



HAL
open science

Use of Data Science tools for Assessing Inland Water Surface and Quality on Regional Scales Through High-Resolution Sentinel-2 Remote Sensing Images

Mauricio Cezar Rebello Cordeiro

► **To cite this version:**

Mauricio Cezar Rebello Cordeiro. Use of Data Science tools for Assessing Inland Water Surface and Quality on Regional Scales Through High-Resolution Sentinel-2 Remote Sensing Images. Hydrology. Université Paul Sabatier - Toulouse III, 2022. English. NNT : 2022TOU30204 . tel-04161699

HAL Id: tel-04161699

<https://theses.hal.science/tel-04161699>

Submitted on 13 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE

**En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE
Délivré par l'Université Toulouse 3 - Paul Sabatier**

**Présentée et soutenue par
Mauricio Cezar REBELLO CORDEIRO**

Le 6 décembre 2022

**Exploitation des outils de la science des données pour le suivi de
la quantité et de la qualité des eaux de surface à l'échelle
régionale à partir de données de télédétection haute résolution
Sentinel-2**

**Ecole doctorale : SDU2E - Sciences de l'Univers, de l'Environnement et de
l'Espace**

Spécialité : Surfaces et interfaces continentales, Hydrologie

**Unité de recherche :
GET - Geosciences Environnement Toulouse**

**Thèse dirigée par
Jean-Michel MARTINEZ**

Jury

**M. Osmar ABÍLIO JR., Rapporteur
M. Jean-Stéphane BAILLY, Rapporteur
Mme Brigitte VINÇON-LEITE, Examinatrice
Mme Sabine SAUVAGE, Examinatrice
M. Jean-Michel MARTINEZ, Directeur de thèse
M. Bruno LARTIGES, Président**

Université Fédérale



Toulouse Midi-Pyrénées

THÈSE

En vue de l'obtention du
DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par l'Université Toulouse 3 - Paul Sabatier

Présentée et soutenue par

Mauricio Cezar REBELLO CORDEIRO

Le 6 décembre 2022

**Use of Data Science Tools for Assessing Inland Water Surface and
Quality on Regional Scales Through High-Resolution Sentinel-2
Remote Sensing Images**

Ecole doctorale : SDU2E - Sciences de l'Univers, de l'Environnement et de
l'Espace

Spécialité : Surfaces et interfaces continentales, Hydrologie

Unité de recherche :

GET - Geosciences Environnement Toulouse

Thèse dirigée par

Jean-Michel MARTINEZ

Jury

M. Osmar ABÍLIO JR., Rapporteur

M. Jean-Stéphane BAILLY, Rapporteur

Mme Brigitte VINÇON-LEITE, Examinatrice

Mme Sabine SAUVAGE, Examinatrice

M. Jean-Michel MARTINEZ, Directeur de thèse

M. Bruno LARTIGES, Président

Abstract

As demand for freshwater increases worldwide, water resources management is a priority for most developed and developing countries. Water security has been elected by the United Nations (UN) as one of the Sustainable Development Goals for the 2030 Agenda. In this regard, water monitoring is essential for understanding water changes in terms of quantity and quality supporting decision-makers. However, periodic monitoring of surface water remains challenging due to the high costs involved in performing field measurements, due to maintaining dedicated human resources, and laboratory analysis and the difficulty accessing remote locations. Moreover, sparse punctual and infrequent measurements do not capture the complex spatial and temporal variations that may occur in freshwater resources.

In this context, the present thesis proposes the use of data science tools such as Machine Learning (supervised and unsupervised), Cloud Computing, and High-Performance Computing (HPC) to leverage Remote Sensing (RS) as a potential tool that can be used to complement or even replace traditional ground-based surveys while providing advantages such as cost-effectiveness, higher temporal frequency, and spatial continuity.

Until the present, it lacks operational tools and products focusing on inland water bodies monitoring, in contrast to several products available for land or oceanic applications. This absence can be partially explained by the complex interactions among optically active constituents and the need for higher spatial and spectral resolution from sensors to monitor smaller water bodies, that leads to larger volumes of data.

To bridge some of these gaps, this thesis focuses on different aspects toward the development of a comprehensive water quality assessment framework that can be replicated in

different regions of the globe in regional scale. Three main axes have been identified: unsupervised and automatic water bodies detection; improved classification of optical water types for water quality parameters retrieval; and large-scale monitoring and spatial-temporal analysis.

The first part of this thesis focused in developing a fully automated water detection algorithm for optical imagery. The `waterdetect` algorithm uses a multidimensional clustering approach to provide robust detection regardless of land coverage type and specific calibration. It is capable of detecting reservoirs as small as 0.5 ha with fine accuracy (κ greater than 0.8). Moreover it is coupled with a random sampling and a subsequent machine learning classifier to deliver high performance and low memory consumption. After been validated over continental France (Cordeiro et al., 2021), it has been applied to several land coverage and challenging conditions (e.g., Greenland, Gabon, Chad, Mexico, Colombia, among others) and to track volume of smaller reservoirs (Peña-Luque et al., 2021; Tottrup et al., 2022).

The second part proposes using unsupervised machine learning to obtain water classification schemes to improve water quality inversion accuracy. Agglomerative clustering has been used to classify spectra from more than 1,000 field measurements acquired over 14 different watersheds before this thesis, and the solution has been optimized to decrease Suspended Particle Matter (SPM) retrieval errors. It is shown that by calibrating SPM models for just four optical water types (OWT), the overall accuracy matches the one obtained by calibrating one model individually for each river/basin. These finds make it possible to adopt one simple classification scheme for regional and global application, regardless of the inherent characteristics of each watershed.

Finally, in the third part, a comprehensive methodology has been proposed aiming at large-scale generalization and analysis of inland water surface and water quality parameters. In this part, a case study has been performed to demonstrate the applicability of RS in regional-scale water

assessment. The study focused in evaluating the effects of the major drought that hit Brazil's southeast in 2021, covering an area of 320,000 km². Several analyses considering water quantity and water quality (turbidity and Chl-*a*) were performed and validated with available field data. The results showed that RS allows to finely monitor water surface variation of thousands of water bodies that lack proper monitoring and to map spatial anomalies, revealing that the drought did not affect the region equally.

In the water quality domain, the case study successfully mapped seasonal variation of water quality parameters in the Paranapanema basin. Median turbidity and Chl-*a* were calculated for each month throughout the period of analysis, considering all water bodies above 1 ha. The median Chl-*a* ranged from 3.6 mg/m³ in May 2018 to 16.3 mg/m³ in September 2021, when water storage levels were lower, evidencing the drought can substantially affect eutrophic state.

In summary, the present thesis demonstrates the feasibility of using RS imagery for inland water assessment by addressing some of its scientific bottlenecks. Moreover, it represents a step forward in terms of large-scale and operational water monitoring in terms of quantity and quality and it also opens the path to better practices in water resources management. Part of the work developed herein (i.e., waterdetect) is being used operationally by the National Centre for Space Studies (CNES) in the French Data Center for Continental Surfaces (THEIA) processing workflow.

Resumé

Alors que la demande en eau douce augmente dans le monde entier, la gestion des ressources en eau est une priorité pour tous les pays développés ou en voie de développement. La sécurité hydrique a été choisie par les Nations Unies comme l'un des objectifs de développement durable de l'Agenda 2030. À cet égard, la surveillance de l'eau est essentielle pour comprendre ses changements en termes de quantité et de qualité afin de fournir des données robustes à destination des décideurs. Cependant, la surveillance périodique des ressources en eaux, notamment de surface, reste un défi en raison des coûts élevés liés à la réalisation de mesures sur le terrain, du maintien de ressources humaines dédiées, d'analyses en laboratoire et de la difficulté d'accéder à des sites éloignés. En outre, les mesures ponctuelles et peu fréquentes ne permettent pas de saisir les variations spatiales et temporelles complexes qui peuvent se produire au sein des lacs et rivières.

Dans ce contexte, la présente thèse propose l'utilisation d'outils de science des données tels que l'apprentissage automatique (supervisé et non supervisé), le cloud computing et le calcul haute performance (HPC, en anglais) pour tirer parti de la télédétection (RS, en anglais) comme un outil potentiel qui peut être utilisé pour compléter ou même remplacer les mesures conventionnelles au sol tout en offrant divers avantages tels qu'un bon rapport coût-efficacité, une fréquence temporelle plus élevée et une continuité spatiale.

Jusqu'à présent, il manque d'outils et de produits opérationnels axés sur la surveillance des masses d'eau intérieures, contrairement à plusieurs produits disponibles pour des applications terrestres ou océaniques. Cette absence peut s'expliquer en partie par les interactions complexes entre les constituants optiquement actifs et le besoin d'une résolution spatiale et spectrale plus

élevée des capteurs pour surveiller les petits plans d'eau, ce qui conduit à de plus grands volumes de données.

Afin de combler certaines de ces lacunes, cette thèse se concentre sur différents aspects du développement d'une chaîne d'évaluation complète de la qualité de l'eau qui peut être reproduit dans différentes régions du globe à l'échelle régionale. Trois axes principaux ont été identifiés: la détection non supervisée et automatique des masses d'eau ; le classement amélioré des types d'eau optiques pour la inversion des paramètres de qualité de l'eau ; et la surveillance à grande échelle et l'analyse spatio-temporelle.

Afin de combler certaines de ces lacunes, la première partie de cette thèse s'est concentrée sur le développement d'un algorithme de détection des surfaces en eaux entièrement automatisé pour l'imagerie optique. L'algorithme utilise une approche de regroupement multidimensionnel pour fournir une détection robuste indépendamment du type de couverture du sol et sans nécessité d'une calibration spécifique à la région d'étude. Il est capable de détecter des réservoirs aussi petits que 0,5 ha avec une précision fine (κ supérieur à 0,8). De plus, il est associé à un échantillonnage aléatoire et à une *machine learning* subséquent afin de délivrer des performances élevées et une faible consommation de mémoire. Après avoir été validé sur la France continentale (Cordeiro et al., 2021), il a été appliqué à plusieurs couvertures terrestres et conditions difficiles (par exemple, Groenland, Gabon, Tchad, Mexique, Colombie, entre autres) et pour suivre le volume de réservoirs plus petits (Peña-Luque et al., 2021; Tottrup et al., 2022).

La deuxième partie propose d'utiliser le *machine learning* non supervisé pour classer les masses d'eaux intérieures en différents types afin d'améliorer la précision de l'inversion de la qualité de l'eau. Le regroupement agglomératif a été utilisé pour classer les spectres provenant de plus de 1 000 mesures sur le terrain obtenus dans 14 bassins versants avant la thèse, et la solution

a été optimisée pour réduire les erreurs de récupération des matières en suspension (SPM, en anglais). Il est démontré qu'en calibrant les modèles SPM pour seulement quatre classes optiques d'eau (OWT, en anglais), la précision globale correspond à celle obtenue en calibrant un modèle individuellement pour chaque rivière/bassin. Ces résultats permettent d'adopter un schéma de classification simple pour une application régionale et globale, indépendamment des caractéristiques inhérentes à chaque bassin versant.

Pour finir, dans la troisième partie, une méthodologie complète a été proposée afin de généraliser et d'analyser à grande échelle les paramètres de surface et de qualité des eaux intérieures. Dans cette partie, une étude de cas a été réalisée pour démontrer l'applicabilité de la télédétection dans l'évaluation des eaux à l'échelle régionale. L'étude s'est concentrée sur l'évaluation des effets de la grande sécheresse qui a frappé le sud-est du Brésil en 2021, couvrant une zone de 320 000 km². Plusieurs analyses portant sur la quantité et la qualité de l'eau (turbidité et Chl-*a*) ont été réalisées et validées à l'aide des données de terrain disponibles. Les résultats ont montré que le RS permet de surveiller finement la variation de la surface de l'eau de milliers de masses d'eau qui manquent de surveillance adéquate et d'analyser les anomalies spatiales, révélant que la sécheresse n'a pas affecté la région de manière similaire.

Dans le domaine de la qualité de l'eau, l'étude de cas a évalué la variation saisonnière des paramètres de qualité de l'eau dans le bassin du Paranapanema. La turbidité médiane et la Chl-*a* ont été calculées pour chaque mois tout au long de la période d'analyse, en considérant tous les plans d'eau supérieurs à 1 ha. Le Chl-*a* médian variait de 3,6 mg/m³ en mai 2018 à 16,3 mg/m³ en septembre 2021, lorsque les niveaux de stockage de l'eau étaient plus bas, ce qui montre que la sécheresse peut affecter considérablement l'état eutrophe.

En résumé, la présente thèse démontre la faisabilité de l'utilisation des images de télédétection pour évaluer des eaux intérieures en abordant certains de ses obstacles scientifiques. En outre, elle représente un pas en avant en matière de surveillance de l'eau à grande échelle en termes de quantité et de qualité et elle ouvre également la voie à de meilleures pratiques en matière de gestion des ressources en eau. Une partie des travaux développés dans cette thèse (par exemple, le waterdetect) est utilisée de manière opérationnelle par le Centre National d'Etudes Spatiales (CNES) dans la chaîne de traitement des surfaces continentales (THEIA).

Acknowledgements

This thesis has been possible thanks to the support of several institutions and people involved.

