0565, DOI: 10.1007/BF00994018, URL: https://link.springer.com/article/10.1007/BF00994018.

Costanza, Robert et al. (Mar. 1997), "The value of the world's ecosystem services and natural capital", en, *in*: *Nature* 387.6630, pp. 253–260, ISSN: 1476-4687, DOI: 10.1038/387253a0, URL: https://www.nature.com/articles/387253a0https://www.nature.com/articles/387253a0.pdf.

Cybenko, G. (Dec. 1989), "Approximation by superpositions of a sigmoidal function", *in*: *Mathematics of Control, Signals, and Systems* 2.4, pp. 303–314, ISSN: 09324194, DOI: 10.1007/BF02551274.

Dai, Wenxia et al. (Oct. 2018), "A new method for 3D individual tree extraction using multispectral airborne LiDAR point clouds", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 144, pp. 400–411, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2018.08.010.

Damodaran, Bharath Bhushan et al. (2018), *DeepJDOT: Deep Joint Distribution Optimal Transport for Unsupervised Domain Adaptation*.

Dash, M et al. (1997), "Feature Selection for Classification", *in*: *IDA ELSEVIER Intelligent Data Analysis* 1, pp. 131–156, URL: www.elsevier.comAocate/ida.

De Gélis, Iris (Apr. 2023), "Deep learning for change detection in 3D point clouds", PhD thesis, Université de Bretagne Sud, URL: https://theses.hal.science/tel-04449411.

Demantke, Jérôme et al. (2012), "Dimensionality Based Scale Selection in 3D LiDAR Point Clouds", *in*: *The international archives of the photogrammetry, remote sensing and spatial information sciences* 38, pp. 97–102, URL: https://hal.archives-ouvertes.fr/hal-02384758.

Demantké, Jérôme et al. (2012), "Streamed vertical rectangle detection in terrestrial laser scans for facade database production", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 1, pp. 99–104.

Deng, Shuang et al. (2021), "GA-NET: Global Attention Network for Point Cloud Semantic Segmentation", *in*: *IEEE Signal Processing Letters* 28, pp. 1300–1304, ISSN: 15582361, DOI: 10.1109/LSP.2021.3082851.

DeVries, Terrance et al. (Aug. 2017), "Improved Regularization of Convolutional Neural Networks with Cutout", *in*: *arXiv preprint arXiv:1708.04552*, URL: http://arxiv.org/abs/1708.04552.

Ding, Kai et al. (Feb. 2018), "An improved quadrilateral fitting algorithm for the water column contribution in airborne bathymetric lidar waveforms", *in*: *Sensors (Switzerland)* 18.2, ISSN: 14248220, DOI: 10.3390/s18020552.

Dong, Nanqing et al. (2018), "Few-Shot Semantic Segmentation with Prototype Learning", *in*: *BMVC* 3.4, pp. 1–13.

Dong, Weihua et al. (Aug. 2017), "Selection of LiDAR geometric features with adaptive neighborhood size for urban land cover classification", *in*: *International Journal of Applied Earth Observation and Geoinformation* 60, pp. 99–110, ISSN: 1569-8432, DOI: 10.1016/J.JAG.2017.04.003.

Dosovitskiy, A et al. (2020), "Transformers for image recognition at scale", *in*: *arXiv preprint arXiv:2010.11929*.

Drake, Jason B. et al. (Aug. 2002), "Sensitivity of large-footprint lidar to canopy structure and biomass in a neotropical rainforest", *in*: *Remote Sensing of Environment* 81.2-3, pp. 378–392, ISSN: 0034-4257, DOI: 10.1016/S0034-4257(02)00013-5.

Dubayah, Ralph et al. (June 2020), "The Global Ecosystem Dynamics Investigation: High-resolution laser ranging of the Earth's forests and topography", *in*: *Science of Remote Sensing* 1, p. 100002, ISSN: 2666-0172, DOI: 10.1016/J.SRS.2020.100002.

Duncanson, Laura et al. (June 2020), "Biomass estimation from simulated GEDI, ICESat-2 and NISAR across environmental gradients in Sonoma County, California", *in*: *Remote Sensing of Environment* 242, p. 111779, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2020.111779.

Eitel, Jan U.H. et al. (Dec. 2016), *Beyond 3-D: The new spectrum of lidar applications for earth and ecological sciences*, DOI: 10.1016/j.rse.2016.08.018.

Ekhtari, Nima et al. (June 2018), "Classification of airborne multispectral lidar point clouds for land cover mapping", *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 11.6, pp. 2068–2078, ISSN: 21511535, DOI: 10.1109/JSTARS.2018.2835483.

Enríquez, Susana et al. (Oct. 2019), "Structural complexity governs seagrass acclimatization to depth with relevant consequences for meadow production, macrophyte diversity and habitat carbon storage capacity", *in*: *Scientific Reports 2019 9:1* 9.1, pp. 1–14, ISSN: 2045-2322, DOI: 10.1038/s41598-019-51248-z, URL: https://www.nature.com/articles/s41598-019-51248-z.

Eren, Firat et al. (Mar. 2018), "Bottom characterization by using airborne lidar bathymetry (ALB) waveform features obtained from bottom return residual analysis", *in*: *Remote Sensing of Environment* 206, pp. 260–274, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2017.12.035.

Fang, Jinli et al. (Dec. 2022), "Gaussian convolution decomposition for non-Gaussian shaped pulsed LiDAR waveform", *in*: *Measurement Science and Technology* 34.3, p. 035203, ISSN: 0957-0233, DOI: 10.1088/1361-6501/ACA3C6, URL: https://iopscience.iop.org/article/10.1088/1361-6501/aca3c6https://iopscience.iop.org/article/10.1088/1361-6501/aca3c6/meta.

Feigels, Victor J (1992), "Lidars for Oceanological Research: Criteria for Comparison, Main Limitations, Perspectives", *in*: *SPIE Ocean Optics* 1750.9, pp. 473–484, URL: http://spiedl.org/terms.

Feng, Hengxin et al. (May 2022), "Enrich Features for Few-Shot Point Cloud Classification", *in*: *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2022-May, IEEE, pp. 2285–2289, ISBN: 978-1-6654-0540-9, DOI: 10.1109/ICASSP43922.2022.9747562, URL: https://ieeexplore.ieee.org/document/9747562/.

Fernandez-Diaz, Juan Carlos et al. (Feb. 2014), "Early results of simultaneous terrain and shallow water bathymetry mapping using a single-wavelength airborne LiDAR sensor", *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.2, pp. 623–635, ISSN: 19391404, DOI: 10.1109/JSTARS.2013.2265255.

Fernandez-Diaz, Juan Carlos et al. (Nov. 2016), "Capability assessment and performance metrics for the titan multispectral mapping lidar", *in*: *Remote Sensing* 8.11, ISSN: 20724292, DOI: 10.3390/rs8110936.

Feygels, Viktor I et al. (2003), "Airborne lidar system with variable-field-of-view receiver for water optical properties measurement", *in*: *Ocean Remote Sensing and Imaging II*, ed. by Robert J Frouin et al., vol. 5155, SPIE, 12 – 21, DOI: 10.1117/12.506976, URL: https://doi.org/10.1117/12.506976.

Gao, Jay (Feb. 2009), "Bathymetric mapping by means of remote sensing: methods, accuracy and limitations", en, *in*: *Progress in Physical Geography: Earth and Environment* 33.1, pp. 103–116, ISSN: 0309-1333, 1477-0296, DOI: 10.1177/0309133309105657, URL: http://journals.sagepub.com/doi/10.1177/0309133309105657https://pdfs.semanticscholar.org/627a/27ce12e9aa4135f8e352059b9c06f56ddfc4.pdf.

Gao, Yongbin et al. (Feb. 2023), "LFT-Net: Local Feature Transformer Network for Point Clouds Analysis", *in*: *IEEE Transactions on Intelligent Transportation Systems* 24.2, pp. 2158–2168, ISSN: 15580016, DOI: 10.1109/TITS.2022.3140355.

Garcia, Victor et al. (Nov. 2018), "Few-Shot Learning with Graph Neural Networks", *in*: *6th International Conference on Learning Representations, ICLR 2018*, URL: http://arxiv.org/abs/1711.04043.

Garstka, Jens et al. (2016), "Evaluation of local 3-D point cloud descriptors in terms of suitability for object classification", *in*: *ICINCO 2016 - Proceedings of the 13th International Conference on Informatics in Control, Automation and Robotics*, vol. 2, SciTePress, pp. 540–547, ISBN: 9789897581984, DOI: 10.5220/0006011505400547.

Gastellu-Etchegorry, Jean-Philippe et al. (2016), "Simulation of satellite, airborne and terrestrial LiDAR with DART (I): Waveform simulation with quasi-Monte Carlo ray tracing", *in*: *Remote Sensing of Environment* 184, pp. 418–435, ISSN: 0034-4257, DOI: https://doi.org/10.1016/j.rse.2016.07.010, URL: https://www.sciencedirect.com/science/article/pii/S003442571630267X.

Gatziolis, Demetrios et al. (2008), *A Guide to LIDAR Data Acquisition and Processing for the Forests of the Pacific Northwest*, tech. rep.

Gélis, Iris de et al. (July 2021), "Change Detection in Urban Point Clouds: An Experimental Comparison with Simulated 3D Datasets", *in*: *Remote Sensing 2021, Vol. 13, Page 2629* 13.13, p. 2629, ISSN: 2072-4292, DOI: 10.3390/RS13132629, URL: https://www.mdpi.com/2072-4292/13/13/2629/htmhttps://www.mdpi.com/2072-4292/13/13/2629.

Géron, Aurélien (2022), *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*," O'Reilly Media, Inc."

Ghahramani, Zoubin (2004), "Unsupervised learning", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 3176, pp. 72–112, ISSN: 16113349, DOI: 10.1007/978-3-540-28650-9{_}5/COVER, URL: https://link.springer.com/chapter/10.1007/978-3-540-28650-9_5.

Girardeau-Montaut, Daniel (2022), *CloudCompare (version 2.12.4) [GPL software]. (2022). Retrieved from http://www.cloudcompare.org/*.

Gong, Wei et al. (Sept. 2015), "Investigating the Potential of Using the Spatial and Spectral Information of Multispectral LiDAR for Object Classification", *in*: *Sensors* 15.9, pp. 21989–22002, ISSN: 1424-8220, DOI: 10.3390/s150921989, URL: http://www.mdpi.com/1424-8220/15/9/21989.

Goodfellow, Ian et al. (2016), *Deep Learning*, MIT Press.

Goodfellow, Ian J et al. (2014), "Generative Adversarial Nets", *in*: *Advances in Neural Information Processing Systems* 27, URL: http://www.github.com/goodfeli/adversarial.

Graham, Benjamin et al. (2018), "3D Semantic Segmentation with Submanifold Sparse Convolutional Networks", *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition 2018*, pp. 9224–9232, URL: https://github.com/facebookresearch/.

Grande, Maitane et al. (2009), *Assessment of the discrimination potential of bathymetric LIDAR and multispectral imagery for intertidal and subtidal habitats*, URL: https://www.researchgate.net/profile/Jose-Fernandes-35/publication/255962103_Assessment_of_the_discrimination_potential_of_bathymetric_LIDAR_and_multispectral_imagery_for_intertidal_and_subtidal_habitats/links/0deec520fcfed50f14000000/Assessment-of-the-di.

Gross, Hermann et al. (2006), "Extraction of lines from laser point clouds", *in*: *Symposium of ISPRS Commission III: Photogrammetric Computer Vision PCV06. International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36.3, pp. 86–91.

Guan, Haiyan et al. (2012), "Random forests-based feature selection for land-use classification using lidar data and orthoimagery", *in*: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39, pp. 203–208.

Guenther, Gary C (1985), *Airborne lidar bathymetry*, tech. rep., URL: jsessionid=649D439DC44B91B08ADEA6C65doi=10.1.1.137.8332&rep=rep1&type=pdf.

Guenther, Gary C et al. (2000), "Meeting the Accuracy Challenge in Airborne Lidar Bathymetry", *in*: *EARSeL eProceedings* 1.1, pp. 1–27.

Guiotte, Florent et al. (Nov. 2020), "Semantic Segmentation of LiDAR Points Clouds: Rasterization beyond Digital Elevation Models", *in*: *IEEE Geoscience and Remote Sensing Letters* 17.11, pp. 2016–2019, ISSN: 15580571, DOI: 10.1109/LGRS.2019.2958858.

Guo, Meng Hao et al. (2021a), "PCT: Point cloud transformer", *in*: *Computational Visual Media* 7.2, pp. 187–199, ISSN: 20960662, DOI: 10.1007/s41095-021-0229-5, URL: https://doi.org/10.1007/s41095-021-0229-5.

Guo, Yadong et al. (2023), "Water-land classification for single-wavelength airborne LiDAR bathymetry based on waveform feature statistics and point cloud neighborhood analysis", *in*: *International Journal of Applied Earth Observation and Geoinformation* 118, p. 103268, ISSN: 1569-8432, DOI: https://doi.org/10.1016/j.jag.2023.103268, URL: https://www.sciencedirect.com/science/article/pii/S1569843223000900.

Guo, Yulan et al. (2021b), "Deep Learning for 3D Point Clouds: A Survey", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.12, pp. 4338–4364, ISSN: 1939-3539, DOI: 10.1109/TPAMI.

2020.3005434, URL: https://arxiv.org/pdf/1912.12033https://ieeexplore.ieee.org/abstract/document/9127813.

Guyon, Isabelle et al. (2003), "An Introduction to Variable and Feature Selection", *in*: *Journal of Machine Learning Research* 3, pp. 1157–1182.

Guyot, Alexandre et al. (Sept. 2019), "Airborne Hyperspectral Imaging for Submerged Archaeological Mapping in Shallow Water Environments", *in*: *Remote Sensing 2019, Vol. 11, Page 2237* 11.19, p. 2237, ISSN: 2072-4292, DOI: 10.3390/RS11192237, URL: https://www.mdpi.com/2072-4292/11/19/2237/htmhttps://www.mdpi.com/2072-4292/11/19/2237.

Hackel, Timo et al. (June 2016), "Fast Semantic Segmentation Of 3d Point Clouds With Strongly Varying Density", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* III-3, pp. 177–184, ISSN: 2194-9050, DOI: 10.5194/isprs-annals-III-3-177-2016, URL: https://isprs-annals.copernicus.org/articles/III-3/177/2016/.

Hackel, Timo et al. (Apr. 2017), "Semantic3D.net: A new Large-scale Point Cloud Classification Benchmark", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1/W1.1*W1*, pp. 91–98, URL: http://arxiv.org/abs/1704.03847.

Hamraz, Hamid et al. (Mar. 2019), "Deep learning for conifer/deciduous classification of airborne LiDAR 3D point clouds representing individual trees", en, *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 158, pp. 219–230, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2019.10.011, URL: https://linkinghub.elsevier.com/retrieve/pii/S0924271619302485https://arxiv.org/ftp/arxiv/papers/1802/1802.08872.pdf.

Hancock, Steven et al. (July 2015), "Waveform lidar over vegetation: An evaluation of inversion methods for estimating return energy", *in*: *Remote Sensing of Environment* 164, pp. 208–224, ISSN: 00344257, DOI: 10.1016/j.rse.2015.04.013.

Hansen, Signe Schilling et al. (Oct. 2021), "Classification of Boulders in Coastal Environments Using Random Forest Machine Learning on Topo-Bathymetric LiDAR Data", en, *in*: *Remote Sensing* 13.20, pp. 4101–4126, ISSN: 2072-4292, DOI: 10.3390/rs13204101, URL: https://www.mdpi.com/2072-4292/13/20/4101.

Hastie, Trevor et al. (2009), "Multi-class AdaBoost", *in*: *Statistics and Its Interface* 2.3, pp. 349–360, ISSN: 1938-7997, DOI: 10.4310/SII.2009.V2.N3.A8, URL: https://www.intlpress.com/site/pub/pages/journals/items/sii/content/vols/0002/0003/a008/abstract.php.

Hazirbas, Caner et al. (2017), "FuseNet: Incorporating depth into semantic segmentation via fusion-based CNN architecture", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 10111 LNCS, pp. 213–228, ISSN: 16113349,

DOI: 10.1007/978-3-319-54181-5{_}14/COVER, URL: https://link.springer.com/chapter/10.1007/978-3-319-54181-5_14.

He, Kaiming et al. (2016), *Deep Residual Learning for Image Recognition*, URL: http://image-net.org/challenges/LSVRC/2015/.

He, Shuting et al. (2023), "Prototype Adaption and Projection for Few- and Zero-Shot 3D Point Cloud Semantic Segmentation", *in*: *IEEE Transactions on Image Processing* 32, pp. 3199–3211, ISSN: 1057-7149, DOI: 10.1109/TIP.2023.3279660, URL: https://ieeexplore.ieee.org/document/10138737/.

Heinzel, Johannes et al. (2011), "Exploring full-waveform LiDAR parameters for tree species classification", *in*: *International Journal of Applied Earth Observation and Geoinformation* 13.1, pp. 152–160, ISSN: 1872826X, DOI: 10.1016/j.jag.2010.09.010.

Hermosilla, Pedro et al. (Dec. 2018), "Monte Carlo convolution for learning on non-uniformly sampled point clouds", *in*: *ACM Transactions on Graphics* 37.6, pp. 1–12, ISSN: 0730-0301, DOI: 10.1145/3272127.3275110, URL: https://dl.acm.org/doi/10.1145/3272127.3275110.

Hilldale, Robert C. et al. (Apr. 2008), "Assessing the ability of airborne LiDAR to map river bathymetry", *in*: *Earth Surface Processes and Landforms* 33.5, pp. 773–783, ISSN: 1096-9837, DOI: 10.1002/ESP.1575, URL: https://onlinelibrary.wiley.com/doi/full/10.1002/esp.1575https://onlinelibrary.wiley.com/doi/abs/10.1002/esp.1575https://onlinelibrary.wiley.com/doi/10.1002/esp.1575.

Himmelsbach, M et al. (2009), "Real-time Object Classification in 3D Point Clouds Using Point Feature Histograms", *in*: *The 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp. 994–1000, ISBN: 9781424438044.

Hmida, Sahar Ben et al. (2017), "Crop Biophysical Properties Estimation Based on LiDAR Full-Waveform Inversion Using the DART RTM", *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 10.11, pp. 4853–4868, DOI: 10.1109/JSTARS.2017.2763242.

Hochreiter, Sepp et al. (Nov. 1997), "Long Short-Term Memory", *in*: *Neural Computation* 9.8, pp. 1735–1780, ISSN: 08997667, DOI: 10.1162/NECO.1997.9.8.1735.

Höfle, Bernhard et al. (Dec. 2007), "Correction of laser scanning intensity data: Data and model-driven approaches", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 62.6, pp. 415–433, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2007.05.008.

Hofton, Michalle A. et al. (July 2000), "Decomposition of laser altimeter waveforms", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 38.4 II, pp. 1989–1996, ISSN: 01962892, DOI: 10.1109/36.851780.

Hu, Qingyong et al. (June 2020), "RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds", *in*: *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, pp. 11105–11114, ISBN: 978-1-7281-7168-5, DOI: 10.1109/CVPR42600.2020.01112, URL: https://ieeexplore.ieee.org/document/9156466/.

Hu, Shanjiang et al. (Nov. 2019), "Classification of sea and land waveforms based on deep learning for airborne laser bathymetry", *in*: *Hongwai yu Jiguang Gongcheng/Infrared and Laser Engineering* 48.11, ISSN: 10072276, DOI: 10.3788/IRLA201948.1113004.

Hua, Binh-Son et al. (June 2018), "Pointwise Convolutional Neural Networks", *in*: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 984–993, ISBN: 978-1-5386-6420-9, DOI: 10.1109/CVPR.2018.00109, URL: https://ieeexplore.ieee.org/document/8578207/.

Huang, Qiangui et al. (2018), "Recurrent Slice Networks for 3D Segmentation of Point Clouds", *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2626–2632, URL: https://github.com/qianguih/RSNet.

Huang, Rong et al. (July 2021), "GraNet: Global relation-aware attentional network for semantic segmentation of ALS point clouds", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 177, pp. 1–20, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2021.04.017.

Hui, Le et al. (2021), "Superpoint Network for Point Cloud Oversegmentation", *in*: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5510–5519.

Im, Jungho et al. (Apr. 2008), "Object-Based Land Cover Classification Using High-Posting-Density LiDAR Data", *in*: *GIScience & Remote Sensing* 45.2, pp. 209–228, ISSN: 1548-1603, DOI: 10.2747/1548-1603.45.2.209, URL: https://www.tandfonline.com/doi/full/10.2747/1548-1603.45.2.209.

Irish, J L et al. (1998), "Coastal engineering applications of high-resolution lidar bathymetry", *in*: *Coastal Engineering* 35, pp. 47–71.

Jaboyedoff, Michel et al. (Mar. 2012), "Use of LIDAR in landslide investigations: A review", *in*: *Natural Hazards* 61.1, pp. 5–28, ISSN: 0921030X, DOI: 10.1007/s11069-010-9634-2.

Jakubowski, Marek K et al. (Jan. 2013), "Tradeoffs between lidar pulse density and forest measurement accuracy", en, *in*: *Remote Sensing of Environment* 130, pp. 245–253, ISSN: 0034-4257, DOI: 10.1016/j.rse.2012.11.024, URL: https://www.sciencedirect.com/science/article/pii/S0034425712004567.

Janowski, Lukasz et al. (May 2022), "Automatic classification and mapping of the seabed using airborne LiDAR bathymetry", *in*: *Engineering Geology* 301, ISSN: 00137952, DOI: 10.1016/J.ENGGEO.2022.106615.

Jaritz, Maximilian et al. (Feb. 2023), "Cross-Modal Learning for Domain Adaptation in 3D Semantic Segmentation", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45.2, pp. 1533–1544, ISSN: 0162-8828, DOI: 10.1109/TPAMI.2022.3159589, URL: https://ieeexplore.ieee.org/document/9737217/.

Jiang, Jincen et al. (Oct. 2021), "Unsupervised Representation Learning for 3D Point Cloud Data", *in*: URL: http://arxiv.org/abs/2110.06632.

Jing, Zhuangwei et al. (June 2021), "Multispectral LiDAR Point Cloud Classification Using SE-PointNet++", *in*: *Remote Sensing* 13, pp. 2516–2535, ISSN: 2072-4292, DOI: 10.3390/rs13132516, URL: https://www.mdpi.com/2072-4292/13/13/2516.

Jutzi, Boris et al. (2003), "Estimation and measurement of backscattered signals from pulsed laser radar", *in*: *SPIE Image and Signal Processing for Remote Sensing* 4885.8, pp. 256–267.

Jutzi, Boris et al. (Nov. 2006), "Range determination with waveform recording laser systems using a Wiener Filter", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 61.2, pp. 95–107, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2006.09.001.

— (2007), "Simulation and analysis of full-waveform laser data of urban objects", *in*: *2007 IEEE Urban Remote Sensing Joint Event*, pp. 1–5.

Kashani, Alireza G et al. (Nov. 2015), "A Review of LIDAR Radiometric Processing: From Ad Hoc Intensity Correction to Rigorous Radiometric Calibration", en, *in*: *Sensors* 15.11, pp. 28099–28128, DOI: 10.3390/s151128099, URL: https://www.mdpi.com/1424-8220/15/11/28099https://www.mdpi.com/1424-8220/15/11/28099/pdf.