First and foremost, I would like to thank my actual employer, the Brazilian National Water and Sanitation Agency (ANA), for providing me with the funds and time to accomplish my Ph.D. studies. Without this support, it would be impractical even to start the thesis.

I am grateful to my supervisor Dr. Jean-Michel Martinez for buying the idea since our first talks, hosting me at the *Géosciences Environnement Toulouse* (GET) laboratory in Toulouse, and for his guidance throughout the thesis. The “steering comitée” formed by Brigitte Vinçon Letite, Laurent Kergoat, and Santiago Peña-Luque, has also been supportful and has encouraged me whenever I needed it. Thanks to the Centre for Spatial Studies (CNES) for providing the computational workforce needed and other institutions involved in the Hydrology for Amazon Basin (Hybam) project that shared research data for the current study.

I am also grateful to my colleagues for providing a good working environment, constructive discussions, and overall assistance. Special thanks to David Guimarães and Daniel Moreira, which were always ready to help me in their knowledge fields, and to all the good friends I have met in Toulouse with whom I shared both fun and hard times.

Lastly, but certainly not least, an exceptional thought goes to my family. My parents that believed in the project since the first time until the very end. My wife and children that didn't hesitate to move their lives entirely to a new country, leaving behind family and friends, to follow me in this journey. Fernanda, Lucas and Luisa, I love you.

Table of contents

Abstract	i
Resumé	iv
Acknowledgements	viii
Table of contents	ix
List of Abbreviations and Acronyms	xii
Introduction (English)	14
1. Global Water Challenges	14
2. Remote Sensing	16
2.1 Remote Sensing Applied to Water Resources	19
3. Data science tools for Remote Sensing	23
3.1 Machine Learning	24
3.2 Deep Learning.....	28
4. Generalization and Operational Tools.....	29
5. Objectives	31
6. Thesis Structure	36
7. Development context of the Thesis	39
Introduction (Français).....	41
1. Les enjeux mondiaux liés à l'eau	41
2. Télédétection	44
2.1 La télédétection appliquée aux ressources hydriques	46
3. Les outils de science des données pour la télédétection.....	51
3.1 <i>Machine Learning</i>	52
3.2 Deep Learning.....	57
4. Généralisation et outils opérationnels.....	58
5. Objectifs.....	60
6. Structure de la thèse.....	66
7. Contexte du développement de la thèse	69
Chapter I - Water Detection	71

1. Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors	76
Chapter II - Optical Water Classification.....	94
1. Article Draft - Optical water type classification for suspended Particulate Matter Retrieval Over inland tropical waters.....	94
1.1 Abstract	94
1.2 Introduction.....	95
1.3 Materials and methods	100
1.4 Results.....	113
1.5 Discussion	142
1.6 Conclusion	145
Chapter III – Generalization and Analysis: A case study of Brazil’s 2021 drought.....	148
1. Water Detection and Assessment Methodologies	150
1.1 Large scale water detection methodology.....	150
1.2 Water surface and quality Assessment Methodology	159
2. Article Draft - Assessing the Impacts on Water Surface and Quality during Drought Events Using High Resolution Optical Satellite Imagery: Parana River 2021 drought (Brazil).	168
2.1 Abstract	168
2.2 Introduction.....	169
2.3 Materials and Methods.....	172
2.4 Results.....	184
2.5 Discussion	203
2.6 Conclusion	214
Conclusion and Prospects.....	217
Conclusion et Perspectives.....	227
References	238
List of Figures	246
List of Tables.....	255
Annexes.....	257
I. Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level	257

II. Surface Water Dynamics from Space: A Round Robin Intercomparison of Using Optical and SAR High-Resolution Satellite Observations for Regional Surface Water Detection 284

III. Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, PlanetScope and Sharpened imagery306

List of Abbreviations and Acronyms

ANA	Agência Nacional de Águas e Saneamento Básico / Brazilian National Water and Sanitation Agency
ANN	Artificial Neural Networks
CNES	Centre National d'Études Spatiales / French National Centre for Space Studies
DL	Deep Learning
DT	Decision Trees
EM	Electromagnetic Energy
ESA	European Space Agency
FMask	Function of Mask cloud detector
GDP	Gross Domestic Product
GEE	Google Earth Engine
GET	Géosciences Environnement Toulouse / Geosciences Environment Toulouse laboratory
GMT	Greenwich Mean Time
GPU	Graphics Processing Unit
GRS	Glint Removal for Sentinel atmospheric correction algorithm
GSW	Global Surface Water
HDI	Human Development Index
HPC	High Performance Computing
HTTP	Hypertext Transfer Protocol
IRD	Institute de Recherche pour le Développement / French National Research Institute for Sustainable Development
iWSSM	International Centre for Water Security and Sustainable Management
KNN	K-Nearest Neighbors
LANDSAT	Land Remote-Sensing Satellite
MACCS	Multi-sensor Atmospheric Correction and Cloud Screening
MAJA	Maccs-Atcor Joint Algorithm. Multi-temporal cloud detection and atmospheric correction algorithm
ML	Machine Learning
MNDWI	Modified Normalized Difference Water Index
MSI	Multispectral Instrument (Sentinel-2 satellite sensor)
MSS	Multispectral Scanner (Landsat satellite sensor)
NASA	National Aeronautics and Space Administration
NIR	Near-infrared wavelength region
OAC	Optically Active Constituents
OBS2CO	Observation de la Couleur des eaux Continentales project

OFB	Office français de la biodiversité / French Office for Biodiversity
OWT	Optical Water Tipology
PSD	Particle Size Distribution
PBIAS	Percent BIAS metric
RF	Random Forests
RMSE	Root mean squared error
RMSLE	Root mean squared log error
RS	Remote Sensing
SAR	Synthetic Aperture Radar
SDG	UNESCO's Sustainable Development Goals for Agenda 2030
Sen2Cor	Sentinel 2 level 2A processor
SPM	Suspended Particle Matter
SPOT	Satellite Pour l'Observation de la Terre / Earth Observation Satellite
STAC	Spatial Temporal Assets Catalog
SVM	Support Vector Machines
SWIR	Short wave infrared wavelength region
THEIA	French data center for continental surfaces
UFAM	Universidade Federal do Amazonas / Federal University of Amazonas
UN	United Nations organization
UNB	Universidade de Brasília / Brasilia University
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
WRT	Water residence time
WWAP	UNESCO's World Water Assessment Programme
XML	Extensible Markup Language

Introduction (English)

1. Global Water Challenges

Water is the most important natural resource on Earth, as it is necessary to sustain all kinds of life. Besides, it is vital for human, social and economic development, food and energy production, and maintaining a healthy environment. Unfortunately, although more than 70% of the world's surface is covered by water, most of this total is salty water within the oceans (98%). The remaining freshwater (2%) is mostly stored in ice caps or deep subterranean aquifers, leaving just a tiny fraction of the total (0.024%) on the surface available for human consumption (Gleick et al., 1993).

Estimates show that global freshwater withdrawal has increased by six times over the past 100 years, reaching more than 4,000 km³/year. Agriculture (69%) and industry (19%) are the most water-demanding activities. Much of this growth is credited to a combination of population growth, economic development, and shifting consumption patterns (UNESCO-WWAP, 2021).

The world's population, which is currently 7.6 billion people, is expected to reach 10 billion by 2050 and 11 billion by 2100, according to projections from the United Nations (UNDP, 2020). In addition to population growth, a rise in living standards, noticeable by the increase of the Human Development Index (HDI) over the years, also translate into higher pressure on natural resources, predominantly freshwater (UNDP, 2022a).

Moreover, the urban population increased from 2.3 billion people in 1990 to 4.0 billion in 2015, which represents 51% of the total population, and this concentration is expected to grow up in the next years, reaching up to 68% by 2050, according to the World Urbanization Prospects (UN, 2018). In this context, understanding the spatial distribution of water resources in terms of quality

and quantity is also crucial, as this movement towards urbanization puts even more pressure on localized, high-density areas with scarce water sources.

Despite different projections of future trends in water use and the uncertainty of its global increase, a growing fraction of the population will face some form of water scarcity in the coming years (UNESCO-WWAP, 2021). Furthermore, climate-related changes can also affect evaporation and rainfall patterns due to temperature rise leading to more extreme water-related events such as floods and droughts throughout the globe.

In addition to the pressure for additional water resources caused by the aforementioned trends, water quality has also deteriorated due to the release of various contaminants in ground and surface water resources, led by increased human activities. Half of the world's rivers and lakes are polluted, according to Makarigakis and Jimenez-Cisneros (2019), and this situation can also be largely degraded by climate changes.

To face these challenges and to include sustainable water resource management in the global political agenda, the United Nations has devoted one of its Sustainable Development Goals (SGD), also known as Global Goals, for the 2030 Agenda, to ensure availability and sustainable management of water and sanitation for all (6th goal) (UNDP, 2022b).

One aspect to consider is that most world's water-related issues are not driven by an actual lack of water. Instead, they result from poor and inefficient water resources governance (UNESCO-iWSSM, 2019). In this context, water monitoring is essential for understanding water changes in quantity and quality. The United Nations World Water Development Report (UNESCO-WWAP, 2021) argues that to achieve sustainable and long-term water resources management, it is essential to incorporate better measurement and monitoring into decision-making frameworks. This way, stakeholders can make equitable comparisons of the multiple values of water.

Moreover, periodic monitoring of water bodies and integrated management of river basins are essential for water resource conservation. However, there are many challenges to overcome regarding field monitoring, especially for non-developed countries, such as:

- 1) Time and costs involved in performing field measurements and laboratory analysis;
- 2) Punctual samples do not always capture spatial and temporal variations; and
- 3) Continuous field monitoring is not feasible in remote, difficult-access regions.

Therefore, the use of alternative data sources from remote sensors, especially from orbital satellites, has been rapidly increasing in the last decades and is helping the scientific community to better understand the complexity of water resources management on larger scales. In addition, these technologies have shown the potential to complement or replace traditional ground-based surveys and provide additional advantages such as cost-effectiveness, higher frequency, and better spatial continuity (Barbosa et al., 2019).

2. Remote Sensing

According to Kairu (1982), the term remote sensing can be defined as “*the collection and interpretation of information about an object, a region or event without being in physical contact with the object*”. A simple photograph followed by the analysis of its contents is then considered a kind of remote sensing. The sensors can be mounted on any platform, such as an airplane or a research balloon. Spatial remote sensing for earth observations is not a new concept. It started effectively with the launch of LANDSAT-1 in 1972 by the National Aeronautics and Space Administration (NASA), followed by the French Space Agency (CNES) with the launch of SPOT-1 in 1986.

Although considerable progress has been made in remote sensing for earth observations since then, the potential of satellite imagery remains under-exploited in numerous specialized

fields, especially by public actors and decision-makers responsible for elaborating and applying public policies. In the early days, the cost of images and the lack of experts with the proper knowledge for processing and analyzing these data were some of the causes for this under exploitation of remote sensing data (Baghdadi and Zribi, 2016).

However, the capacity of remote sensing to survey the land surface and environmental conditions over large areas has increased over the last few years, mostly due to the deployment of several earth observation satellites that provide inexpensive data all over the globe. Moreover, universities have adapted their curriculum and distinct courses, such as Geology, Geography, Ecology and others, now consider remote sensing-related axes.

In practical terms, the information collected by a sensor is the energy emitted by the object, frequently called as target, in different parts of the electromagnetic (EM) spectrum. The region of interest of the EM spectrum in remote sensing goes from short wavelengths (visible portion), which corresponds to the EM spectrum that the human eye can view, generally placed between 380 to 700nm, to infrared and longer wavelengths (microwaves).

The radiation emitted or reflected by the different targets on Earth will directly depend on the incoming energy and the absorption/reflection of this energy by the target. This interaction is specific for each type of material (Figure 1-a) and its physico-chemical properties. There exist two types of sensors, depending on the energy source. The passive sensors rely on external sources of energy, usually the sun, and the active sensors have their own source of energy, as in the case of radars.

Once the downwelling energy hits the target, a portion of this energy can be absorbed by the object. This absorption varies over the range of wavelengths in the EM spectrum and the portion of energy not absorbed is scattered. The backscattered energy is measured by the sensor in different

wavelength regions, defined as the spectral bands of the sensor. The relationship between the upwelling radiance (measured by the sensor) and the downward irradiance is called reflectance and can be estimated for each wavelength across the spectrum. Considering a single energy source and that the absorption and scattering processes are a function of the material, each target will have a unique spectral pattern called spectral signature (Figure 1b).