Khan, Salman et al. (Jan. 2021), "Transformers in Vision: A Survey", *in*: DOI: 10.1145/3505244, URL: http://arxiv.org/abs/2101.01169http://dx.doi.org/10.1145/3505244.

Khosla, Prannay et al. (2020), "Supervised Contrastive Learning", *in*: *Advances in Neural Information Processing Systems* 33, pp. 18661–18673, URL: https://t.ly/supcon.

Kim, Minsu et al. (Dec. 2016), "Modeling of airborne bathymetric lidar waveforms", *in*: *Journal of Coastal Research* 76.spi, pp. 18–30, ISSN: 15515036, DOI: 10.2112/SI76-003.

Kingma, Diederik P et al. (2013), "Auto-encoding variational bayes", *in*: *arXiv preprint arXiv:1312.6114*.

Kinzel, Paul J et al. (Feb. 2013), "Mapping River Bathymetry With a Small Footprint Green LiDAR: Applications and Challenges", en, *in*: *JAWRA Journal of the American Water Resources Association* 49.1, pp. 183–204, ISSN: 1093474X, DOI: 10.1111/jawr.12008, URL: http://doi.wiley.com/10.1111/jawr.12008https://wwwbrr.cr.usgs.gov/gstl/kinzel_papers/Kinzel_JAWRA_2013.pdf.

Klonowski, Wojciech M et al. (Nov. 2007), "Retrieving key benthic cover types and bathymetry from hyperspectral imagery", *in*: *Journal of Applied Remote Sensing* 1.1, p. 11505, ISSN: 1931-3195, 1931-3195,

DOI: 10.1117/1.2816113, URL: https://www.spiedigitallibrary.org/journals/journal-of-applied-remote-sensing/volume-1/issue-1/011505/Retrieving-key-benthic-cover-types-and-bathymetry-from-hyperspectral-imagery/10.1117/1.2816113.fullhttps://www.spiedigitallibrary.org/journals/journal-o.

Koetz, B et al. (2006), "Inversion of a lidar waveform model for forest biophysical parameter estimation", *in*: *IEEE Geoscience and Remote Sensing Letters* 3.*1*, pp. 49–53, DOI: 10.1109/LGRS.2005.856706.

Kogut, Tomasz et al. (May 2019), "Improvement of Full Waveform Airborne Laser Bathymetry Data Processing based on Waves of Neighborhood Points", *in*: *Remote Sensing 2019, Vol. 11, Page 1255* 11.*10*, p. 1255, ISSN: 2072-4292, DOI: 10.3390/RS11101255, URL: https://www.mdpi.com/2072-4292/11/10/1255/htmhttps://www.mdpi.com/2072-4292/11/10/1255.

Komatsu, Teruhisa et al. (Dec. 2003), "Use of multi-beam sonar to map seagrass beds in Otsuchi Bay on the Sanriku Coast of Japan", en, *in*: *Aquatic Living Resources* 16.*3*, pp. 223–230, ISSN: 0990-7440, 1765-2952, DOI: 10.1016/S0990-7440(03)00045-7, URL: https://www.cambridge.org/core/journals/aquatic-living-resources/article/abs/use-of-multibeam-sonar-to-map-seagrass-beds-in-otsuchi-bay-on-the-sanriku-coast-of-japan/353C22680546C29D1A0642B7928D001E.

Kopilevich, Yu I et al. (2008), "Mathematical modeling of the input signals of oceanological lidars", *in*: *J. Opt. Technol.* 75.*5*, pp. 321–326.

Kopilevich, Yuri I et al. (2005), "Measurement of ocean water optical properties and seafloor reflectance with scanning hydrographic operational airborne lidar survey (SHOALS): I. Theoretical background", *in*: *Remote Sensing of the Coastal Oceanic Environment*, ed. by Robert J Frouin et al., vol. 5885, SPIE, p. 58850D, DOI: 10.1117/12.618923, URL: https://doi.org/10.1117/12.618923.

Korpela, Ilkka et al. (Aug. 2023), "Influence of phenology on waveform features in deciduous and coniferous trees in airborne LiDAR", *in*: *Remote Sensing of Environment* 293, p. 113618, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2023.113618.

Kotsiantis, S. B. et al. (2007), "Supervised machine learning: A review of classification techniques", *in*: *Emerging artificial intelligence applications in computer engineering* 160.*1*, pp. 3–24.

Krishnan, Sriram et al. (2011), "OpenTopography: A Services Oriented Architecture for Community Access to LIDAR Topography", *in*: *Proceedings of the 2nd International Conference on Computing for Geospatial Research & Applications*, COM.Geo '11, New York, NY, USA: Association for Computing Machinery, ISBN: 9781450306812, DOI: 10.1145/1999320.1999327, URL: https://doi.org/10.1145/1999320.1999327.

Kukko, Antero et al. (Aug. 2012), "Multiplatform Mobile Laser Scanning: Usability and Performance", *in*: *Sensors 2012, Vol. 12, Pages 11712-11733* 12.*9*, pp. 11712–11733, ISSN: 1424-8220, DOI: 10.3390/S120911712,

URL: https://www.mdpi.com/1424-8220/12/9/11712/htmhttps://www.mdpi.com/1424-8220/12/9/11712.

Kumpumäki, Teemu et al. (Dec. 2015), "Data-Driven Approach to Benthic Cover Type Classification Using Bathymetric LiDAR Waveform Analysis", en, *in*: *Remote Sensing* 7.10, pp. 13390–13409, DOI: 10.3390/rs71013390, URL: https://www.mdpi.com/2072-4292/7/10/13390https://www.mdpi.com/2072-4292/7/10/13390/pdf.

Kutser, Tiit et al. (Dec. 2006), "Assessing suitability of multispectral satellites for mapping benthic macroalgal cover in turbid coastal waters by means of model simulations", en, *in*: *Estuarine, Coastal and Shelf Science* 67.3, pp. 521–529, ISSN: 0272-7714, DOI: 10.1016/j.ecss.2005.12.004, URL: http://www.sciencedirect.com/science/article/pii/S0272771405004221.

Kutser, Tiit et al. (Nov. 2020), "Remote sensing of shallow waters – A 50 year retrospective and future directions", en, *in*: *Remote Sensing of Environment* 240, p. 111619, ISSN: 0034-4257, DOI: 10.1016/j.rse.2019.111619, URL: https://www.sciencedirect.com/science/article/pii/S003442571930639X.

Lague, D et al. (Nov. 2020), "Topo-bathymetric airborne LiDAR for fluvial-geomorphology analysis", en, *in*: *Developments in Earth Surface Processes*, ed. by Paolo Tarolli et al., vol. 23, Remote Sensing of Geomorphology, Elsevier, pp. 25–54, DOI: 10.1016/B978-0-444-64177-9.00002-3, URL: https://www.sciencedirect.com/science/article/pii/B9780444641779000023https://www.sciencedirect.com/science/article/abs/pii/B9780444641779000023.

Lai, Xin et al. (2022), "Stratified Transformer for 3D Point Cloud Segmentation", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 8500–8509.

Landrieu, Loic et al. (2018), "Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs", *in*: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4558–4567, ISSN: 10636919, DOI: 10.1109/CVPR.2018.00479.

Landrieu, Loic et al. (2019), "Point Cloud Oversegmentation with Graph-Structured Deep Metric Learning", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7440–7449.

Landrieu, Loïc et al. (Oct. 2017), "A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 132, pp. 102–118, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2017.08.010.

Lang, K J (1988), "The development of the time-delay neural network architecture for speech recognition", *in*: *Technical Report CMU-CS-88-152*.

Lang, Nico et al. (Jan. 2022), "Global canopy height regression and uncertainty estimation from GEDI LIDAR waveforms with deep ensembles", *in*: *Remote Sensing of Environment* 268, p. 112760, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2021.112760.

Laslier, Marianne et al. (Mar. 2019), "Mapping Riparian Vegetation Functions Using 3D Bispectral LiDAR Data", *in*: *Water* 11.3, pp. 483–502, ISSN: 2073-4441, DOI: 10.3390/w11030483, URL: https://www.mdpi.com/2073-4441/11/3/483.

Launeau, Patrick et al. (Apr. 2018), "Full-Waveform LiDAR Pixel Analysis for Low-Growing Vegetation Mapping of Coastal Foredunes in Western France", en, *in*: *Remote Sensing* 10.5, pp. 669–683, ISSN: 2072-4292, DOI: 10.3390/rs10050669, URL: https://www.mdpi.com/2072-4292/10/5/669https://www.mdpi.com/2072-4292/10/5/669/pdf.

Leatherman, Stephen P. et al. (Apr. 2000), "Sea level rise shown to drive coastal erosion", en, *in*: *Eos, Transactions American Geophysical Union* 81.6, p. 55, ISSN: 0096-3941, DOI: 10.1029/00EO00034, URL: http://doi.wiley.com/10.1029/00EO00034.

Lecun, Y et al. (1989), "Handwritten Digit Recognition with a Back-Propagation Network", *in*: *Advances in neural information processing systems* 2.

Lecun, Yann et al. (1995), "Convolutional Networks for Images, Speech, and Time-Series", *in*: *The handbook of brain theory and neural networks* 3361.10, p. 1995.

Lecun, Yann et al. (1998), "Gradient-based Learning Applied To Document Recognition - Proceedings of the IEEE", *in*: *Proceedings of the IEEE* 86.11, pp. 2278–2324, DOI: 10.1109/5.726791{\"{i}}, URL: https://hal.science/hal-03926082.

Lecun, Yann et al. (May 2015), "Deep learning", *in*: *Nature* 521.7553, pp. 436–444, ISSN: 14764687, DOI: 10.1038/nature14539.

Leigh, Holly W et al. (2016), "Using dual-wavelength, full-waveform airborne lidar for surface classification and vegetation characterization", *in*: *Journal of Applied Remote Sensing* 10.4, p. 45001, DOI: 10.1117/1.JRS.10.045001, URL: https://doi.org/10.1117/1.JRS.10.045001.

Lesser, M. P. et al. (Dec. 2007), "Bathymetry, water optical properties, and benthic classification of coral reefs using hyperspectral remote sensing imagery", *in*: *Coral Reefs* 26.4, pp. 819–829, ISSN: 0722-4028, DOI: 10.1007/s00338-007-0271-5, URL: http://link.springer.com/10.1007/s00338-007-0271-5.

Letard, M et al. (Mar. 2022a), "Classification of Land-Water Continuum Habitats Using Exclusively Airborne Topobathymetric Lidar Green Waveforms and Infrared Intensity Point Clouds", en, *in*: *Remote Sensing* 14.2, pp. 341–371, ISSN: 2072-4292, DOI: 10.3390/rs14020341, URL: https://www.mdpi.com/2072-4292/14/2/341https://www.mdpi.com/2072-4292/14/2/341/pdf.

Letard, M. et al. (May 2022b), "Using Bispectral Full-Waveform LiDAR to Map Seamless Coastal Habitats in 3D", *in*: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XLIII-B3-2022.*B3-2022*, pp. 463–470, ISSN: 2194-9034, DOI: 10.5194/isprs-archives-XLIII-B3-2022-463-2022, URL: https://isprs-archives.copernicus.org/articles/XLIII-B3-2022/463/2022/.

Letard, Mathilde et al. (Sept. 2021a), "Classification of coastal and estuarine ecosystems using full-waveform topo-bathymetric lidar data and artificial intelligence", *in*: *OCEANS 2021: San Diego – Porto*, vol. 2021-September, IEEE, pp. 1–10, ISBN: 978-0-692-93559-0, DOI: 10.23919/OCEANS44145.2021.9705797, URL: https://ieeexplore.ieee.org/document/9705797/.

Letard, Mathilde et al. (July 2021b), "Towards 3D Mapping of Seagrass Meadows with Topo-Bathymetric Lidar Full Waveform Processing", *in*: *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, IEEE, pp. 8069–8072, ISBN: 978-1-6654-0369-6, DOI: 10.1109/IGARSS47720.2021.9554262, URL: https://ieeexplore.ieee.org/document/9554262/.