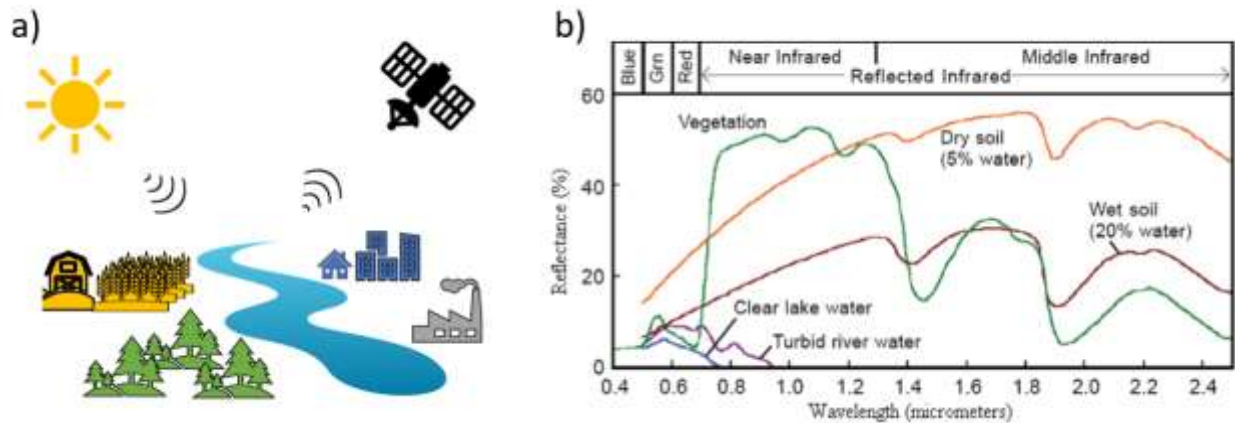


Figure 1: (a) Remote sensing schema. (b) Spectral signatures for distinct natural targets. Extracted from Mercan and Alam (2011).

Many other factors can interfere with the remote measurements, such as interactions of both the downwelling and upwelling light with atmospheric particles and the sun's relative position. A detailed presentation of these physical concepts is beyond the scope of the present thesis but can be found in Guo et al. (2020), Kairu (1982), and Baghdadi and Zribi (2016).

The observation of these reflectance characteristics through space and time makes it possible for the scientists and researchers to model different phenomena and to better understand and characterize our environment (Baghdadi and Zribi, 2016).

2.1 Remote Sensing Applied to Water Resources

Satellite platforms have been used for more than 50 years to understand complex processes such as global land use, vegetation analysis, and ocean productivity or to establish connections between industrial production and carbon cycling (Topp et al., 2020).

The focus on understanding these global coarse processes can be explained by the spatial, spectral, and temporal resolutions provided by earlier platforms. As a reference, the LANDSAT-1 platform, launched in 1972, was the first satellite designed specifically for Earth-observing applications and it was loaded with a Multispectral Scanner System (MSS) with 83m of spatial resolution and just 4 spectral bands (visible and near-infrared) with a revisit time of 18 days. Other ancient oceanic and atmospheric satellites had even coarser spatial resolutions and higher temporal frequencies.

Concerning the use of remotely sensed images for inland waters assessment, Topp et al. (2020) argue that, despite an increasing number of publications, there are limited global products towards inland water quality of rivers, lakes and reservoirs, in comparison to land and oceanic products. This slow evolution can be partially explained by the challenges related to the remote sensing of continental waterbodies such as limited capacity of sensors to detect small bodies and the complex interactions among optically active constituents (OACs). OACs are dissolved and particulate water constituents that can absorb and scatter light (e.g., particulate matter, chlorophyll, colored dissolved organic matter, and others), giving water its characteristic color.

One example of this complexity can be seen in Figure 2, which shows the mean spectral response of several rivers considered in the present study, measured by field hyperspectral radiometers, and how this reflectance can vary depending on its constituents, geographic region,

etc. These observations contrast with the overall assumption of null water-leaving radiance at the near-infrared (NIR) wavelengths, commonly adapted for clear oceanic waters (i.e. Case 1 Water).

Thankfully, there has been an increased capacity of remote-sensing Earth observation satellites in recent years. In this regard, it is worth noting the Copernicus program, designed by the European Space Agency (ESA), with missions that focus on different aspects of Earth observation: Atmospheric, Oceanic and Land monitoring (Suhet and Hoersch, 2015).

The satellites of the Sentinel-2 mission, part of the Copernicus observation program, are equipped with optical Multispectral Instrument (MSI) that offers high spatial, spectral and temporal resolutions at a global scale without cost to the user. The first satellite of the Sentinel-2 family (Sentinel-2A) was launched by the European Space Agency (ESA) in 2015, followed by its twin satellite (Sentinel-2B) launched in 2017. Both are polar-orbiting satellites and operate phased at 180° to each other to provide a high revisit time of 2-3 days in mid-latitudes and 5 days revisit time near the equator (Suhet and Hoersch, 2015). The Sentinel-2 has been selected to address inland water quality issues, which is the scope of this thesis, as it provides a resolution that is suitable to

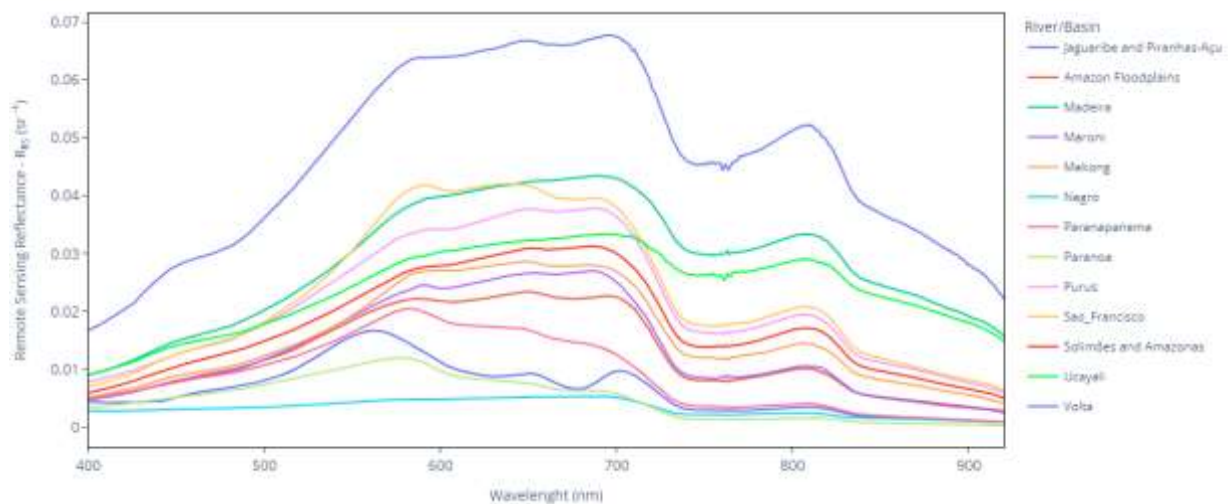


Figure 1: The mean spectral profiles of several rivers/basins considered in the present study measured with field radiometers.

better detect narrow and linear features as river channels and smaller water bodies that would not be visible with other medium resolution sensors.

Moreover, it can capture a wide range of spectral wavelengths separated into 13 spectral bands ranging from 443 to 2200 nm, and spatial resolutions ranging from 10 to 60m depending on the spectral band. Another information that must be considered when using satellite reflectance to retrieve water quality parameters is the spectral resolution provided by each band as it can interfere in the reflectance actually “seen” by the satellite. The Sentinel-2 MSI sensor has bandwidths ranging from 15 to 180 nm, making it possible to identify different water colors and then infer the properties of their constituents. However, it is important to note that the methodologies described hereafter are not satellite specific and could be applicable to other platforms with different technical specifications. **Error! Reference source not found.** presents the spectral response functions for the 13 bands available in the Sentinel-2 MSI sensor and their corresponding spatial resolutions.

This extended capacity provided by newer satellites enables scientists to explore different possibilities concerning the inland water domain, including surface water coverage (Buma et al., 2018; Cordeiro et al., 2021; Feyisa et al., 2014; Markert et al., 2018; Pekel et al., 2016; Peña-Luque et al., 2021; Souza et al., 2019), flood and inundation mapping (Kordelas et al., 2019, 2018; Martinis et al., 2011; Wieland and Martinis, 2019), water quality monitoring (Ansper and Alikas, 2018; Delegido et al., 2014; Frampton et al., 2013; Lins et al., 2017; Toming et al., 2016; Uudeberg et al., 2020; Yadav et al., 2019), suspended sediment assessment (Condé et al., 2019; Espinoza Villar et al., 2013; Martinez et al., 2009; Yopez et al., 2018), among others.

However, the growing number of scientific publications on the water resources domain, does not translate into operational products in regional or global scale, which is crucial for decision-makers (Topp et al., 2020)

One exception that is worth mentioning is the Global Surface Water (GSW) project (Pekel et al., 2016), which was developed to assess water surfaces worldwide and track their long-term changes, making use of Landsat-8 optical imagery with 30m resolution. However, the objective of the GSW project is to track long-term monthly changes, but to perform operational water quality assessment through remotely sensed images, higher spatial and temporal resolutions are required. In this way, the GSW dataset cannot be used for continuous monitoring.

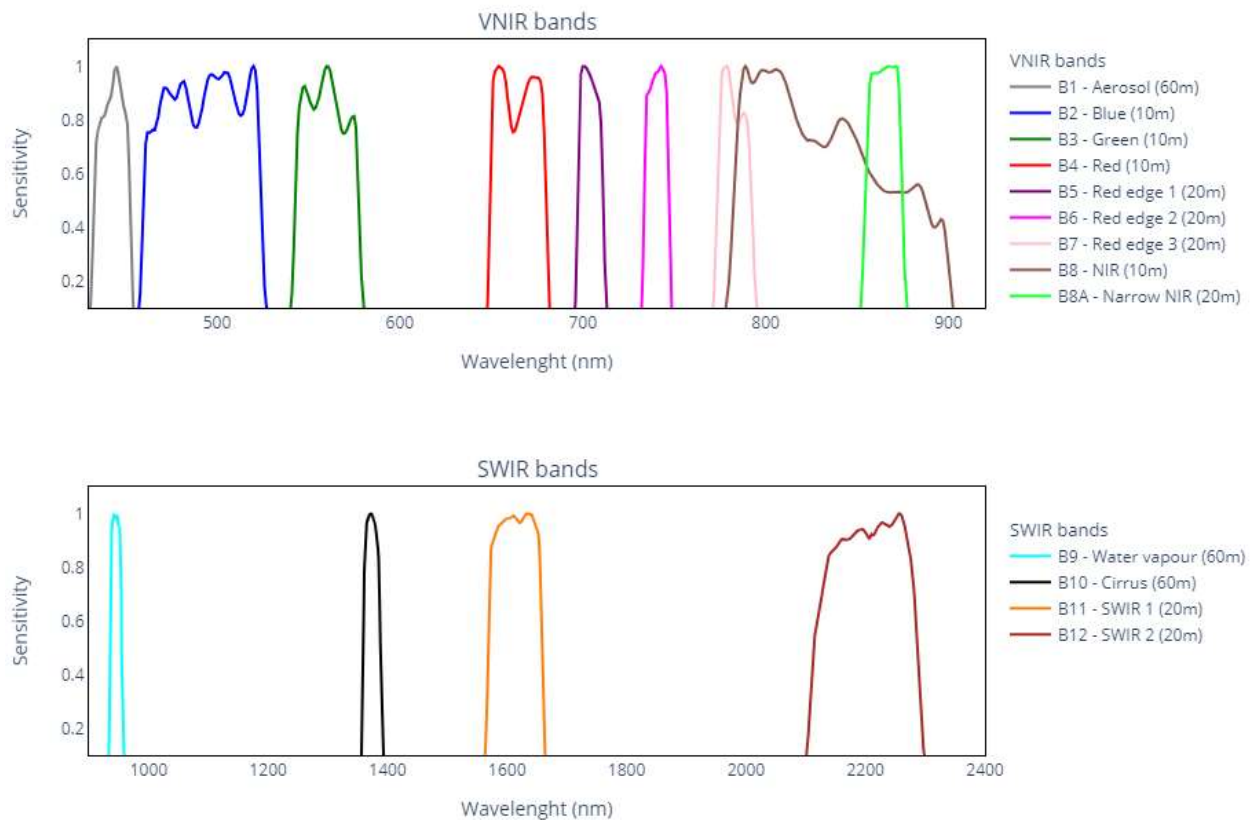


Figure 2: Spectral response functions for Sentinel-2/MSI, Visual-NIR, and SWIR bands and their corresponding spatial resolutions.

Among the barriers that limit inland water quality assessment through satellite remote sensing, Chami et al. (2015) highlight the difficulty of controlling the variation of atmospheric effects. Historic correction procedures rely upon zero water leaving radiance beyond the visible spectrum. This assumption cannot be generalized to optically complex inland waters, as the

presence of inorganic and organic matter increases the light backscattering processes in the NIR spectrum (Figure 2).

Moreover, optical sensors are subject to the atmospheric conditions such as clouds, cloud shadows and other surface occurrences that can interfere in the detection of the water pixels, such as snow, for example. Thus, besides an atmospheric correction that considers the presence of optically complex water, another key element to provide operational products is the application of robust masking methods that are able to correctly identify (and possibly discard) undesired elements in the scene.

3. Data science tools for Remote Sensing

In the previous section, we broadly discussed the uses of satellite remote sensing images for earth observations and how they can be a valuable tool to assess the physical, chemical, and biological status of the inland water bodies, despite their challenges.

In the early days, remote sensing was mainly based on optical photography (confined between the visible and infrared spectral range) captured from aerial platforms. The photo-interpretation of these images was performed by trained technicians responsible for identifying elements and patterns without the aid of automated processes (Baghdadi and Zribi, 2016). However, this process is time-consuming and unfeasible, considering the amount of data available in recent years. In addition, remote sensing has evolved from primarily optical to using the entire spectrum of wavelengths, from the visible to microwaves, turning the images into n -dimensional cubes that are difficult to be interpreted manually.

In this context, considerable progress in the development of computer-assisted methods for processing and analysis of remote sensing data is being made, and data science tools such as Data Mining, Machine Learning (ML), or Deep Learning (DL) are now playing a central role.

3.1 Machine Learning

Machine Learning is a subfield of the artificial intelligence domain, together with Deep Learning. The term was coined in 1959 by Arthur Samuel, an American pioneer in the fields of computer gaming and artificial intelligence, and describes the ability to make predictions or decisions without being explicitly programmed to do so (Mitchell, 1997).

To understand what ML is, one may think of a problem that has to be solved programmatically. Sticking to the remote sensing context, imagine, for example, we need to develop a program that can identify clouds in a satellite image. The straightforward approach would be to map, beforehand, all the characteristics a cloud would have (e.g., color, temperature, reflection intensity, among others) and how these characteristics translate into band reflectance values and write clauses to verify if each pixel lies within those rules. This is called rule-based approach. However, considering the diversity of clouds and their interaction with the ground (e.g., a semi-transparent cirrus), it can be cumbersome, or even impractical, to map all possible values to program this algorithm accurately.

In the ML approach, instead of mapping all possible values that can occur, the ML algorithm is presented with data previously labeled as cloud or non-cloud by a human, and the computer “learns” from the given data how to distinguish clouds. This method, where data is labeled previously, is called supervised learning. Other approaches when no label is provided are also available and are called unsupervised learning (e.g., clustering and dimensionality reduction) and will be discussed further. The fundamental differences between rule-based and ML programming are highlighted in **Error! Reference source not found.**

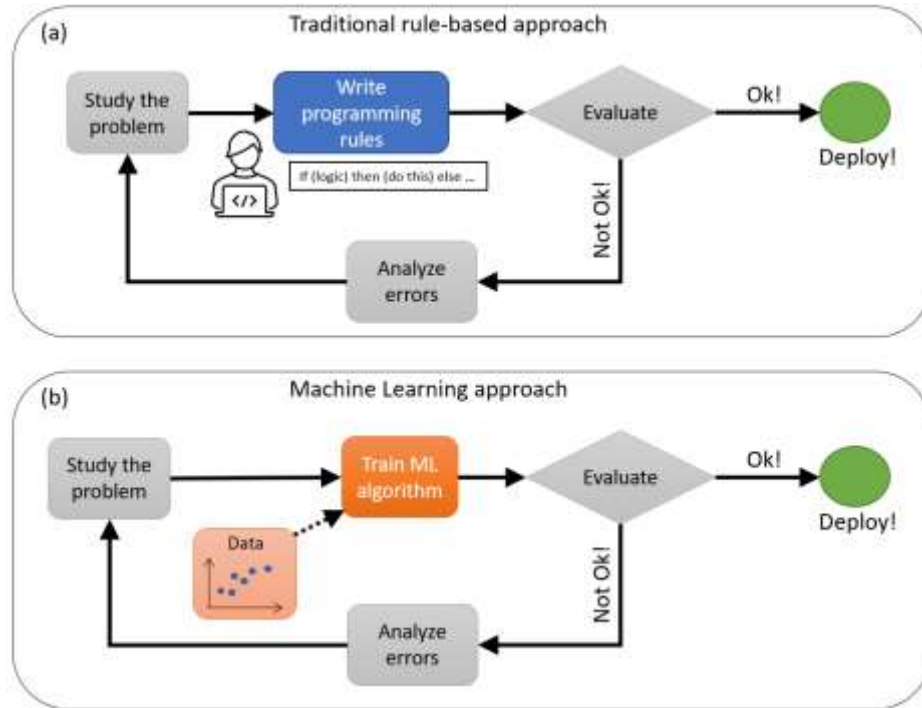


Figure 3: Schematic comparison of traditional programming approach (a) and machine learning approach (b). Adapted from Géron (2017).

One advantage of ML methods over traditional approaches is that they do not require previous knowledge concerning the intrinsic data properties and statistics. Instead, the relationship among input variables (also called input features) is obtained automatically during the training process, where internal parameters are calibrated to improve accuracy.

The ML-based techniques have proven useful for a very large number of applications, such as land cover classification, object detection, empirical retrieval algorithms, among others and distinct Earth observation domains (e.g., land, ocean, and atmosphere) (Lary et al., 2016). Therefore, ML methods are increasingly used for interpreting all kinds of remote-sensing images.

To extract relevant information from multispectral images, ML techniques can be divided into two main groups: 1) supervised learning; and 2) unsupervised learning.

Supervised Learning

In the supervised learning context, the goal is to estimate variables (qualitative classes, in the case of classification problems, or continuous values in regression problems) from samples in a data set, given that some of these variables are provided for a subset called training dataset. Based on the training dataset and according to the supervised algorithm being used, the computer can establish a relationship between the data and the desired variables internally. Once the model is trained (i.e., internal parameters calibrated during the training or learning phase), it can be used to estimate the variables for all the samples in the data set, even if we cannot fully understand the internal relationship (Baghdadi and Zribi, 2016).

Because of these characteristics, ML is usually used in domains such as difficult-to-program applications or when our theoretical knowledge is still incomplete, but for which we do have a significant number of observations and other data (Lary et al., 2016).

In the example presented before, concerning cloud detection, the algorithm can establish a relationship considering as many input variables as we want. For example, we could provide the totality of multispectral bands and the time of acquisition. The variables that are passed to the ML algorithm during the learning process are called input features, and the program will use them to predict the outputs of new samples.

The cloud detection application is a classification example, as the expected output is qualitative (cloud or non-cloud). The same approach could be used to retrieve a quantitative output, such as land temperature or evapotranspiration, for example. In this last case, we consider it a regression application. Several methods exist in the supervised learning context. Among them, the most commonly used ones in remote sensing are: support vector machines (SVM), decision trees (DT), random forests (RF), artificial neural networks (ANN), k-nearest neighbors (KNN), among

others (Lary et al., 2016). A detailed explanation of such methods is outside the scope of the present thesis, however, comprehensive descriptions are available in the literature (Géron, 2017; Vanderplas, 2016).

Unsupervised Learning

Unsupervised techniques are commonly used when we have no prior knowledge about the data being analyzed (i.e., no labels), and we want the computer to extract meaningful correlations from it. For example, suppose we want to identify how many different crops appear in a given scene. Partition algorithms, also called clustering algorithms, can be used to automatically group the data into subsets without intervention or training data. Several clustering algorithms exist in the literature, and the most commonly used in the remote sensing domain are k-Means, agglomerative clustering and DBScan. **Error! Reference source not found.** shows an example of agglomerative clustering applied to several pixels in a Sentinel-2 scene to separate different classes of pixels automatically. This kind of application will be further explored in Chapter 1 – Water Detection.

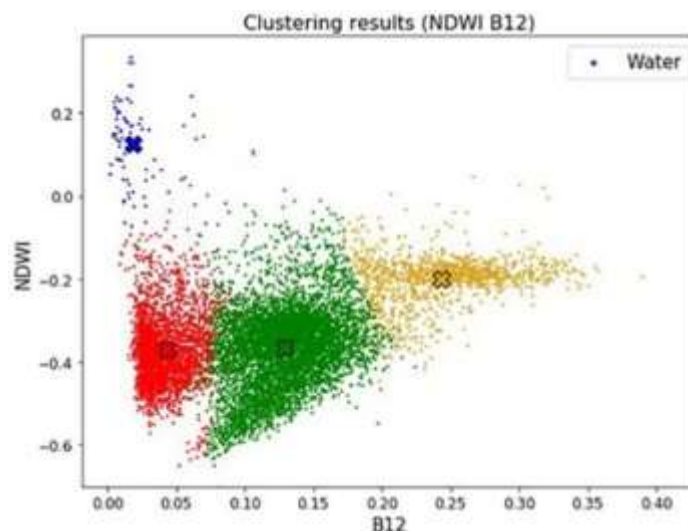


Figure 4: Scatter plot of the Normalized Difference Water Index (NDWI, Y-axis) and SWIR reflectance (X-axis) in sr^{-1} for several pixels in a Sentinel-2 scene, showing clustering results in different colors.

Another group of techniques used within unsupervised learning are called visualization or dimensionality reduction methods. In dimensionality reduction, the task is to simplify the data, preserving as much information as possible. This can be done by eliminating correlated input features or merging several features into one new synthetic feature. This process of crafting input variables to achieve better results is called feature extraction or feature engineering and it is a key step in machine learning applications (Géron, 2017).

3.2 Deep Learning

In the previous section, we have mentioned ANN as one of the tools for supervised learning with ML. Although the practical application of ANNs is recent, its theoretical conception date back to 1943, when McCulloch and Walter Pitts, a neurophysiologist and a cybernetician, respectively, proposed the first mathematical model a neural network (Abrahart et al., 2004). An ANN is formed by simple processing units called perceptrons, which are inspired by the human neurons that connect to other perceptrons, creating a network that would mimic the brain.

However, due to difficulties to train an ANN, its use remained neglected by the scientific community for a long time. More recently, with the increase in computational power, especially provided graphics processing unity (GPUs), the availability of training data, and some technical improvements in the training step (e.g., backpropagation) the ANNs received a lot of attention.

In this context, we define DL models as ANN models composed of several stacked layers, typically more than three. The more stacked layers, the model becomes more profound, and more complex interactions can be learned, hence the name Deep Learning. These deep architectures proved successful in extracting information from abstract concepts and outperformed classical ML methods in a number of applications, such as speech recognition and image recognition (Hoeser and Kuenzer, 2020).

According to Hoerer and Kuenzer (2020), the use of DL reached the Earth Observation community in 2012. Since then, the number of publications has doubled each year with a wide range of applications, such as super-resolution, data fusion, denoising, weather forecasting, scene recognition, classification, object detection and image segmentation.

4. Generalization and Operational Tools

Although multispectral visible optical remote sensing is now well known by scientists and actors in charge of managing the environment, putting together all these pieces and making them general enough to work in different regions with diverse land covers and atmospheric conditions remains a challenge for the inland water domain. In addition, we faced a radical increase in the volume of data (Big Data) acquired by higher-resolution sensors, which stands as a new issue in the processing of satellite data.

Within global surface water detection, an effort towards this generalization has been recently proposed by the World Water project (Tottrup et al., 2022). It was conducted an inter-comparison round robin with the objective to assess several optical and radar methodologies aiming to provide high-resolution regional surface water detection in different challenging regions across the globe.

Another difficulty towards generalization concerns the development of global retrieval algorithms that make it possible to assess water quality parameters independently of the monitored water body. In oceanographic research, the delivery of synoptic maps of global chlorophyll-*a* (Chl-*a*) concentration from radiometric sensors mounted on satellites has been a reality for decades. Concerning inland waters, even though many different models have been proposed in the literature for different parameters, the optical complexity in these waters frequently limits its operational use to localized scales (Neil et al., 2019).

Besides that, in recent years, massive amounts of satellite data have been generated, dramatically expanding our ability to understand our environment from both spatial and temporal perspectives. However, this unprecedented volume of data (the volume of remote sensed imagery collected by satellites or drones could easily reach Terabytes or Petabytes levels) brought to light the challenges and bottlenecks involved in storing, processing, analyzing and ultimately visualizing all this information (Guo et al., 2020).

This means that methodologies developed locally are not feasible to be applied to larger scales, because storage, memory and processing constraints will just prevent them from working correctly, or they will not produce the results in the expected time, making them obsolete. And that's another reason why most inland water quality products and publications are usually constrained to smaller areas or specific water bodies, and it makes them difficult to be applied on regional or global scales to help us understand the spatial context of these complex interactions.

In this context, new computational paradigms developed in other disciplines, such as Data Mining, Grid Computing, Cloud Computing and High-Performance Computing (HPC) are being incorporated into the set of optical remote sensing tools to provide processing support to address these challenges (Guo et al., 2020).

These approaches are necessary depending on the tasks being performed as they unlock the possibility of processing amounts of data unimaginable to be done by a single processing instance. However, being able to process these huge amounts of data is just one step. The satellite image is still a complex product so it is also fundamental to develop methodologies that can be used to extend the analysis to another spatial and temporal scale.

5. Objectives

Remote sensing can be an efficient tool to enable large-scale monitoring to address water quantity and quality issues. However, creating an operational framework that can assess inland water quality from satellite remote sensing involves many steps and the careful selection of several parameters, so it can all work in an integrated manner. From selecting the satellite platform to identifying the atmospheric correction model and developing a robust inversion algorithm, all these steps must be orchestrated in a meaningful way, as depicted in Figure .

In this context, this thesis proposes to work on different aspects of this chain, making use of existing methodologies and proposing new advancements towards robustness, scale, and automation, through the extensive application of data science techniques.

Therefore, the overall objective is to provide a comprehensive framework that can be replicated in different regions of the globe for inland water quality studies through remotely sensed imagery. Among a diversity of topics involved, the aspects to be explored in detail were deliberately selected during the development of the thesis and were separated into three main axes listed below:

- 1- Unsupervised and automatic water bodies detection;
- 2- Improved classification of optical water types for inland waters parameters retrieval;
and
- 3- Satellite-based large-scale monitoring and spatial-temporal trend analysis.

The objectives within each defined axis are detailed in the following sections.