Letard, Mathilde et al. (2023), "Bathymetric LiDAR Waveform Decomposition with Temporal Attentive Encoder-Decoders", *in*: *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 4435–4438, DOI: 10.1109/IGARSS52108.2023.10281841.

Li, Jiawei et al. (2022), "Self-Supervised Point Cloud Learning in Few-Shot Scenario by Point Up-Sampling and Mutual Information Neural Estimation", *in*: *2022 7th International Conference on Computer and Communication Systems, ICCCS 2022*, Institute of Electrical and Electronics Engineers Inc., pp. 304–309, ISBN: 9781665450607, DOI: 10.1109/ICCCS55155.2022.9846258.

Li, Yangyan et al. (Jan. 2018), "PointCNN: Convolution On X-Transformed Points", *in*: *Advances in neural information processing systems 31*, URL: http://arxiv.org/abs/1801.07791.

Liang, Gang et al. (2022), "MVCNN: A Deep Learning-Based Ocean–Land Waveform Classification Network for Single-Wavelength LiDAR Bathymetry", *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, pp. 1–19, ISSN: 1939-1404, DOI: 10.1109/JSTARS.2022.3229062, URL: https://ieeexplore.ieee.org/document/9984840/.

Liang, Xinlian et al. (May 2016), "Terrestrial laser scanning in forest inventories", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 115, pp. 63–77, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2016.01.006.

Lim, Bryan et al. (Apr. 2020), "Time Series Forecasting With Deep Learning: A Survey", *in*: DOI: 10.1098/rsta.2020.0209, URL: http://arxiv.org/abs/2004.13408http://dx.doi.org/10.1098/rsta.2020.0209.

Lim, Ee Hui et al. (Oct. 2009), "3D terrestrial LIDAR classifications with super-voxels and multi-scale Conditional Random Fields", *in*: *Computer-Aided Design* 41.10, pp. 701–710, ISSN: 0010-4485, DOI: 10.1016/J.CAD.2009.02.010.

Lin, Yaping et al. (June 2021), "Local and global encoder network for semantic segmentation of Airborne laser scanning point clouds", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 176, pp. 151–168, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2021.04.016.

Lindberg, Eva et al. (Mar. 2012), "Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data", en, *in*: *Remote Sensing of Environment* 118, pp. 151–161, ISSN: 00344257, DOI: 10.1016/j.rse.2011.11.015, URL: https://linkinghub.elsevier.com/retrieve/pii/S003442571100410X.

Liu, Chang et al. (2022a), "A Robust Deconvolution Method of Airborne LiDAR Waveforms for Dense Point Clouds Generation in Forest", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 60, ISSN: 15580644, DOI: 10.1109/TGRS.2021.3061166.

Liu, Gangping et al. (Mar. 2022b), "Full-waveform LiDAR echo decomposition based on dense and residual neural networks", *in*: *Applied Optics* 61.9, F15, ISSN: 1559-128X, DOI: 10.1364/AO.444910, URL: https://opg.optica.org/abstract.cfm?URI=ao-61-9-F15.

Liu, Maohua et al. (June 2021a), "Tree species classification of LiDAR data based on 3D deep learning", *in*: *Measurement* 177, p. 109301, ISSN: 0263-2241, DOI: 10.1016/J.MEASUREMENT.2021.109301.

Liu, Yaxin et al. (Oct. 2021b), "A deep learning method for LiDAR bathymetry waveforms processing", *in*: *2021 International Conference on Neural Networks, Information and Communication Engineering*, ed. by Zhiyong Zhang, vol. 11933, SPIE, p. 45, ISBN: 9781510647374, DOI: 10.1117/12.2615112, URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/11933/2615112/A-deep-learning-method-for-LiDAR-bathymetry-waveforms-processing/10.1117/12.2615112.full.

Lu, Xiaomei et al. (2021), "New Ocean Subsurface Optical Properties From Space Lidars: CALIOP/CALIPSO and ATLAS/ICESat-2", *in*: *Earth and Space Science* 8.10, e2021EA001839, DOI: https://doi.org/10.1029/2021EA001839, URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EA001839.

Lu, Xiaomei et al. (Sept. 2022), "Nearshore bathymetry and seafloor property studies from Space lidars: CALIPSO and ICESat-2", *in*: *Optics Express* 30.20, p. 36509, ISSN: 10944087, DOI: 10.1364/oe.471444.

Lu, Xiaomei et al. (Aug. 2023), "Lidar attenuation coefficient in the global oceans: insights from ICESat-2 mission", *in*: *Opt. Express* 31.18, pp. 29107–29118, DOI: 10.1364/OE.498053, URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-31-18-29107.

Lu, Yan et al. (2012), "Simplified Markov Random Fields for Efficient Semantic Labeling of 3D Point Clouds", *in*: *IEEE/RSJ International Conference on Intelligent Robots and Systems*. Pp. 2690–2697, ISBN: 9781467317368.

Lundberg, Scott M et al. (2017), "A unified approach to interpreting model predictions", *in*: *Advances in Neural Information Processing Systems*, vol. 2017-Decem, pp. 4766–4775, URL: https://github.com/slundberg/shap.

Ma, Yi et al. (Aug. 2019), "Bathymetry retrieval method of LiDAR waveform based on multi-Gaussian functions", *in*: *Journal of Coastal Research* 90, pp. 324–331, ISSN: 15515036, DOI: 10.2112/SI90-041.1.

Mader, D. et al. (June 2019), "Detection and extraction of water bottom topography from laserbathymetry data by using full-waveform-stacking techniques", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, 2/W13, International Society for Photogrammetry and Remote Sensing, pp. 1053–1059, DOI: 10.5194/isprs-archives-XLII-2-W13-1053-2019.

Mader, D. et al. (Oct. 2023), "Volumetric nonlinear ortho full-waveform stacking in airborne LiDAR bathymetry for reliable water bottom point detection in shallow waters", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 204, pp. 145–162, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2023.08.014.

Mader, David et al. (Apr. 2021), "Potential of a Non-linear Full-Waveform Stacking Technique in Airborne LiDAR Bathymetry: Demonstration of Full-Waveform Stacking Techniques on Data from the Elbe River", *in*: *Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 89.2, pp. 139–158, ISSN: 25122819, DOI: 10.1007/s41064-021-00147-y.

Magruder, Lori A et al. (2010), "Lidar waveform stacking techniques for faint ground return extraction", *in*: *Journal of Applied Remote Sensing* 4.1, p. 43501, DOI: 10.1117/1.3299657, URL: https://doi.org/10.1117/1.3299657.

Mallet, Clment et al. (Dec. 2010), "A marked point process for modeling lidar waveforms", *in*: *IEEE Transactions on Image Processing* 19.12, pp. 3204–3221, ISSN: 10577149, DOI: 10.1109/TIP.2010.2052825.

Mallet, Clément et al. (Aug. 2009), "Full-waveform topographic lidar: State-of-the-art", fr, *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.1, pp. 1–16, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2008.09.007, URL: https://linkinghub.elsevier.com/retrieve/pii/S0924271608000993http://documents.irevues.inist.fr/bitstream/handle/2042/16841/01-Mallet(coul).pdf?sequence=1.

Mallet, Clément et al. (Mar. 2011), "Relevance assessment of full-waveform lidar data for urban area classification", en, *in*: *ISPRS Journal of Photogrammetry and Remote Sensing*, Advances in LIDAR Data Processing and Applications 66.*6, Supplement*, S71–S84, ISSN: 0924-2716, DOI: 10.1016/j.isprsjprs.2011.09.008, URL: https://www.sciencedirect.com/science/article/pii/S0924271611001055.

Mandlburger, G. et al. (Oct. 2013), "Analyzing near water surface penetration in laser bathymetry – A case study at the River Pielach", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, 5W2, Copernicus GmbH, pp. 175–180, DOI: 10.5194/isprsannals-II-5-W2-175-2013.

Mandlburger, G. et al. (2016), "Evaluation of a novel uav-borne topo-bathymetric laser profiler", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 2016-January, International Society for Photogrammetry and Remote Sensing, pp. 933–939, DOI: 10.5194/isprsarchives-XLI-B1-933-2016.

Mandlburger, Gottfried et al. (Feb. 2015), "Topo-Bathymetric LiDAR for Monitoring River Morphodynamics and Instream Habitats—A Case Study at the Pielach River", en, *in*: *Remote Sensing* 7.5, pp. 6160–6195, DOI: 10.3390/rs70506160, URL: https://www.mdpi.com/2072-4292/7/5/6160.

Mandlburger, Gottfried et al. (Mar. 2020), "Concept and performance evaluation of a Novel UAV-Borne Topo-Bathymetric LiDAR sensor", *in*: *Remote Sensing* 12.*6*, ISSN: 20724292, DOI: 10.3390/rs12060986.

Mano, K. et al. (Aug. 2020), "The measurement accuracy and measurement characteristics of green LiDAR drone", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 43, B1, International Society for Photogrammetry and Remote Sensing, pp. 479–483, DOI: 10.5194/isprs-archives-XLIII-B1-2020-479-2020.

Mao, Yongqiang et al. (June 2022a), "Beyond single receptive field: A receptive field fusion-and-stratification network for airborne laser scanning point cloud classification", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 188, pp. 45–61, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2022.03.019.

Mao, Yongqiang et al. (Sept. 2022b), "Bidirectional Feature Globalization for Few-shot Semantic Segmentation of 3D Point Cloud Scenes", *in*: *2022 International Conference on 3D Vision (3DV)*, IEEE, pp. 505–514, ISBN: 978-1-6654-5670-8, DOI: 10.1109/3DV57658.2022.00062, URL: https://ieeexplore.ieee.org/document/10044434/.

Mao, Yongqiang et al. (2022c), "Semantic segmentation for point cloud scenes via dilated graph feature aggregation and pyramid decoders", *in*: *arXiv preprint arXiv:2204.04944*.

Martínez, M L et al. (Mar. 2007), "The coasts of our world: Ecological, economic and social importance", en, *in*: *Ecological Economics*, Ecological Economics of Coastal Disasters 63.*2*, pp. 254–272, ISSN: 0921-

8009, DOI: 10.1016/j.ecolecon.2006.10.022, URL: https://www.sciencedirect.com/science/article/pii/S0921800906005465.

Matikainen, Leena et al. (June 2017), "Object-based analysis of multispectral airborne laser scanner data for land cover classification and map updating", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 128, pp. 298–313, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2017.04.005.

McKean, Jim et al. (Nov. 2009), "Remote Sensing of Channels and Riparian Zones with a Narrow-Beam Aquatic-Terrestrial LIDAR", *in*: *Remote Sensing* 1.4, pp. 1065–1096, ISSN: 2072-4292, DOI: 10.3390/rs1041065, URL: http://www.mdpi.com/2072-4292/1/4/1065.

McKenzie, Len J et al. (July 2020), "The global distribution of seagrass meadows", en, *in*: *Environmental Research Letters* 15.7, p. 074041, ISSN: 1748-9326, DOI: 10.1088/1748-9326/ab7d06, URL: https://iopscience.iop.org/article/10.1088/1748-9326/ab7d06.

Mei, Guofeng et al. (Feb. 2022), "Unsupervised Learning on 3D Point Clouds by Clustering and Contrasting", *in*: URL: http://arxiv.org/abs/2202.02543.

Montes-Hugo, Martin A et al. (Oct. 2016), "Weibull approximation of LiDAR waveforms for estimating the beam attenuation coefficient", *in*: *Opt. Express* 24.20, pp. 22670–22681, DOI: 10.1364/OE.24.022670, URL: https://opg.optica.org/oe/abstract.cfm?URI=oe-24-20-22670.

Morháč, Miroslav et al. (2003), "Efficient algorithm of multidimensional deconvolution and its application to nuclear data processing", *in*: *Digital Signal Processing* 13.1, pp. 144–171, ISSN: 1051-2004, DOI: https://doi.org/10.1016/S1051-2004(02)00011-8, URL: https://www.sciencedirect.com/science/article/pii/S1051200402000118.

Morsy, S. et al. (June 2016), "Airborne Multispectral LiDAR Data for Land-Cover Classification and Land/Water Mapping Using Different Spectral Indexes", *in*: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* III-3, pp. 217–224, ISSN: 2194-9050, DOI: 10.5194/isprsannals-III-3-217-2016, URL: http://www.isprs-ann-photogramm-remote-sens-spatial-inf-sci.net/III-3/217/2016/isprs-annals-III-3-217-2016.pdf.

Morsy, S et al. (2017a), "Clustering of multispectral airborne laser scanning data using Gaussian decomposition", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, 2W7, pp. 269–276, DOI: 10.5194/isprs-archives-XLII-2-W7-269-2017, URL: https://doi.org/10.5194/isprs-archives-XLII-2-W7-269-2017.

— (Apr. 2017b), "Multispectral LiDAR Data for Land Cover Classification of Urban Areas", *in*: *Sensors* 17.5, pp. 958–979, ISSN: 1424-8220, DOI: 10.3390/S17050958, URL: https://www.mdpi.com/1424-8220/17/5/958/htmhttps://www.mdpi.com/1424-8220/17/5/958.

Morsy, Salem et al. (Sept. 2022), "Classification of Multispectral Airborne LiDAR Data Using Geometric and Radiometric Information", *in*: *Geomatics* 2.3, pp. 370–389, ISSN: 2673-7418, DOI: 10.3390/GEOMATICS2030021, URL: https://www.mdpi.com/2673-7418/2/3/21/htmhttps://www.mdpi.com/2673-7418/2/3/21.

Mumby, Pj et al. (Dec. 1997), "Measurement of seagrass standing crop using satellite and digital airborne remote sensing", en, *in*: *Marine Ecology Progress Series* 159, pp. 51–60, ISSN: 0171-8630, 1616-1599, DOI: 10.3354/meps159051, URL: http://www.int-res.com/abstracts/meps/v159/p51-60/https://www.int-res.com/articles/meps/159/m159p051.pdf.

Munoz, Daniel et al. (2008), "Directional Associative Markov Network for 3-D Point Cloud Classification", *in*: *Fourth International Symposium on 3D Data Processing, Visualization and Transmission*.

Mury, Antoine et al. (Nov. 2020), "Wave Attenuation Service by Intertidal Coastal Ecogeosystems in the Bay of Mont-Saint-Michel, France: Review and Meta-Analysis", *in*: *Estuaries and Coastal Zones in Times of Global Change*, pp. 555–572, URL: https://hal.archives-ouvertes.fr/hal-03134051https://hal.archives-ouvertes.fr/hal-03134051/document.

Najafi, Mohammad et al. (2014), "Non-associative higher-order markov networks for point cloud classification", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, PART 5, Springer Verlag, pp. 500–515, DOI: 10.1007/978-3-319-10602-1{_}33.

Narayanan, Ramu et al. (2009), *Classification of SHOALS 3000 bathymetric LiDAR signals using decision tree and ensemble techniques*, p. 462, URL: https://www.researchgate.net/profile/Heungsik-Kim/publication/224129798_Classification_of_SHOALS_3000_bathymetric_LiDAR_signals_using_decision_tree_and_ensemble_techniques/links/54481dd70cf2d62c3052aClassification-of-SHOALS-3000-bathymetric-LiDAR-sign.

Nayegandhi, A. et al. (Feb. 2009), "Small-footprint, waveform-resolving lidar estimation of submerged and sub-canopy topography in coastal environments", *in*: *International Journal of Remote Sensing* 30.4, pp. 861–878, ISSN: 13665901, DOI: 10.1080/01431160802395227.

Nayegandhi, Amar et al. (2006), "Evaluating A Small Footprint, Waveform-resolving Lidar Over Coastal Vegetation Communities", *in*: *Photogrammetric Engineering & Remote Sensing* 72.12, pp. 1407–1417.

Neuenschwander, Amy et al. (2008), "Signal Processing Techniques for Feature Extraction and Classification using Small-Footprint Full-Waveform Airborne LIDAR", *in*: *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, vol. 3, pp. III –676–III –679, DOI: 10.1109/IGARSS.2008.4779438.

Neuenschwander, Amy L (2008), "Evaluation of waveform deconvolution and decomposition retrieval algorithms for ICESat/GLAS data", *in*: *Canadian Journal of Remote Sensing* 34.*sup2*, S240–S246, DOI: 10.5589/m08-044, URL: https://doi.org/10.5589/m08-044.

Neuenschwander, Amy L. (Aug. 2009), "Landcover classification of small-footprint, full-waveform lidar data", *in*: *Journal of Applied Remote Sensing* 3.*1*, p. 033544, ISSN: 1931-3195, DOI: 10.1117/1.3229944, URL: http://remotesensing.spiedigitallibrary.org/article.aspx?doi=10.1117/1.3229944.

Nie, Sheng et al. (July 2017), "Above-ground biomass estimation using airborne discrete-return and full-waveform LiDAR data in a coniferous forest", en, *in*: *Ecological Indicators* 78, pp. 221–228, ISSN: 1470160X, DOI: 10.1016/j.ecolind.2017.02.045, URL: https://linkinghub.elsevier.com/retrieve/pii/S1470160X17301085.

Niemeyer, J. et al. (July 2012), "Conditional Random Fields For Lidar Point Cloud Classification In Complex Urban Areas", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* I-3, pp. 263–268, ISSN: 2194-9050, DOI: 10.5194/isprsannals-I-3-263-2012, URL: https://isprs-annals.copernicus.org/articles/I-3/263/2012/.

Niemeyer, Joachim et al. (2011), "Conditional random fields for urban scene classification with full waveform LiDAR data", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6952 LNCS, pp. 233–244, ISSN: 03029743, DOI: 10.1007/978-3-642-24393-6{_}20/COVER, URL: https://link.springer.com/chapter/10.1007/978-3-642-24393-6_20.

Noordermeer, Lennart et al. (Mar. 2023), "Monitoring tree occupancy and height in the Norwegian alpine treeline using a time series of airborne laser scanner data", *in*: *International Journal of Applied Earth Observation and Geoinformation* 117, p. 103201, ISSN: 1569-8432, DOI: 10.1016/J.JAG.2023.103201.

Osada, Robert et al. (2002), "Shape Distributions", *in*: *ACM Transactions on Graphics* 21.*4*, pp. 807–832.

Oshiro, Thais Mayumi et al. (2012), "How Many Trees in a Random Forest?", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 7376 LNAI, pp. 154–168, DOI: 10.1007/978-3-642-31537-4{_}13, URL: http://link.springer.com/10.1007/978-3-642-31537-4_13.

Pal, M. (Jan. 2005), "Random forest classifier for remote sensing classification", *in*: *International Journal of Remote Sensing* 26.*1*, pp. 217–222, ISSN: 0143-1161, DOI: 10.1080/01431160412331269698, URL: https://www.tandfonline.com/doi/full/10.1080/01431160412331269698.

Pan, Zhigang et al. (Feb. 2015), "Performance Assessment of High Resolution Airborne Full Waveform LiDAR for Shallow River Bathymetry", en, *in*: *Remote Sensing* 7.5, pp. 5133–5159, DOI: 10.3390/rs70505133, URL: https://www.mdpi.com/2072-4292/7/5/5133.

Pan, Zhigang et al. (2016), "Fusion of LiDAR Orthowaveforms and Hyperspectral Imagery for Shallow River Bathymetry and Turbidity Estimation", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 54.7, pp. 4165–4177, DOI: 10.1109/TGRS.2016.2538089.

Park, Chunghyun et al. (2022), "Fast Point Transformer", *in*: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16949–16958, URL: http://cvlab.postech.ac.kr/research/FPT.

Parrish, Christopher E. et al. (2011), "Empirical comparison of full-waveform lidar algorithms: Range extraction and discrimination performance", *in*: *Photogrammetric Engineering and Remote Sensing* 77.8, pp. 825–838, ISSN: 00991112, DOI: 10.14358/PERS.77.8.825.

Parrish, Christopher E. et al. (Dec. 2016), "Post-Sandy Benthic Habitat Mapping Using New Topobathymetric Lidar Technology and Object-Based Image Classification", en, *in*: *Journal of Coastal Research* 76, pp. 200–208, ISSN: 0749-0208, DOI: 10.2112/SI76-017, URL: http://www.bioone.org/doi/10.2112/SI76-017.

Pasqualini, V et al. (Dec. 1998), "Mapping of Posidonia oceanicausing Aerial Photographs and Side Scan Sonar: Application off the Island of Corsica (France)", en, *in*: *Estuarine, Coastal and Shelf Science* 47.3, pp. 359–367, ISSN: 0272-7714, DOI: 10.1006/ecss.1998.0361, URL: http://www.sciencedirect.com/science/article/pii/S0272771498903614.

Pastol, Yves (Mar. 2011), "Use of Airborne LIDAR Bathymetry for Coastal Hydrographic Surveying: The French Experience", *in*: *Journal of Coastal Research* 62.62, pp. 6–18, ISSN: 0749-0208, DOI: 10.2112/SI{_}62{_}2, URL: http://www.bioone.org/doi/abs/10.2112/SI_62_2.

Pauly, Mark (2003), *Point primitives for interactive modeling and processing of 3D geometry*, Hartung-Gorre.

Pedregosa, Fabian et al. (2011), "Scikit-learn: Machine learning in Python", *in*: *Journal of machine learning research* 12.Oct, pp. 2825–2830.

Pe'eri, Shachak et al. (May 2007), "Increasing the existence of very shallow-water LIDAR measurements using the red-channel waveforms", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 45.5, pp. 1217–1223, ISSN: 01962892, DOI: 10.1109/TGRS.2007.894584.

Pelletier, Charlotte et al. (Mar. 2019), "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series", *in*: *Remote Sensing 2019, Vol. 11, Page 523* 11.5, p. 523, ISSN: 2072-4292, DOI: 10.3390/RS11050523, URL: https://www.mdpi.com/2072-4292/11/5/523/htmhttps://www.mdpi.com/2072-4292/11/5/523.

Perez, Luis et al. (2017), "The effectiveness of data augmentation in image classification using deep learning", *in*: *arXiv preprint arXiv:1712.04621*.

Perslev, Mathias et al. (2019), "U-Time: A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging", *in*: *Advances in Neural Information Processing Systems* 32.

Petzold, Theodore J (1972), "Volume scattering functions for selected ocean waters", *in*.

Philpot, William (Nov. 2019), *Airborne Laser Hydrography II*, en, DOI: https://doi.org/10.7298/jxm9-g971, URL: https://ecommons.cornell.edu/handle/1813/66666https://ecommons.cornell.edu/bitstream/1813/66666/2/BBII_AirborneLaserHydrography.pdf.

Pi-Fuei Hsieh et al. (2001), "Effect of spatial resolution on classification errors of pure and mixed pixels in remote sensing", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 39.12, pp. 2657–2663, ISSN: 01962892, DOI: 10.1109/36.975000, URL: http://ieeexplore.ieee.org/document/975000/.

Pirotti et al. (2011), "Analysis of full-waveform LiDAR data for forestry applications: a review of investigations and methods", *in*: *iForest - Biogeosciences and Forestry* 4.3, p. 100, ISSN: 1971-7458, DOI: 10.3832/IFOR0562-004, URL: https://iforest.sisef.org/abstract/?id=ifor0562-004.

Popescu, Sorin C. et al. (Mar. 2008), "A voxel-based lidar method for estimating crown base height for deciduous and pine trees", *in*: *Remote Sensing of Environment* 112.3, pp. 767–781, ISSN: 00344257, DOI: 10.1016/j.rse.2007.06.011.