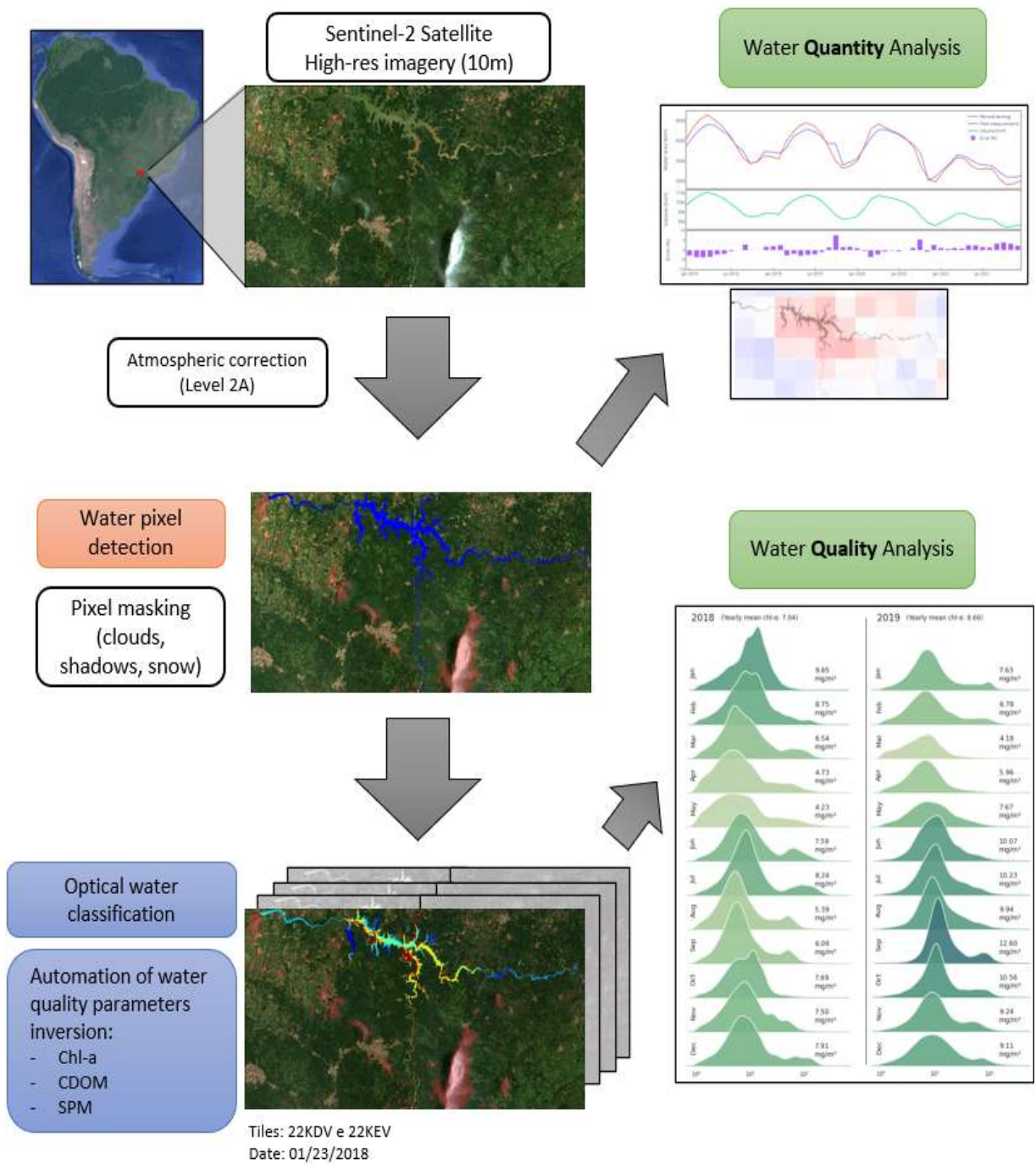


Figure 5: Schematic view of the steps involved in assessing inland water quality from satellite remote sensed imagery that were developed in this thesis. Boxes in orange represent steps covered in axis 1 (water detection), boxes in blue represent axis 2 (water classification) and green boxes represent axis 3 (generalization and analysis).

Axis 1 – Water Bodies Detection

The starting point for any water assessment through a remote sensing framework is correctly separating water pixels from other targets in the scene. During the atmospheric correction of the satellite images that converts them from Level-1C (top of atmosphere reflectance) to Level-2A (surface reflectance), a geophysical mask is typically produced containing land cover pixels classification. The most common Level-2A processors for the Sentinel-2 (MAJA, Sen2Cor, and F-mask) deliver the water mask as a subproduct of the atmospheric correction. However, as inland water is not their focus, the water masks provided by these processors are usually inaccurate, especially for complex inland and smaller water bodies, and should be employed only as a last resort (Cordeiro et al., 2021).

Apart from the standard processors, many studies have been published in the field, but current algorithms usually require human intervention, in the case of supervised learning methods that need training data for calibration, or they lack robustness to be applied in different conditions because of the diversity of optically complex waters (due to its constituents) and land covers.

In this context, the first axis focuses on the challenges involved in water detection from optical satellite imagery. The objective is to provide a methodology that can be applied automatically in different regions with diverse land covers and atmospheric conditions.

Besides robustness, water detection scalability is another scientific issue to be addressed in axis 1, as it can represent a significant constraint depending on the monitored area. As a reference, a single Sentinel-2 scene, which covers an area of 110km x 110km, has more than 100 million pixels, with 13 radiometric bands, and the classification is the first step for further analysis.

Other intervenient factors that can affect the correct water pixel identification are also assessed in this axis. For example, the atmospheric correction to be applied and the masking methods to eliminate undesired pixels such as clouds, shadows, or snow are also discussed in the context of axis 1.

Axis 2 – Water Color Classification

The assessment of water quality, notably suspended sediment matter (SPM), through remote sensing depends on the interaction that occurs between the OACs and the electromagnetic radiation within the visible and near-infrared regions (Barbosa et al., 2019). Mathematical inversion models are then developed, but it is generally accepted that no single model is applicable to all water types occurring across the globe (Kirk, 1994). Furthermore, results are highly coupled to the water types and the field measurements used for the calibration, and it is not always a simple task to define the applicability range of these algorithms (Xue et al., 2019). These issues directly affect the scalability of remote sensing for water resources assessments, as they impose constraints when delivering water quality products on large scales.

To overcome this difficulty, some studies have shown that algorithms can perform better when calibrated to specific ranges of Chl-*a* (Matsushita et al., 2015; Pahlevan et al., 2020; Smith et al., 2018) or SPM (Condé et al., 2019; Yopez et al., 2018). However, an additional difficulty must be considered for an automated regional or global scale assessment. Even if several algorithms are calibrated for different water types, it can be challenging to assure the applicability range and limitations of each model, so it is not known “*a priori*” which one is best suited for a particular place or time. Additionally, the same water body can simultaneously contain different water types (Moore et al., 2014; Neil et al., 2019).

In this regard, studies have shown that using water classification schemas and calibrating models for each defined optical water type (OWT) can contribute to an improvement of retrieval algorithms accuracy (Lubac and Loisel, 2007; Moore et al., 2014; Vantrepotte et al., 2012). However, although these papers can demonstrate an improvement in parameter retrieval if OWTs are considered, the inversion accuracy is not the final objective of the classification process, it is a secondary goal.

Considering this, the objective of the second axis is to propose a framework that can classify OWTs in order to maximize a retrieval's accuracy. Using machine learning strategies such as those employed in axis 1, like unsupervised agglomerative clustering, axis 2 proposes to separate spectral curves into water types, not by spatial distribution or river basin. The main idea is to explore the hyperspectral response of different rivers and combine them automatically into distinct OWTs, so the parameter retrieval is optimal.

A comprehensive database of more than 1,000 in situ radiometric measurements, coupled with other laboratory water quality measurements will be used. Due to data constraints, the methodology will be developed to retrieve Suspended Particle Matter (SPM), but the developed framework can be extended to other water quality parameters, subject to data availability.

Axis 3 – Generalization and Analysis

The challenges towards robustness and generalization are not constrained to this third axis. As already presented, the first two axes were also motivated by the need for more tools that permit operational usage of satellite remote sensing images for inland water assessment.

However, developing methods that can be applied in different regions is just a first step. Applying such methods to a regional scale and analyzing such amount of data adds up another layer of complexity due to volume constraints.

In this axis, concepts of Big Data, such as cloud computing and parallel processing, are explored, to enable the processing of thousands of images that would not be feasible with conventional tools.

Once these data are correctly processed, another scientific question arises and needs to be addressed. How should we analyze and extract meaningful information from this huge amount of data, and how should we present its results?

In this context, this third axis is devoted to exploring methodologies that can permit to scale up the current methods to assess inland water on a regional scale, discuss its shortcomings and propose alternatives. To accomplish this objective, a case study has been proposed to monitor by RS a large area of 320,000km² in Brazil's Paraná Basin. This region was selected for its national importance and because it faced a major drought in 2021. Water surface extents, as well as turbidity and Chl-*a* have been assessed from 2018 to 2021 using Sentinel-2 imagery, and their results were analyzed accordingly.

6. Thesis Structure

Considering the aforementioned objectives, this thesis is organized in three main chapters, as detailed below:

Chapter 1 - Water Detection

This chapter focuses on the challenge of distinguishing water pixels from other objects in satellite images. This chapter presents the paper “**Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors**” (Cordeiro et al., 2021) published in the Remote Sensing of Environment journal. The paper proposes a new unsupervised method that uses machine learning methods, notably

clustering and Naïve Bayes classification, to improve robustness and performance to separate water pixels in high-resolution optical scenes.

A second paper (Annex I), “**Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level**” (Peña-Luque et al., 2021), in which I participated as the third author, focused on further comparison of the proposed clustering methodology to existing ones and compared accuracies of separate optical and radar platforms. These comparisons were performed on regional and reservoirs level, and the effect of time-windowed estimations was also assessed.

Further developments on the same theme towards global generalization and resolution improvements have produced respectively the papers “**Surface water dynamics from space: a round robin inter-comparison of using Optical and SAR high-resolution (10-m) satellite observations for regional surface water detection**” (Tottrup et al., 2022), published in Remote Sensing – MDPI journal and “**Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, PlanetScope and Sharpened imagery [Submitted for publication]**”, submitted for publication in the GIScience & Remote Sensing journal. I participated in the elaboration of both papers with a focus on the same theme but within different environments and sensors, and they are presented in Annex II, and Annex III respectively.

Chapter 2 – Optical Water Classification

This chapter presents a framework to classify water types, according to their reflectance spectra, in order to maximize the SPM retrieval accuracy. It is entitled “Optical Water Classification for Suspended Particle Matter (SPM) retrieval” and uses unsupervised clustering

methods to separate the water spectra in groups. Several input features, normalizations methods and clustering algorithms are tested.

Several articles in the recent years have dealt with water optical class classification but they were mostly focused with chlorophyll-a mapping or for coastal waters. There is a need to address the suspended particulate matter that is the first driver of the water leaving radiance over inland waters as a function of erosion and sediment transport processes in rivers, lakes and reservoirs. In particular, we will seek the optimal number of classes allowing to achieve the best retrieval accuracy and for different spectral resolution. For this, we took benefit of an unprecedented database of hyperspectral field measurements over different rivers that allowed to represent contrasted catchments in terms of climate and geology and hydrological processes.

Chapter 3 – Generalization and Analysis: A case study of Brazil's 2021 drought.

The third chapter presents the methodologies developed towards the generalization and analysis of inland water quality parameters. The first part is an in-depth presentation of the methodologies developed in the current study to extrapolate the water detection and the quality parameters inversion to a regional scale and a long timeframe. In this chapter, automation and robustness are leveraged to a bigger scale (over 300 000 km²) to permit the processing of tens of thousands of satellite imagery without manual intervention.

This introductory section is split in two parts. The first part, “Water Detection Methodology” goes into the details of the improvements implemented over the **waterdetect** package that were not covered in the original “**Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors**” paper. The second part, “Water Assessment Methodology” concerns all the

postprocessing that has been performed in the resulting water masks and corresponding analysis steps.

Following the detailed presentation of the methodologies that are used within this chapter, a case study is presented in the form of an article, that employ the proposed methods to assess the effects of a major drought that occurred in the Brazilian's Paraná River basin in 2021, in terms of water quantity and water quality of thousands of inland water bodies.

Concerning water extension, all the region that was considered as in critical situation by the Brazilian authorities was assessed. An area of 320,000km² that includes 17 major reservoirs has been analyzed from January 2018 to December 2021. A total of 12,000 Sentinel-2 images with 10m of spatial resolution have been used to assess the effects of the drought on more than 20,000 reservoirs as small as 1 ha. Besides the analysis stratified by reservoirs' dimensions, that would not be feasible with field measurements, a spatial anomaly analysis is also performed, evidencing how the drought affected differently the regions of the basin.

For the water quality assessment, a sub-basin of the Paraná Basin, called Paranapanema basin has been selected due to availability of sun-glint corrected data (to be discussed further). Although smaller in area extension, the same time frame from January 2018 to December 2021 has been assessed. The water quality parameters turbidity and Chl-*a* have been estimated through inversion algorithms from the literature. Again, a stratified analysis has been performed to assess difference in these parameters in function of reservoirs' size and how they responded to the major drought.

7. Development context of the Thesis

The present work has been developed as part of the cooperation agreement "Hidrosat" between the French Institute de Recherche pour le Developpement (IRD), Géosciences

Environnement Toulouse (GET) laboratory, and the Brazilian National Water and Sanitation Agency (ANA), which financed my doctorate studies, to build up capacity in spatial hydrology. More particularly, my work regarding the use of Sentinel-2/MSI sensor imagery has been developed in the context of the OBS2CO project (acronym for OBServation de la Couleur des eaux Continentales) and of the French data center for continental surfaces (THEIA) that involves several institutions to provide access to the water sciences community to high-resolution water quality products over lakes, reservoirs, and rivers.

Besides IRD and ANA, other institutions such as the French National Center for Space Studies (CNES), the French Office for Biodiversity (OFB), Brasilia University (UNB), Federal University of Amazonas (UFAM) takes part into these initiatives. The final objective is to develop operational workflows to process remotely sensed time-series and make them available to the community. Among its several work fronts, we can highlight: 1) the development of optical and water quality databases; 2) inversion algorithms; 3) products validation; and 4) publishing platforms. The results obtained from this pipeline permit the development of research and practical applications. Trends in eutrophication, impacts of hydrological extreme events, erosion monitoring, sediment transport and deposition (siltation), and environmental catastrophes are some examples of subjects that can be addressed using the tools that are being developed in the thesis.

In this thesis, I worked closely with this multidisciplinary team and made significant improvements in terms of methodology towards large-scale processing and analysis.

Introduction (Français)

1. Les enjeux mondiaux liés à l'eau

L'eau est la ressource naturelle la plus importante sur Terre, car elle est nécessaire au maintien de toutes les formes de vie. En outre, elle est vitale pour le développement humain, social et économique, la production de nourriture et d'énergie, et le maintien d'un environnement sain. Malheureusement, bien que plus de 70 % de la surface du globe soit recouverte d'eau, la majeure partie de ce total est constituée d'eau salée dans les océans (98 %). L'eau douce restante (2%) est principalement stockée dans les calottes glaciaires ou dans des aquifères souterrains profonds, ce qui ne laisse qu'une infime partie du total (0,024%) à la surface disponible pour la consommation humaine (Gleick et al., 1993).

Les données montrent que les prélèvements mondiaux d'eau douce ont été multipliés par six au cours des 100 dernières années, pour atteindre plus de 4 000 km³/an. L'agriculture (69 %) et l'industrie (19 %) sont les activités les plus consommatrices des ressources en eaux. Une grande partie de cette augmentation est attribuée à la combinaison de la croissance démographique, du développement économique et de l'évolution des modes de consommation (UNESCO-WWAP, 2021).

La population mondiale, qui est actuellement de 7,6 milliards d'habitants, devrait atteindre 10 milliards d'ici 2050 et 11 milliards d'ici 2100, selon les projections de l'ONU (UNDP, 2020). Outre la croissance démographique, l'élévation de l'espérance de vie, perceptible par l'augmentation de l'indice de développement humain (HDI en anglais) au fil des ans, se traduit également par une pression accrue sur les ressources naturelles, principalement l'eau douce (UNDP, 2022a).

De plus, la population urbaine est passée de 2,3 milliards de personnes en 1990 à 4,0 milliards en 2015, ce qui représente 51% de la population totale, et cette concentration devrait s'accroître dans les prochaines années, pour atteindre jusqu'à 68% en 2050, selon le rapport *World Urbanization Prospects* (United Nations, 2018). Dans ce contexte, il est également crucial de comprendre la répartition spatiale des ressources en eau en termes de qualité et de quantité, car ce mouvement d'urbanisation exerce une pression encore plus forte sur des zones restreintes et à forte densité et donc sujettes à d'épisodes de restrictions d'usages de l'eau.

Bien que différentes études aient projeté des tendances différentes dans l'utilisation future de l'eau, et que l'augmentation réelle de l'utilisation mondiale de l'eau reste incertaine, il est un fait qu'une fraction croissante de la population sera confrontée à des pénuries d'eau dans les années à venir (UNESCO-WWAP, 2021). En outre, les changements climatiques peuvent également affecter les processus d'évaporation et de précipitations en raison de l'augmentation de la température, renforçant des événements extrêmes liés à l'eau, tels que des crues et des sécheresses dans le monde entier.

Outre la pression supplémentaire exercée sur les ressources en eau par les tendances évoquées ci-dessus, la qualité de l'eau s'est également détériorée en raison de la présence de divers contaminants dans les ressources en eau souterraine et de surface, induite par l'intensification des activités humaines. La moitié des rivières et des lacs du monde sont pollués, selon Makarigakis et Jimenez-Cisneros (2019), et cette situation peut également être largement dégradée par les changements climatiques.

Pour faire face à ces défis et inscrire la gestion durable des ressources en eau dans l'agenda politique mondial, l'ONU a consacré l'un de ses Objectifs de Développement Durable (SGD en

anglais), également appelés Objectifs Globaux, de l'Agenda 2030, à assurer la disponibilité et la gestion durable de l'eau et de l'assainissement pour tous (6ème objectif) (UNDP, 2022b).

Il faut notamment tenir compte du fait que la plupart des problèmes liés à l'eau dans le monde ne sont pas dus à un manque réel d'eau. Ils résultent plutôt d'une gouvernance défaillante et inefficace des ressources en eau (UNESCO-iWSSM, 2019). Dans ce contexte, la surveillance de l'eau est essentielle pour comprendre ses changements en termes de quantité et de qualité. Le rapport *United Nations World Water Development* (UNESCO-WWAP, 2021) sur la mise en valeur des ressources globales en eau affirme que pour parvenir à une gestion durable et à long terme des ressources en eau, il est essentiel d'intégrer de meilleures évaluations et un meilleur suivi pour les processus décisionnels des gestionnaires.

En outre, la surveillance périodique des masses d'eau et la gestion intégrée des bassins hydrographiques sont essentielles à la conservation de ces ressources. Cependant, il existe de nombreux défis à relever en matière de surveillance sur le terrain, en particulier pour les pays en voie de développement, tels que :

1) Le temps et les coûts nécessaires pour effectuer des mesures sur le terrain et des analyses en laboratoire ;

2) Les échantillons ponctuels ne permettent pas toujours de saisir les variations spatiales et temporelles ; et

3) La surveillance continue sur le terrain n'est pas réalisable dans les régions éloignées et difficiles d'accès.

Par conséquent, l'utilisation de sources de données alternatives provenant de capteurs de télédétection, en particulier de satellites orbitaux, a rapidement augmenté au cours des dernières

décennies et appuie la communauté scientifique à mieux comprendre l'évolution des ressources en eaux. En outre, ces technologies ont montré qu'elles pouvaient compléter ou remplacer les inventaires traditionnels au sol et offrir des avantages supplémentaires tels qu'un meilleur rapport coût-efficacité, une fréquence plus élevée et une meilleure continuité spatiale (Barbosa et al., 2019).

2. Télédétection

Selon Kairu (1982), le terme télédétection peut être défini comme « la collecte et l'interprétation d'informations sur un objet, une région ou un événement sans être en contact physique avec l'objet ». Une simple photographie suivie de l'analyse de son contenu est alors considérée comme une étude de télédétection. Les capteurs peuvent être montés sur n'importe quelle plateforme, comme un avion ou un ballon de recherche. La télédétection spatiale pour l'observation de la terre n'est pas un concept nouveau. Elle a effectivement débuté avec le lancement de LANDSAT-1 en 1972 par la *National Aeronautics and Space Administration* (NASA), suivi par le Centre national d'études spatiales (CNES) avec le lancement de SPOT-1 en 1986.

Bien que des progrès considérables aient été réalisés depuis lors dans le domaine de la télédétection pour l'observation de la terre, le potentiel de l'imagerie satellitaire reste sous-exploité dans de nombreux domaines spécialisés, notamment par les acteurs publics et les décideurs chargés d'appliquer les politiques publiques. Au début de cet ère satellitaire, le coût des images et le manque d'experts ayant les connaissances appropriées pour traiter et analyser ces données, étaient quelques-unes des causes de cette sous-exploitation des données de télédétection (Baghdadi et Zribi, 2016).

Cependant, la capacité de la télédétection à étudier la surface terrestre et les conditions environnementales sur de vastes zones a augmenté au cours des dernières années, principalement grâce au lancement de plusieurs satellites d'observation de la terre qui fournissent des données peu

coûteuses sur l'ensemble du globe. De plus, les universités ont adapté leur programme d'études et des formations distinctes, telles que la géologie, la géographie, l'écologie et d'autres, inégalement désormais des axes liés à la télédétection.

En termes pratiques, l'information collectée par un capteur correspond à l'énergie émise par l'objet, fréquemment appelé cible, dans différentes parties du spectre électromagnétique (EM en anglais). La région d'intérêt du spectre EM en télédétection va des courtes longueurs d'onde (partie visible), qui correspond au spectre EM que l'œil humain peut voir, généralement placé entre 380 et 700nm, à l'infrarouge et aux plus grandes longueurs d'onde (micro-ondes).

La radiation émise ou réfléchiée par les différentes cibles sur Terre va dépendre directement de l'énergie entrante et de l'absorption/réflexion de cette énergie par la cible. Cette interaction est spécifique à chaque type de matériau (Figure 7a) et à ses propriétés physico-chimiques. Il existe deux types de capteurs, en fonction de la source d'énergie. Les capteurs passifs dépendent de sources d'énergie externes, généralement le soleil, et les capteurs actifs ont leur propre source d'énergie, comme dans le cas des radars.

Lorsque l'énergie descendante atteint la cible, une partie de cette énergie peut être absorbée par l'objet. Cette absorption varie selon la gamme de longueurs d'onde du spectre électromagnétique et la partie de l'énergie non absorbée est diffusée. L'énergie rétrodiffusée est mesurée par le capteur dans différentes régions de longueurs d'onde, définies comme les bandes spectrales du capteur. La relation entre la radiance ascendant (mesuré par le capteur) et l'énergie descendante qui touche la cible est appelée réflectance et peut être estimée pour chaque longueur d'onde à travers le spectre. Si l'on considère qu'il n'y a qu'une seule source d'énergie et que les processus d'absorption et de diffusion sont fonction du matériau, chaque cible aura un modèle spectral unique, appelé signature spectrale (Figure 7b).

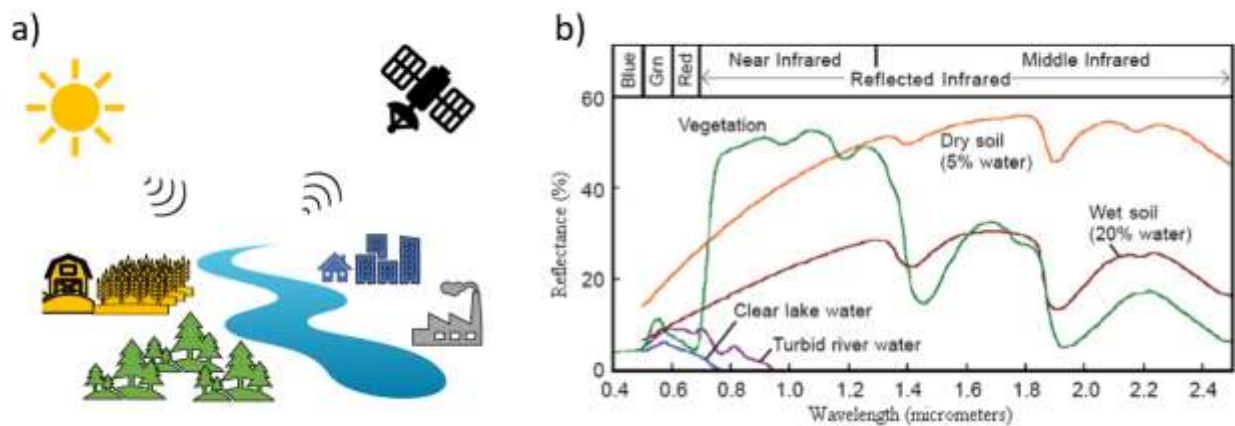


Figure 7: (a) Principe général de la télédétection passive, (b) Signatures spectrales de différentes cibles naturelles. Extrait de Mercan et Alam (2011).

De nombreux autres facteurs peuvent interférer dans les mesures à distance, comme les interactions de la lumière descendante et ascendante avec les particules de l'atmosphère et la position relative du soleil. Une présentation détaillée de ces concepts physiques dépasse le cadre de la présente thèse mais peut être consultée dans Guo et al. (2020), Kairu (1982), et Baghdadi et Zribi (2016).

L'observation de ces caractéristiques de réflectance à travers l'espace et le temps permet aux scientifiques et chercheurs de modéliser différents phénomènes et de mieux comprendre et caractériser notre environnement (Baghdadi et Zribi, 2016).

2.1 La télédétection appliquée aux ressources hydriques

Les plateformes satellitaires sont utilisées depuis plus de 50 ans pour comprendre des processus complexes tels que l'utilisation des sols à l'échelle mondiale, l'analyse de la végétation, la productivité des océans ou pour établir des liens entre la production industrielle et le cycle du carbone (Topp et al., 2020).