Puech, Christian et al. (2014), "Airborne lidar for natural environments: research and applications in France", *in*: *Revue Française de Photogrammétrie et de Télédétection* 200, pp. 54–58, DOI: https://doi.org/10.52638/rfpt.2012.62, URL: http://LiDAR.teledetection.fr/,.

Qi, Charles R. et al. (2017a), "PointNet++: Deep hierarchical feature learning on point sets in a metric space", *in*: *Advances in Neural Information Processing Systems*, vol. 2017-Decem, pp. 5100–5109.

Qi, Charles R et al. (2017b), "PointNet: Deep learning on point sets for 3D classification and segmentation", *in*: *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-Janua, pp. 77–85, ISBN: 9781538604571, DOI: 10.1109/CVPR.2017.16.

Qin, Yuchu et al. (2012), "Toward an optimal algorithm for LiDAR waveform decomposition", *in*: *IEEE Geoscience and Remote Sensing Letters* 9.3, pp. 482–486, ISSN: 1545598X, DOI: 10.1109/LGRS.2011.2172676.

Quadros, N D et al. (2008), "Integration of bathymetric and topographic LiDAR: A preliminary investigation", en, *in*: *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci* 37, pp. 1299–1304, URL: https://www.isprs.org/proceedings/xxxvii/congress/8_pdf/13_ths-19/01.pdf.

Redmon, Joseph et al. (2016), "You Only Look Once: Unified, Real-Time Object Detection", *in*: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, URL: http://pjreddie.com/yolo/.

Redmon, Joseph et al. (2017), *YOLO9000: Better, Faster, Stronger*, URL: http://pjreddie.com/yolo9000/.

— (Apr. 2018), "YOLOv3: An Incremental Improvement", *in*: URL: http://arxiv.org/abs/1804.02767.

Rees, W. G. (2001), *Physical Principles of Remote Sensing*.

Reitberger, J. et al. (Nov. 2009), "3D segmentation of single trees exploiting full waveform LIDAR data", en, *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 64.6, pp. 561–574, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2009.04.002, URL: https://linkinghub.elsevier.com/retrieve/pii/S0924271609000495.

Ribeiro, Marco Tulio et al. (2016), ""Why Should I Trust You?": Explaining the Predictions of Any Classifier", *in*: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pp. 1135–1144.

Richardson, Jeffrey J et al. (Oct. 2011), "Strengths and limitations of assessing forest density and spatial configuration with aerial LiDAR", en, *in*: *Remote Sensing of Environment* 115.10, pp. 2640–2651, ISSN: 00344257, DOI: 10.1016/j.rse.2011.05.020, URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425711002033.

Richter, K. et al. (June 2021), "Water turbidity estimation from lidar bathymetry data by full-waveform analysis - Comparison of two approaches", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 43, B2-2021, International Society for Photogrammetry and Remote Sensing, pp. 681–688, DOI: 10.5194/isprs-archives-XLIII-B2-2021-681-2021.

— (May 2022), "Determination of 3D Water Turbidity Parameter Fields from Lidar Bathymetry Data by Volumetric Data Analysis", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 43, B2-2022, International Society for Photogrammetry and Remote Sensing, pp. 945–951, DOI: 10.5194/isprs-archives-XLIII-B2-2022-945-2022.

Richter, Katja et al. (2017), "An Approach to Determining Turbidity and Correcting for Signal Attenuation in Airborne Lidar Bathymetry", *in*: *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 85.1, pp. 31–40, ISSN: 2363-7145, DOI: 10.1007/s41064-016-0001-0, URL: https://doi.org/10.1007/s41064-016-0001-0.

Robert, Damien et al. (June 2023), "Efficient 3D Semantic Segmentation with Superpoint Transformer", *in*: *arXiv preprint arXiv:2306.08045*, URL: http://arxiv.org/abs/2306.08045.

Ronneberger, Olaf et al. (2015), "U-net: Convolutional networks for biomedical image segmentation", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9351, pp. 234–241, ISSN: 16113349, DOI: 10.1007/978-3-319-24574-4{_}28/COVER, URL: https://link.springer.com/chapter/10.1007/978-3-319-24574-4_28.

Roscher, Ribana et al. (2020), "Explainable Machine Learning for Scientific Insights and Discoveries", *in*: *IEEE Access* 8, pp. 42200–42216, ISSN: 21693536, DOI: 10.1109/ACCESS.2020.2976199.

Rumelhart, D. E. et al. (1986), "Learning representations by back-propagating errors", *in*: *Nature* 323.*6088*, pp. 533–536.

Rusu, Radu Bogdan et al. (2009), "Fast Point Feature Histograms (FPFH) for 3D Registration", *in*: *2009 IEEE International Conference on Robotics and Automation*, [IEEE], pp. 3212–3217, ISBN: 9781424427895.

Sagi, Omer et al. (July 2018), "Ensemble learning: A survey", *in*: *WIREs Data Mining and Knowledge Discovery* 8.*4*, pp. 1–18, ISSN: 1942-4787, DOI: 10.1002/widm.1249, URL: https://onlinelibrary.wiley.com/doi/10.1002/widm.1249.

Sandidge, Juanita C. et al. (Sept. 1998), "Coastal Bathymetry from Hyperspectral Observations of Water Radiance", en, *in*: *Remote Sensing of Environment* 65.*3*, pp. 341–352, ISSN: 00344257, DOI: 10.1016/S0034-4257(98)00043-1, URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425798000431.

Saylam, Kutalmis et al. (June 2017), "Assessment of depth and turbidity with airborne Lidar bathymetry and multiband satellite imagery in shallow water bodies of the Alaskan North Slope", *in*: *International Journal of Applied Earth Observation and Geoinformation* 58, pp. 191–200, ISSN: 1872826X, DOI: 10.1016/j.jag.2017.02.012.

Saylam, Kutalmis et al. (Aug. 2020), "Multi-sensor approach to improve bathymetric lidar mapping of semi-arid groundwater-dependent streams: Devils River, Texas", *in*: *Remote Sensing* 12.*15*, ISSN: 20724292, DOI: 10.3390/RS12152491.

Scheeres, Janneke et al. (May 2023), "Distinguishing forest types in restored tropical landscapes with UAV-borne LIDAR", *in*: *Remote Sensing of Environment* 290, p. 113533, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2023.113533.

Schmidt, Alena et al. (2014), "Contextual Classification of Full Waveform Lidar Data in the Wadden Sea", *in*: *IEEE Geoscience and Remote Sensing Letters* 11.*9*, pp. 1614–1618, ISSN: 1558-0571, DOI: 10.1109/LGRS.2014.2302317, URL: https://ieeexplore.ieee.org/abstract/document/6739122.

Schmohl, S. et al. (May 2019), "Submanifold Sparse Convolutional Networks For Semantic Segmentation Of Large-Scale ALS Point Clouds", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 4, 2/W5, Copernicus GmbH, pp. 77–84, DOI: 10.5194/isprs-annals-IV-2-W5-77-2019.

Schwarz, Roland et al. (Aug. 2017), "Exponential decomposition with implicit deconvolution of lidar backscatter from thewater column", *in*: *Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 85.3, pp. 159–167, ISSN: 25122819, DOI: 10.1007/s41064-017-0018-z.

Schwarz, Roland et al. (Apr. 2019), "Design and evaluation of a full-wave surface and bottom-detection algorithm for LiDAR bathymetry of very shallow waters", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 150, pp. 1–10, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2019.02.002.

Selvaraju, Ramprasaath R et al. (2020), "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization", *in*: *International Journal of Computer Vision* 128, pp. 336–359, DOI: 10.1007/s11263-019-01228-7, URL: http://gradcam.cloudcv.org.

Settles, Burr (2009), "Active Learning Literature Survey", *in*: URL: https://minds.wisconsin.edu/handle/1793/60660.

Shaker, Ahmed et al. (Dec. 2019), "Automatic land-water classification using multispectral airborne LiDAR data for near-shore and river environments", en, *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 152, pp. 94–108, ISSN: 0924-2716, DOI: 10.1016/j.isprsjprs.2019.04.005, URL: https://www.sciencedirect.com/science/article/pii/S0924271619301017.

Shapley, Lloyd S. (Apr. 1953), "A Value for N-Person Games", *in*: *Contributions to the Theory of Games (AM-28), Volume II*, ed. by H W Kuhn et al., Princeton University Press, DOI: 10.7249/P0295, URL: https://www.rand.org/pubs/papers/P295.html.

Shen, Xiang et al. (June 2017), "Decomposition of LiDAR waveforms by B-spline-based modeling", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 128, pp. 182–191, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2017.03.006.

Shi, Yongjie et al. (2022), "Deep Unsupervised Domain Adaptation with Time Series Sensor Data: A Survey", *in*: *Sensors* 22.15, ISSN: 1424-8220, DOI: 10.3390/s22155507, URL: https://www.mdpi.com/1424-8220/22/15/5507.

Shinohara, Takayuki et al. (Aug. 2020), "FWNet: Semantic Segmentation for Full-Waveform LiDAR Data Using Deep Learning", en, *in*: *Sensors* 20.12, p. 3568, DOI: 10.3390/s20123568, URL: https://www.mdpi.com/1424-8220/20/12/3568https://www.mdpi.com/1424-8220/20/12/3568/pdfhttps://www.mdpi.com/1424-8220/20/12/3568/htm.

Smeeckaert, Julien et al. (Nov. 2013), "Large-scale classification of water areas using airborne topographic lidar data", *in*: *Remote Sensing of Environment* 138, pp. 134–148, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2013.07.004.

Sokolova, Marina et al. (2006), "Beyond Accuracy, F-score, and ROC: A Family of Discriminant Measures for Performance Evaluation", *in*: *Australasian joint conference on artificial intelligence*, pp. 1015–1021, URL: www.aaai.org.

Song, J.-H. et al. (2002), "Assessing the possibility of land-cover classification using lidar intensity data", *in*: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 34, pp. 259–262, ISSN: 16821750.

Song, Shalei et al. (Mar. 2019), "A new waveform decomposition method for multispectral LiDAR", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 149, pp. 40–49, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2019.01.014.

Stumpf, Richard P. et al. (Jan. 2003), "Determination of water depth with high-resolution satellite imagery over variable bottom types", en, *in*: *Limnology and Oceanography* 48.*1part2*, pp. 547–556, ISSN: 00243590, DOI: 10.4319/lo.2003.48.1{_}part{_}2.0547, URL: http://doi.wiley.com/10.4319/lo.2003.48.1_part_2.0547.

Sun, Yung Da et al. (2017), "A hybrid seabed classification method using airborne laser bathymetric data", *in*: *Journal of Marine Science and Technology (Taiwan)* 25.*3*, pp. 358–364, ISSN: 10232796, DOI: 10.6119/JMST-016-1230-1.

Sundararajan, Mukund et al. (Mar. 2017), "Axiomatic Attribution for Deep Networks", *in*: URL: http://arxiv.org/abs/1703.01365.

Syvitski, James P M et al. (Apr. 2005), "Dynamics of the Coastal Zone", en, *in*: *Coastal Fluxes in the Anthropocene: The Land-Ocean Interactions in the Coastal Zone Project of the International Geosphere-Biosphere Programme*, ed. by Christopher J Crossland et al., Global Change — The IGBP Series, Berlin, Heidelberg: Springer, pp. 39–94, ISBN: 978-3-540-27851-1, URL: https://doi.org/10.1007/3-540-27851-6_2http://link.springer.com/content/pdf/10.1007%2F3-540-27851-6_2.pdf.

Szegedy, Christian et al. (2015), *Going Deeper With Convolutions*.

Tchapmi, Lyne et al. (May 2018), "SEGCloud: Semantic segmentation of 3D point clouds", *in*: *Proceedings - 2017 International Conference on 3D Vision, 3DV 2017*, Institute of Electrical and Electronics Engineers Inc., pp. 537–547, ISBN: 9781538626108, DOI: 10.1109/3DV.2017.00067.

Teo, Tee Ann et al. (June 2017), "Analysis of Land Cover Classification Using Multi-Wavelength LiDAR System", *in*: *Applied Sciences* 7.*7*, p. 663, ISSN: 2076-3417, DOI: 10.3390/APP7070663, URL:

https://www.mdpi.com/2076-3417/7/7/663/htmhttps://www.mdpi.com/2076-3417/7/7/663.

Thomas, Hugues et al. (2018), "Semantic classification of 3d point clouds with multiscale spherical neighborhoods", *in*: *Proceedings - 2018 International Conference on 3D Vision, 3DV 2018*, pp. 390–398, ISBN: 9781538684252, DOI: 10.1109/3DV.2018.00052.

Thomas, Hugues et al. (2019), "KPConv: Flexible and deformable convolution for point clouds", *in*: *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2019-Octob, pp. 6410–6419, ISBN: 9781728148038, DOI: 10.1109/ICCV.2019.00651.

Tombari, Federico et al. (2010), "Unique Signatures of Histograms for Local Surface Description", *in*: *ECCV 2010*, pp. 356–369, URL: http://www.vision.deis.unibo.it.

Topouzelis, Konstantinos et al. (Dec. 2018), "Seagrass mapping in Greek territorial waters using Landsat-8 satellite images", en, *in*: *International Journal of Applied Earth Observation and Geoinformation* 67, pp. 98–113, ISSN: 0303-2434, DOI: 10.1016/j.jag.2017.12.013, URL: http://www.sciencedirect.com/science/article/pii/S0303243417303264.

Triebel, Rudolph et al. (2006), "Robust 3D Scan Point Classification using Associative Markov Networks", *in*: *Proceedings 2006 IEEE International Conference on Robotics and Automation*, pp. 2603–2608.

Tuell, Grady H et al. (2005), "Measurement of ocean water optical properties and seafloor reflectance with scanning hydrographic operational airborne lidar survey (SHOALS): II. Practical results and comparison with independent data", *in*: *Remote Sensing of the Coastal Oceanic Environment*, ed. by Robert J Frouin et al., vol. 5885, SPIE, 58850E, DOI: 10.1117/12.619215, URL: https://doi.org/10.1117/12.619215.

Tulldahl, H Michael et al. (1999), "Analytical waveform generation from small objects in lidar bathymetry", *in*: *Applied Optics* 38.6, pp. 1021–1039.

Tulldahl, H Michael et al. (Dec. 2012), "Classification of aquatic macrovegetation and substrates with airborne lidar", en, *in*: *Remote Sensing of Environment* 121, pp. 347–357, ISSN: 0034-4257, DOI: 10.1016/j.rse.2012.02.004, URL: https://www.sciencedirect.com/science/article/pii/S0034425712000880.

Turner, R K et al. (Nov. 2015), "Conceptual Framework", en, *in*: *Coastal Zones Ecosystem Services: From Science to Values and Decision Making*, ed. by R Kerry Turner et al., Studies in Ecological Economics, Cham: Springer International Publishing, pp. 11–40, ISBN: 978-3-319-17214-9, URL: https://doi.org/10.1007/978-3-319-17214-9_2.

Vallet, Julien (2011), "La lasergrammétrie aéroportée, approches théoriques, concepts, méthodes", *in*: *Collection EDYTEM. Cahiers de géographie* 12.*1*, pp. 29–40, ISSN: 1762-4304, DOI: 10.3406/edyte.2011.1175.

Vandapel, N. et al. (2004), "Natural terrain classification using 3-d ladar data", *in*: *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*, vol. 2004, 5, IEEE, pp. 5117–5122, ISBN: 0-7803-8232-3, DOI: 10.1109/ROBOT.2004.1302529, URL: http://ieeexplore.ieee.org/document/1302529/.

Vaswani, Ashish et al. (2017), "Attention is All you Need", *in*: *Advances in Neural Information Processing Systems* 30.

Vosselman, G et al., eds. (2010), *Airborne and terrestrial laser scanning*, CRC Press (Taylor & Francis), pp. 83–91, ISBN: 978-1904445-87-6, URL: https://research.utwente.nl/en/publications/airborne-and-terrestrial-laser-scanning-2.

Vosselman, George et al. (June 2017), "Contextual segment-based classification of airborne laser scanner data", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 128, pp. 354–371, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2017.03.010.

Wagner, Wolfgang (Nov. 2010), "Radiometric calibration of small-footprint full-waveform airborne laser scanner measurements: Basic physical concepts", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 65.*6*, pp. 505–513, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2010.06.007.

Wagner, Wolfgang et al. (Apr. 2006), "Gaussian decomposition and calibration of a novel small-footprint full-waveform digitising airborne laser scanner", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 60.*2*, pp. 100–112, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2005.12.001.

Walker, Ronald E. et al. (Apr. 1999), "Lidar equations for turbid media with pulse stretching", *in*: *Applied Optics* 38.*12*, p. 2384, ISSN: 0003-6935, DOI: 10.1364/ao.38.002384.

Wang, Chi-Kuei et al. (Dec. 2007), "Using airborne bathymetric lidar to detect bottom type variation in shallow waters", en, *in*: *Remote Sensing of Environment* 106.*1*, pp. 123–135, ISSN: 0034-4257, DOI: 10.1016/j.rse.2006.08.003, URL: http://www.sciencedirect.com/science/article/pii/S0034425706003002.

Wang, Chisheng et al. (Mar. 2015), "A comparison of waveform processing algorithms for single-wavelength LiDAR bathymetry", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 101, pp. 22–35, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2014.11.005.

Wang, Lei et al. (2019a), "Graph Attention Convolution for Point Cloud Semantic Segmentation", *in*: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10296–10305.

Wang, Qingwang et al. (Mar. 2020a), "A Discriminative Tensor Representation Model for Feature Extraction and Classification of Multispectral LiDAR Data", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 58.3, pp. 1568–1586, ISSN: 15580644, DOI: 10.1109/TGRS.2019.2947081.

Wang, Wei et al. (Jan. 2019b), "A Survey of Zero-Shot Learning", *in*: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2, ISSN: 21576912, DOI: 10.1145/3293318, URL: https://dl.acm.org/doi/10.1145/3293318.

Wang, Yaqing et al. (June 2020b), "Generalizing from a Few Examples", *in*: *ACM Computing Surveys (CSUR)* 53.3, ISSN: 15577341, DOI: 10.1145/3386252, URL: https://dl.acm.org/doi/10.1145/3386252.

Wang, Yue et al. (Oct. 2019c), "Dynamic graph Cnn for learning on point clouds", *in*: *ACM Transactions on Graphics* 38.5, pp. 146–158, ISSN: 15577368, DOI: 10.1145/3326362.

Watkins, Ray H et al. (2023), "Validation of ICESat-2 Derived Data Products on Freshwater Lakes: Bathymetry, Diffuse Attenuation Coefficient for Downwelling Irradiance (Kd), and Particulate Backscatter Coefficient (bbp)", *in*: *IEEE Geoscience and Remote Sensing Letters* 20, pp. 1–5, DOI: 10.1109/LGRS.2023.3261551.

Webster, Tim et al. (Feb. 2020), "Calculating macroalgal height and biomass using bathymetric LiDAR and a comparison with surface area derived from satellite data in Nova Scotia, Canada", *in*: *Botanica Marina* 63.1, pp. 43–59, ISSN: 1437-4323, DOI: 10.1515/bot-2018-0080, URL: https://www.degruyter.com/document/doi/10.1515/bot-2018-0080/html.

Wedding, Lisa M. et al. (Nov. 2008), "Using bathymetric lidar to define nearshore benthic habitat complexity: Implications for management of reef fish assemblages in Hawaii", *in*: *Remote Sensing of Environment* 112.11, pp. 4159–4165, ISSN: 00344257, DOI: 10.1016/j.rse.2008.01.025, URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425708002137.

Wei, Mingqiang et al. (Jan. 2023), "AGConv: Adaptive Graph Convolution on 3D Point Clouds", *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–18, ISSN: 0162-8828, DOI: 10.1109/tpami.2023.3238516.

Weinmann, M. et al. (Oct. 2013), "Feature relevance assessment for the semantic interpretation of 3D point cloud data", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-5/W2, pp. 313–318, ISSN: 2194-9050, DOI: 10.5194/isprsannals-II-5-W2-313-2013, URL: https://isprs-annals.copernicus.org/articles/II-5-W2/313/2013/.

Weinmann, M et al. (July 2015), "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 105, pp. 286–304, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2015.01.016.

Weiss, Karl et al. (Dec. 2016), "A survey of transfer learning", *in*: *Journal of Big Data* 3.1, pp. 1–40, ISSN: 21961115, DOI: 10.1186/S40537-016-0043-6/TABLES/6, URL: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-016-0043-6.

Wen, Congcong et al. (Mar. 2021), "Airborne LiDAR point cloud classification with global-local graph attention convolution neural network", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 173, pp. 181–194, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2021.01.007.

Wen, Qingsong et al. (Feb. 2022), "Transformers in Time Series: A Survey", *in*: URL: http://arxiv.org/abs/2202.07125.

West, Karen F. et al. (Sept. 2004), "Context-driven automated target detection in 3D data", *in*: *Automatic Target Recognition XIV* 5426, pp. 133–143, DOI: 10.1117/12.542536, URL: https://www.spiedigitallibrary.org/conference-proceedings-of-spie/5426/0000/Context-driven-automated-target-detection-in-3D-data/10.1117/12.542536.fullhttps://www.spiedigitallibrary.org/conference-proceedings-of-spie/5426/0000/Context-driven-automated-target-detection-in-3D-data/10.1117/12.542536.short.

Wichmann et al. (Aug. 2015), "Evaluating The Potential Of Multispectral Airborne Lidar For Topographic Mapping And Land Cover Classification", *in*: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5*3 W5*, pp. 113–119, ISSN: 2194-9050, DOI: 10.5194/isprsannals-II-3-W5-113-2015, URL: https://isprs-annals.copernicus.org/articles/II-3-W5/113/2015/.

Wilson, Nicholas et al. (Nov. 2019), "Mapping Seafloor Relative Reflectance and Assessing Coral Reef Morphology with EAARL-B Topobathymetric Lidar Waveforms", en, *in*: *Estuaries and Coasts* 45, pp. 923–937, ISSN: 1559-2731, DOI: 10.1007/s12237-019-00652-9, URL: https://doi.org/10.1007/s12237-019-00652-9https://link.springer.com/content/pdf/10.1007%2Fs12237-019-00652-9.pdf.

Wohlkinger, Walter et al. (2011), "Ensemble of Shape Functions for 3D Object Classification", *in*: *2011 IEEE international conference on robotics and biomimetics*, pp. 2987–2992.

Wozencraft, Jennifer et al. (2005), "Airborne Lidar and Integrated Technologies for Coastal Mapping and Nautical Charting", *in*: *Marine Technology Society Journal* 39.3, pp. 27–35.

Wu, Jiaying et al. (June 2011), "A comparison of signal deconvolution algorithms based on small-footprint LiDAR waveform simulation", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 49.6 *PART 2*, pp. 2402–2414, ISSN: 01962892, DOI: 10.1109/TGRS.2010.2103080.

Wu, Zonghan et al. (Jan. 2021), "A Comprehensive Survey on Graph Neural Networks", *in*: *IEEE Transactions on Neural Networks and Learning Systems* 32.1, pp. 4–24, ISSN: 2162-237X, DOI: 10.1109/TNNLS.2020.2978386, URL: https://ieeexplore.ieee.org/document/9046288/.

Xing, Shuai et al. (Dec. 2019), "A depth-adaptive waveform decomposition method for airborne LiDAR bathymetry", *in*: *Sensors* 19.23, ISSN: 14248220, DOI: 10.3390/s19235065.