L'accent mis sur la compréhension de ces processus en échelle globale s'explique par les résolutions spatiales, spectrales et temporelles fournies par les plateformes antérieures. À titre de

référence, le LANDSAT-1, lancée en 1972, a été le premier satellite conçu spécifiquement pour des applications d'observation de la Terre et il était chargé d'un système de balayage multispectral (MSS en anglais) avec 83 m de résolution spatiale et seulement 4 bandes spectrales (visible et proche infrarouge) avec un temps de revisite de 18 jours. D'autres anciens satellites océaniques et atmosphériques avaient des résolutions spatiales encore plus rudimentaires mais une fréquence d'observation élevée.

En ce qui concerne l'utilisation d'images de télédétection pour l'évaluation des eaux intérieures, Topp et al. (2020) affirment que, malgré le nombre croissant de publications, les produits globaux concernant la qualité de ces cours d'eau, lacs et réservoirs sont limités par rapport aux produits terrestres et océaniques. Cette lente évolution peut s'expliquer en partie par les défis liés à la télédétection de masses d'eau continentales ainsi que par la capacité limitée des capteurs à détecter les petites masses d'eau et les interactions complexes de leurs constituants optiquement actifs (OAC en anglais). Les OAC sont des constituants dissous et particulaires de l'eau qui peuvent absorber et diffuser la lumière (par exemple, les matières en suspension, la chlorophylle, la matière organique dissoute colorée et autres), donnant à l'eau sa couleur caractéristique.

Un exemple de cette complexité est illustré par la Figure 8, qui montre la réponse spectrale moyenne de plusieurs rivières considérées dans la présente étude, mesurée par des radiomètres hyperspectraux de terrain, et comment cette réflectance peut varier en fonction de ses constituants, de la région géographique, etc. Ces observations contrastent avec l'hypothèse générale d'une radiance nulle à la sortie de l'eau dans les longueurs d'onde du proche infrarouge (NIR en anglais), qui est communément adaptée aux eaux océaniques claires (Les eaux du cas 1).

Heureusement, ces dernières années, la capacité des satellites d'observation terrestre a fortement augmenté. À cet égard, il convient de noter le programme Copernicus, conçu par

l'European Space Agency (ESA), dont les missions portent sur différents aspects de l'observation de la Terre : Atmosphérique, Océanique et Surveillance des terres (Suhet et Hoersch, 2015).

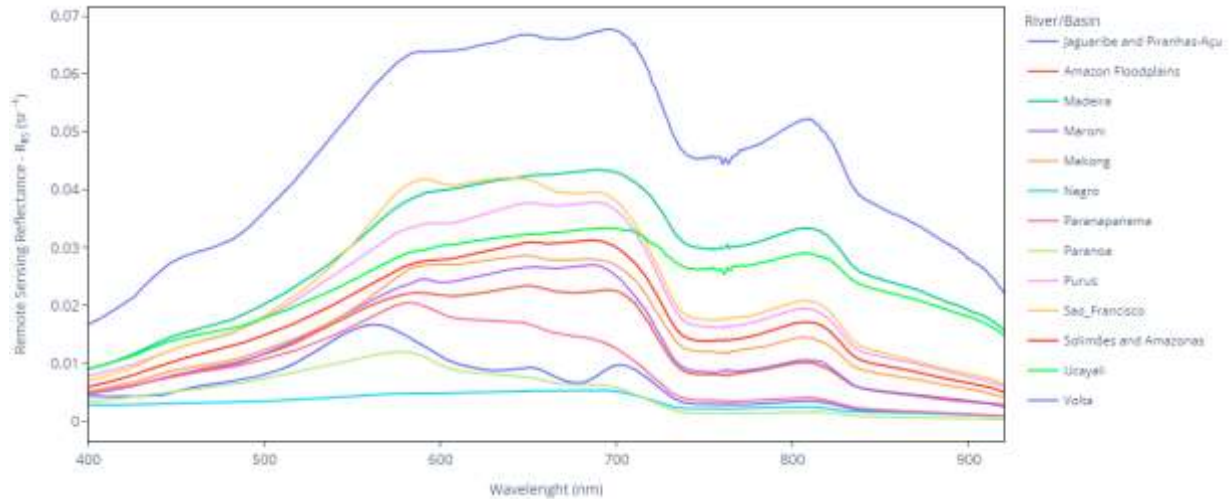


Figure 6: Les profils spectraux moyens de plusieurs rivières/bassins considérés dans la présente étude mesurés avec des radiomètres de terrain.

Les satellites de la mission Sentinel-2, qui fait partie du programme d'observation Copernicus, sont équipés d'un instrument optique multispectral (MSI en anglais) qui offrent de hautes résolutions spatiale, spectrale et temporelle à l'échelle mondiale sans coût pour l'utilisateur. Le premier satellite de la famille Sentinel-2 (Sentinel-2A), a été lancé par l'ESA en 2015, suivi de son satellite jumeau (Sentinel-2B) qui a été lancé en 2017. Tous deux sont des satellites en orbite polaire et fonctionnent en phase à 180 degrés l'un par rapport à l'autre pour offrir une fréquence élevée de revisite de 2 à 3 jours aux latitudes moyennes et de 5 jours près de la ligne équatoriale (Suhet et Hoersch, 2015). Le Sentinel-2 a été choisi pour traiter les questions de qualité des eaux intérieures, ce qui est le champ d'application de cette thèse, car il fournit une résolution compatible avec la détection des canaux de rivière et des plus petits plans d'eau qui ne seraient pas visibles avec d'autres capteurs de résolution moyenne.

De plus, il peut capturer une large gamme de longueurs d'onde spectrales séparées en 13 bandes spectrales allant de 443 à 2200 nm, et des résolutions spatiales allant de 10 à 60m selon la bande spectrale. Une autre information qui doit être prise en compte pour étudier les paramètres de qualité de l'eau est la résolution spectrale car elle définit la capacité à identifier la présence de différents paramètres de qualité des eaux (matières organique dissoute, pigments photosynthétique, particules en suspension). Le capteur MSI de Sentinel-2 a des largeurs de bande allant de 15 à 180 nm, ce qui permet d'identifier et quantifier les différents constituants de l'eau optiquement actifs. Cependant, il est important de noter que les méthodologies développées ci-après ne sont pas spécifiques à un satellite et pourraient être applicables à d'autres plateformes ayant des spécifications techniques différentes. La Figure 9 présente les fonctions de réponse spectrale pour les 13 bandes disponibles dans le capteur MSI de Sentinel-2 et leurs résolutions spatiales correspondantes.

Cette capacité étendue fournie par les satellites les plus récents permet aux scientifiques d'explorer différentes possibilités concernant le domaine des eaux intérieures, notamment la couverture des eaux de surface (Buma et al., 2018 ; Cordeiro et al., 2021 ; Feyisa et al., 2014 ; Markert et al., 2018 ; Pekel et al., 2016 ; Peña-Luque et al., 2021 ; Souza et al., 2019), la cartographie des crues et des inondations (Kordelas et al, 2019, 2018 ; Martinis et al., 2011 ; Wieland et Martinis, 2019), la surveillance de la qualité de l'eau (Ansper et Alikas, 2018 ; Delegido et al., 2014 ; Frampton et al., 2013 ; Lins et al., 2017 ; Toming et al, 2016 ; Udeberg et al., 2020 ; Yadav et al., 2019), l'évaluation des sédiments en suspension (Condé et al., 2019 ; Espinoza Villar et al., 2013 ; Martinez et al., 2009 ; Yopez et al., 2018), entre autres.

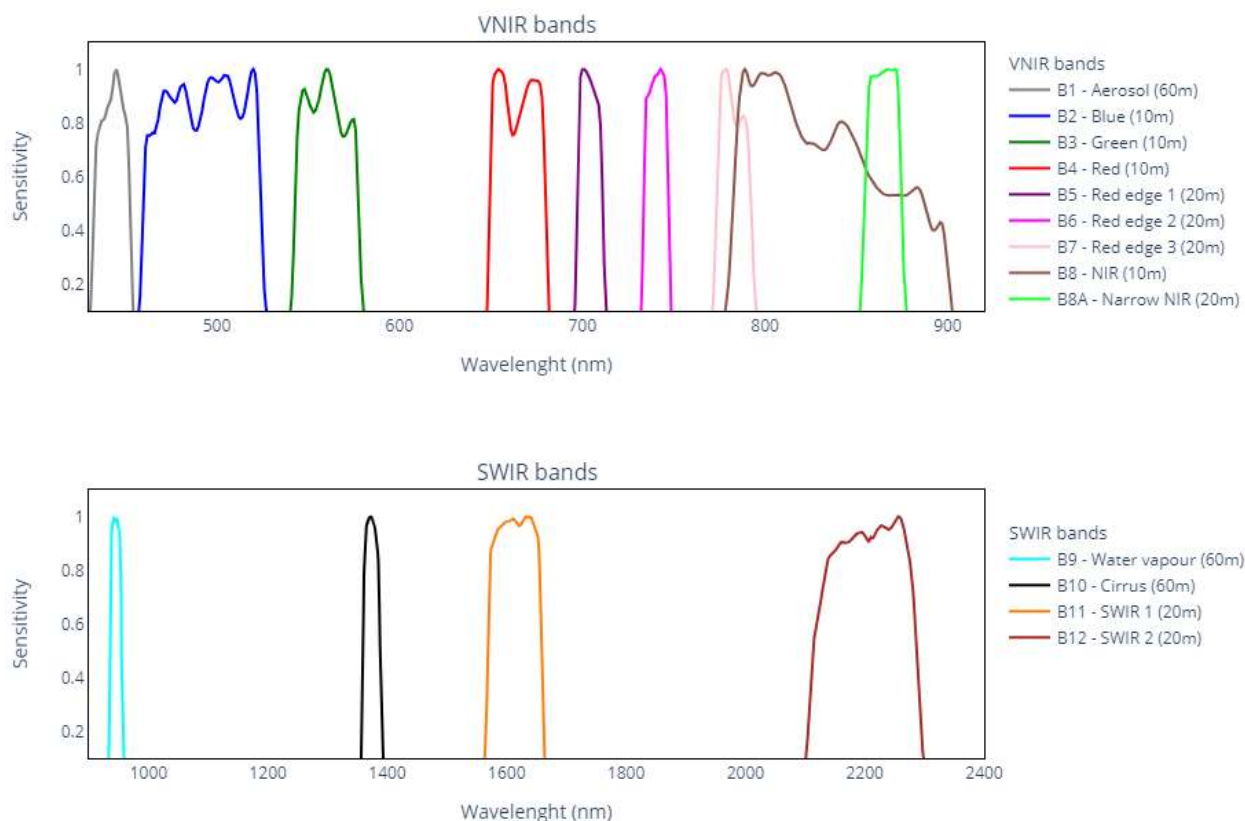


Figure 7: Fonctions de réponse spectrale pour les bandes Sentinel-2/MSI, Visual-NIR et SWIR et leurs résolutions spatiales correspondantes.

Cependant, comme nous l'avons déjà mentionné, ce nombre croissant de publications scientifiques dans le domaine des ressources en eau ne se traduit pas par des produits opérationnels à l'échelle régionale ou mondiale, pourtant cruciaux pour les décideurs (Topp et al., 2020).

Une exception qui mérite d'être mentionnée est le projet *Global Surface Water* (GSW) (Pekel et al., 2016), qui a été développé pour évaluer les surfaces d'eau dans le monde entier et suivre leurs changements à long terme, en utilisant l'imagerie optique Landsat-8 avec une résolution de 30m. Cependant, l'objectif du projet GSW est de suivre les changements mensuels à long terme, mais pour effectuer une évaluation opérationnelle de la qualité de l'eau grâce aux images de télédétection, des résolutions spatiales et temporelles plus élevées sont nécessaires. Ainsi, le groupe de données GSW ne peut pas être utilisé pour une surveillance continue.

Parmi les obstacles qui limitent l'évaluation de la qualité des eaux intérieures par télédétection satellitaire, Chami et al. (2015) soulignent la difficulté de contrôler les altérations atmosphériques, par essence très variables. Les procédures de correction atmosphériques conventionnelles pour les eaux océaniques reposent souvent sur l'hypothèse d'une radiance montante nulle de l'eau au-delà du spectre visible. Cette hypothèse ne peut pas être généralisée aux eaux intérieures optiquement complexes, car la présence de matières inorganiques et organiques augmente les processus de rétrodiffusion de la lumière dans le spectre NIR (Figure 8).

De plus, les capteurs optiques sont soumis aux conditions atmosphériques telles que les nuages, les ombres des nuages et d'autres phénomènes de surface qui peuvent interférer dans la détection des pixels d'eau, comme la neige, par exemple. Ainsi, outre une correction atmosphérique qui tient compte de la présence d'eau optiquement complexe, un autre élément clé pour fournir des produits opérationnels est l'application de méthodes de masquage robustes capables d'identifier correctement (et éventuellement d'éliminer) les éléments indésirables de la scène observée par le satellite.

3. Les outils de science des données pour la télédétection

Dans la section précédente, nous avons largement discuté des utilisations des images de télédétection par satellite pour l'observation de la terre et de la façon dont elles peuvent être un outil précieux pour évaluer l'état physique, chimique et biologique des masses d'eau intérieures, malgré les défis que cela représente.