Xu, Xun et al. (2020), "Weakly Supervised Semantic Point Cloud Segmentation: Towards 10x Fewer Labels", *in*: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13706–13715, URL: https://github.com/alex-xun-xu/WeakSupPointCloudSeg.

Xu, Yifan et al. (2018), "SpiderCNN: Deep learning on point sets with parameterized convolutional filters", *in*: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11212 LNCS, pp. 90–105, ISBN: 9783030012366, DOI: 10.1007/978-3-030-01237-3{_}6.

Yan, Wai Yeung et al. (Jan. 2012), "Improving classification accuracy of airborne LiDAR intensity data by geometric calibration and radiometric correction", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 67.1, pp. 35–44, ISSN: 0924-2716, DOI: 10.1016/J.ISPRSJPRS.2011.10.005.

Yan, Wai Yeung et al. (Mar. 2015), "Urban land cover classification using airborne LiDAR data: A review", *in*: *Remote Sensing of Environment* 158, pp. 295–310, ISSN: 0034-4257, DOI: 10.1016/J.RSE.2014.11.001.

Yang, Fanlin et al. (May 2022), "An airborne LiDAR bathymetric waveform decomposition method in very shallow water: a case study around Yuanzhi Island in the South China Sea", *in*: *Int. J. Appl. Earth Obs. Geoinf.* 109, p. 102788, ISSN: 15698432, DOI: 10.1016/j.jag.2022.102788.

Yang, Zhishuang et al. (Oct. 2018), "Segmentation and Multi-Scale Convolutional Neural Network-Based Classification of Airborne Laser Scanner Data", *in*: *Sensors* 18.10, pp. 3347–3364, ISSN: 1424-8220, DOI: 10.3390/s18103347, URL: http://www.mdpi.com/1424-8220/18/10/3347.

Ye, Xiaoqing et al. (2018), "3D Recurrent Neural Networks with Context Fusion for Point Cloud Semantic Segmentation", *in*: *Proceedings of the European conference on computer vision (ECCV)*, pp. 403–417.

Yuan, Zhimin et al. (2023), "Prototype-Guided Multitask Adversarial Network for Cross-Domain LiDAR Point Clouds Semantic Segmentation", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 61, pp. 1–13, ISSN: 0196-2892, DOI: 10.1109/TGRS.2023.3234542, URL: https://ieeexplore.ieee.org/document/10007866/.

Yun, Sangdoo et al. (2019), "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features", *in*: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6023–6032, URL: https://github.com/clovaai/CutMix-PyTorch..

Zavalas, Richard et al. (Dec. 2014), "Habitat Classification of Temperate Marine Macroalgal Communities Using Bathymetric LiDAR", en, *in*: *Remote Sensing* 6.3, pp. 2154–2175, DOI: 10.3390/rs6032154, URL: https://www.mdpi.com/2072-4292/6/3/2154https://www.mdpi.com/2072-4292/6/3/2154/pdf.

Zeng, Tao et al. (2023), "Recurrent Residual Dual Attention Network for Airborne Laser Scanning Point Cloud Semantic Segmentation", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 61, pp. 1–14, ISSN: 0196-2892, DOI: 10.1109/TGRS.2023.3285207, URL: https://ieeexplore.ieee.org/document/10149028/.

Zhang, Hongyi et al. (Oct. 2017), "mixup: Beyond Empirical Risk Minimization", *in*: *arXiv preprint arXiv:1710.09412*, URL: http://arxiv.org/abs/1710.09412.

Zhang, Jixian et al. (July 2013), "SVM-Based Classification of Segmented Airborne LiDAR Point Clouds in Urban Areas", *in*: *Remote Sensing 2013, Vol. 5, Pages 3749-3775* 5.8, pp. 3749–3775, ISSN: 2072-4292, DOI: 10.3390/RS5083749, URL: https://www.mdpi.com/2072-4292/5/8/3749/htmhttps://www.mdpi.com/2072-4292/5/8/3749.

Zhang, Ka et al. (2022), "A Dual Attention Neural Network for Airborne LiDAR Point Cloud Semantic Segmentation", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 60, pp. 1–17, ISSN: 15580644, DOI: 10.1109/TGRS.2022.3201902.

Zhang, Zhenhua et al. (2023), "Retrieving bbp and POC from CALIOP: A deep neural network approach", *in*: *Remote Sensing of Environment* 287, p. 113482, ISSN: 0034-4257, DOI: https://doi.org/10.1016/j.rse.2023.113482, URL: https://www.sciencedirect.com/science/article/pii/S0034425723000330.

Zhao, Hengshuang et al. (2021a), "Point Transformer", *in*: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 16239–16248, ISBN: 9781665428125, DOI: 10.1109/ICCV48922.2021.01595.

Zhao, Jianhu et al. (May 2017), "Shallow water measurements using a single green laser corrected by building a near water surface penetration model", *in*: *Remote Sensing* 9.5, ISSN: 20724292, DOI: 10.3390/rs9050426.

Zhao, Na et al. (2021b), "Few-shot 3D Point Cloud Semantic Segmentation", *in*: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, IEEE Computer Society, pp. 8869–8878, ISBN: 9781665445092, DOI: 10.1109/CVPR46437.2021.00876.

Zhao, Ruibin et al. (May 2018a), "Classifying airborne LiDAR point clouds via deep features learned by a multi-scale convolutional neural network", *in*: *International Journal of Geographical Information Science* 32.5, pp. 960–979, ISSN: 1365-8816, DOI: 10.1080/13658816.2018.1431840, URL: https://www.tandfonline.com/doi/full/10.1080/13658816.2018.1431840.

Zhao, Xinglei et al. (Feb. 2018b), "Remote sensing of suspended sediment concentrations based on the waveform decomposition of Airborne LiDAR bathymetry", *in*: *Remote Sensing* 10.2, ISSN: 20724292, DOI: 10.3390/rs10020247.

Zhao, Xinglei et al. (2022), "Adaptive Wavelet Threshold Denoising for Bathymetric Laser Full-Waveforms with Weak Bottom Returns", *in*: *IEEE Geoscience and Remote Sensing Letters* 19, ISSN: 15580571, DOI: 10.1109/LGRS.2022.3141057.

Zhao, Zhong Qiu et al. (Nov. 2019), *Object Detection with Deep Learning: A Review*, DOI: 10.1109/TNNLS.2018.2876865.

Zhou, Guoqing et al. (2021), "Comparison Analysis of Five Waveform Decomposition Algorithms for the Airborne LiDAR Echo Signal", *in*: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp. 7869–7880, ISSN: 21511535, DOI: 10.1109/JSTARS.2021.3096197.

Zhou, Jie et al. (Jan. 2020), "Graph neural networks: A review of methods and applications", *in*: *AI Open* 1, pp. 57–81, ISSN: 2666-6510, DOI: 10.1016/J.AIOPEN.2021.01.001.

Zhou, Qian-Yi et al. (Jan. 2018), "Open3D: A Modern Library for 3D Data Processing", *in*: *ArXiv preprint*, URL: http://arxiv.org/abs/1801.09847.

Zhou, Tan et al. (July 2017), "Gold – A novel deconvolution algorithm with optimization for waveform LiDAR processing", *in*: *ISPRS Journal of Photogrammetry and Remote Sensing* 129, pp. 131–150, ISSN: 09242716, DOI: 10.1016/j.isprsjprs.2017.04.021.

Zhou, Tan et al. (Oct. 2019), "waveformlidar: An R Package for Waveform LiDAR Processing and Analysis", *in*: *Remote Sensing 2019, Vol. 11, Page 2552* 11.21, p. 2552, ISSN: 2072-4292, DOI: 10.3390/RS11212552, URL: https://www.mdpi.com/2072-4292/11/21/2552/htmhttps://www.mdpi.com/2072-4292/11/21/2552.

Zhou, Zhi Hua (Jan. 2018), "A brief introduction to weakly supervised learning", *in*: *National Science Review* 5.1, pp. 44–53, ISSN: 2095-5138, DOI: 10.1093/NSR/NWX106, URL: https://dx.doi.org/10.1093/nsr/nwx106.

Zhu, J et al. (Feb. 2012), "Analysis and application of LiDAR waveform data using a progressive waveform decomposition method", en, *in*: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII-5/, pp. 31–36, ISSN: 2194-9034, DOI: 10.5194/isprsarchives-XXXVIII-5-W12-31-2011, URL: https://www.int-arch-photogramm-remote-sens-spatial-inf-sci.net/XXXVIII-5-W12/31/2011/https://d-nb.info/1144553911/34.

Zorzi, S et al. (2019), "Full-Waveform Airborne LiDAR Data Classification Using Convolutional Neural Networks", *in*: *IEEE Transactions on Geoscience and Remote Sensing* 57.10, pp. 8255–8261, ISSN:

1558-0644, DOI: 10.1109/TGRS.2019.2919472, URL: http://ieeexplore.ieee.org/stampPDF/getPDF.jsp?tp=&arnumber=8737008&ref=aHR0cHM6Ly9pZWVleHBsb3JlLmllZWUub3JnL3

Zwally, H. J. et al. (Oct. 2002), "ICESat's laser measurements of polar ice, atmosphere, ocean, and land", *in*: *Journal of Geodynamics* 34.3-4, pp. 405–445, ISSN: 0264-3707, DOI: 10.1016/S0264-3707(02)00042-X.

COLLEGE ECOLOGIE
DOCTORAL GEOSCIENCES
BRETAGNE AGRONOMIE ALIMENTATION

Université de Rennes

Titre : Extraction de connaissances à partir de lidar topo-bathymétrique : apprentissage automatique et réseaux de neurones profonds pour les nuages de points et les formes d'onde.

Mot clés : interface terre/eau, lidar topo-bathymétrique, réseaux de neurones profonds

Résumé : Les interfaces terre-eau, fortement vulnérables au changement climatique et à la pression anthropique, requièrent une surveillance accrue. Toutefois, l'observation ininterrompue des zones submergées et émergées demeure un défi en raison de la présence d'eau. La télédétection lidar topo-bathymétrique constitue une solution adéquate en assurant une représentation continue des zones terre-eau, matérialisée par des nuages de points 3D et des formes d'ondes 1D. Cependant, une pleine exploitation de ces données requiert des outils encore en attente de développement. Cette thèse présente plusieurs méthodes d'extraction de connaissances des données lidar topo-bathymétriques, incluant des approches de classification basées sur des nuages de points bi-spectraux et des formes d'ondes bi-spectrales. En outre, des réseaux de neurones profonds sont conçus pour la segmentation sémantique, la détection et la classification d'objets, ainsi que l'estimation de paramètres physiques de l'eau à partir des formes d'ondes bathymétriques. L'utilisation de modèles de transfert radiatif guide des approches visant à réduire la nécessité de données labélisées, améliorant ainsi le traitement des formes d'ondes lidar dans les eaux très peu profondes ou turbides.

Title: Environmental knowledge extraction from topo-bathymetric lidar: machine learning and deep neural networks for point clouds and waveforms.

Keywords: land/water interface, topo-bathymetric lidar, deep neural networks

Abstract: Land-water interfaces face escalating threats from climate change and human activities, necessitating systematic observation to comprehend and effectively address these challenges. Nevertheless, constraints associated with the presence of water hinder the uninterrupted observation of submerged and emerged areas. Topo-bathymetric lidar remote sensing emerges as a suitable solution, ensuring a continuous representation of land-water zones through 3D point clouds and 1D waveforms. However, fully harnessing the potential of this data requires tools specifically crafted to address its unique characteristics. This thesis introduces methodologies for extracting environmental knowledge from topo-bathymetric lidar surveys. Initially, we introduce methods for classifying land and seabed covers using bi-spectral point clouds or waveform features. Subsequently, we employ deep neural networks for semantic segmentation, component detection and classification, and the estimation of water physical parameters based on bathymetric waveforms. Leveraging radiative transfer models, these approaches alleviate the need for manual waveform labeling, thereby enhancing waveform processing in challenging settings like extremely shallow or turbid waters.