Au début, la télédétection était principalement basée sur la photographie optique (limitée entre la gamme spectrale visible et infrarouge) capturée à partir de plateformes aériennes. La photo-interprétation de ces images était réalisée par des techniciens qualifiés chargés d'identifier les éléments et les motifs sans l'aide de processus automatisés (Baghdadi et Zribi, 2016). Cependant,

ce processus prend beaucoup de temps et est irréalisable compte tenu de la quantité de données disponibles ces dernières années. De plus, la télédétection a évolué, passant d'un mode principalement optique à l'utilisation de l'ensemble du spectre des longueurs d'onde, du visible aux micro-ondes, transformant les images en cubes à nombreuses dimensions difficiles à interpréter manuellement.

Dans ce contexte, des progrès considérables ont été réalisés dans le développement de méthodes assistées par ordinateur pour le traitement et l'analyse des données de télédétection, et les outils de science des données tels que le *Data Mining*, *Machine Learning* (ML) ou *Deep Learning* (DL) jouent désormais un rôle central.

3.1 Machine Learning

Le ML est un sous-domaine du domaine de l'intelligence artificielle, avec le DL. Le terme a été inventé en 1959 par Arthur Samuel, un pionnier américain dans les domaines des jeux vidéo et de l'intelligence artificielle, et décrit la capacité de faire des prédictions ou de prendre des décisions sans être explicitement programmé pour le faire (Mitchell, 1997).

Pour comprendre ce qu'est la ML, on peut penser à un problème qui doit être résolu par un programme. En restant dans le contexte de la télédétection, imaginons, par exemple, que nous devions développer un programme capable d'identifier les nuages sur une image satellite. L'approche la plus simple serait de cartographier, à l'avance, toutes les caractéristiques d'un nuage (par exemple, la couleur, la température, l'intensité de la réflexion, entre autres) et la façon dont ces caractéristiques se traduisent en valeurs de réflectance de bande et d'écrire des clauses pour vérifier si chaque pixel se trouve dans ces règles. C'est ce qu'on appelle l'approche basée sur les règles. Cependant, compte tenu de la diversité des nuages et de leur interaction avec le sol (par

exemple, un cirrus semi-transparent), il peut être fastidieux, voire irréalisable, de cartographier toutes les valeurs possibles pour programmer cet algorithme avec précision.

Dans l'approche ML, au lieu de cartographier toutes les valeurs possibles, on présente à l'algorithme ML des données préalablement étiquetées comme nuage ou non-nuage par un humain, et l'ordinateur "apprend" à partir des données fournies comment distinguer les nuages. Cette méthode, où les données sont préalablement étiquetées, est appelée apprentissage supervisé. D'autres approches, dans lesquelles aucune étiquette n'est fournie, sont également disponibles et sont appelées apprentissage non supervisé (par exemple, le regroupement et la réduction de la dimensionnalité) et seront abordées plus loin. Les différences fondamentales entre la programmation basée sur des règles et la programmation ML sont résumées dans la Figure 10.

L'un des avantages des méthodes ML par rapport aux approches traditionnelles est qu'elles ne nécessitent pas de connaissances préalables concernant les propriétés et les statistiques intrinsèques des données. Au lieu de cela, la relation entre les variables d'entrée (également appelées caractéristiques d'entrée) est obtenue automatiquement au cours du processus de formation, où les paramètres internes sont calibrés pour améliorer la précision.

Les techniques basées sur les ML se sont avérées utiles pour un très grand nombre d'applications, telles que la classification de la couverture terrestre, la détection d'objets, les algorithmes de récupération empirique, entre autres et des domaines distincts d'observation de la terre (par exemple, la terre, l'océan et l'atmosphère) (Lary et al., 2016). Par conséquent, les méthodes ML sont de plus en plus utilisées pour interpréter toutes sortes d'images de télédétection.

Pour extraire des informations pertinentes des images multispectrales, les techniques ML peuvent être divisées en deux groupes principaux : 1) l'apprentissage supervisé ; et 2) l'apprentissage non supervisé.

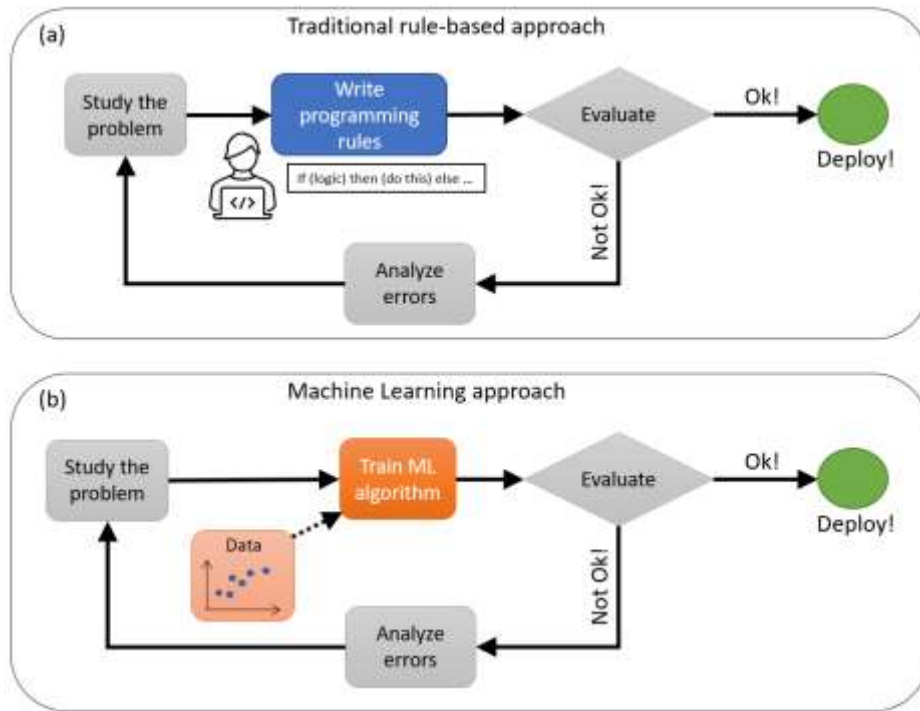


Figure 8: Comparaison schématique de l'approche de programmation traditionnelle (a) et de l'approche d'apprentissage automatique (b). Adapté de Géron (2017).

L'apprentissage supervisé

Dans le contexte de l'apprentissage supervisé, le but est d'estimer des variables (des classes qualitatives, dans le cas de problèmes de classification, ou des valeurs continues dans les problèmes de régression) à partir d'échantillons dans un ensemble de données, étant donné que certaines de ces variables sont fournies pour un sous-ensemble appelé ensemble de données d'entraînement. Sur la base de l'ensemble de données d'apprentissage et selon l'algorithme supervisé utilisé, l'ordinateur peut établir en interne une relation entre les données et les variables souhaitées. Une fois le modèle formé (c'est-à-dire les paramètres internes calibrés pendant la phase de formation ou d'apprentissage), il peut être utilisé pour estimer les variables pour tous les échantillons de l'ensemble de données, même si nous ne pouvons pas comprendre entièrement la relation interne (Baghdadi et Zribi, 2016).

En raison de ces caractéristiques, le ML est généralement utilisé dans des domaines tels que les applications difficiles à programmer ou lorsque nos connaissances théoriques sont encore incomplètes mais pour lesquelles nous disposons d'un nombre important d'observations et d'autres données (Lary et al., 2016).

Dans l'exemple présenté précédemment, concernant la détection des nuages, l'algorithme peut établir une relation en considérant autant de variables d'entrée que nous le souhaitons. Par exemple, nous pourrions fournir la totalité des bandes multispectrales et la date d'acquisition. Les variables qui sont transmises à l'algorithme ML pendant le processus d'apprentissage sont appelées caractéristiques d'entrée, et le programme les utilisera pour prédire les sorties des nouveaux échantillons.

L'application de détection des nuages est un exemple de classification, car le résultat attendu est qualitatif (nuage ou non nuage). La même approche pourrait être utilisée pour récupérer une sortie quantitative, comme la température du sol ou l'évapotranspiration, par exemple. Dans ce dernier cas, nous considérons qu'il s'agit d'une application de régression. Plusieurs méthodes existent dans le contexte de l'apprentissage supervisé. Parmi elles, les plus utilisées en télédétection sont : *Support Vector Machines (SVM)*, *Decision Trees (DT)*, *Random Forests (RF)*, *Artificial Neural Networks (ANN)*, *K-Nearest Neighbors (KNN)*, entre autres (Lary et al., 2016). L'explication détaillée de ces méthodes sort du cadre de la présente introduction, cependant des descriptions complètes sont disponibles dans la littérature (Géron, 2017; Vanderplas, 2016).

L'apprentissage non supervisé

Les techniques non supervisées sont couramment utilisées lorsque nous n'avons aucune connaissance préalable des données analysées (par exemple, aucune étiquette) et que nous voulons que l'ordinateur en extraie des corrélations significatives. Par exemple, supposons que nous

vouliions identifier combien de cultures différentes apparaissent dans une scène donnée. Les algorithmes de partition, également appelés algorithmes de clustering, peuvent être utilisés pour regrouper automatiquement les données en sous-ensembles sans intervention ni données d'entraînement. Plusieurs algorithmes de clustering existent dans la littérature, et les plus

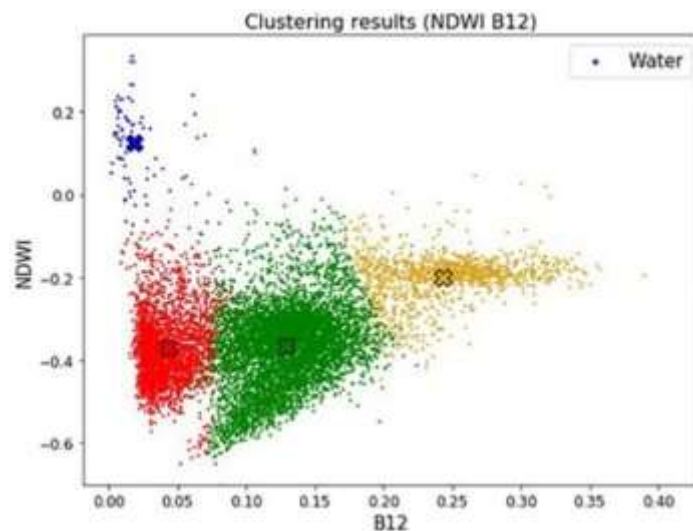


Figure 9: Diagramme de dispersion de l'indice de différence normalisé de l'eau (NDWI, axe Y) et de la réflectance SWIR (axe X) en sr⁻¹ pour les pixels d'une scène Sentinel-2, montrant les résultats du regroupement avec différentes couleurs.

couramment utilisés dans le domaine de la télédétection sont : k-Means, clustering agglomératif et DBScan. La Figure 11 montre un exemple de clustering agglomératif appliqué à plusieurs pixels dans une scène Sentinel-2 pour séparer automatiquement différentes classes de pixels. Ce type d'application sera étudié plus en détail au chapitre 1 - Détection de l'eau.

Un autre groupe de techniques qui appartient à l'apprentissage non supervisé est appelé méthodes de visualisation ou de réduction de la dimensionnalité. Dans la réduction de la dimensionnalité, la tâche consiste à simplifier les données en préservant autant d'informations que possible. Cela peut se faire en éliminant les caractéristiques d'entrée corrélées ou en fusionnant plusieurs caractéristiques en une nouvelle caractéristique synthétique. Ce processus de modification des variables d'entrée pour obtenir de meilleurs résultats est appelé extraction de

caractéristiques ou ingénierie des caractéristiques et constitue une étape clé des applications d'apprentissage automatique (Géron, 2017).

3.2 Deep Learning

Dans la section précédente, nous avons mentionné les ANN comme l'un des outils d'apprentissage supervisé avec le ML. Bien que l'application pratique des ANN soit récente, leur conception théorique remonte à 1943, lorsque McCulloch et Walter Pitts, respectivement neurophysiologiste et cybernéticien, ont proposé le premier modèle mathématique de réseau neuronal (Abrahart et al., 2004). Un ANN est formé d'unités de traitement simples appelées perceptrons, inspirées des neurones humains qui se connectent à d'autres, créant ainsi un réseau qui imiterait le cerveau.

Cependant, en raison des difficultés d'entraînement d'un ANN, son utilisation est restée longtemps négligée par la communauté scientifique. Plus récemment, avec l'augmentation de la puissance de calcul, notamment grâce aux processeurs graphiques (GPU en anglais), la disponibilité des données d'entraînement et certaines améliorations techniques dans l'étape d'entraînement (par exemple, la rétropropagation), les ANN ont reçu beaucoup d'attention.

Dans ce contexte, nous définissons les modèles DL comme des modèles ANN composés de plusieurs couches empilées, généralement plus de trois. Plus il y a de couches empilées, plus le modèle devient profond, et des interactions plus complexes peuvent être apprises, d'où le nom d'apprentissage profond. Ces architectures profondes se sont avérées efficaces pour extraire des informations de concepts abstraits et ont surpassé les méthodes classiques de ML dans un certain nombre d'applications telles que la reconnaissance vocale et la reconnaissance d'images (Hoeser et Kuenzer, 2020).

Selon Hoeser et Kuenzer (2020), l'utilisation de la DL par la communauté de l'observation de la Terre date de 2012. Depuis lors, le nombre de publications a doublé chaque année avec un large éventail d'applications, telles que la super-résolution, la fusion de données, le débruitage, les prévisions météorologiques, la reconnaissance de scènes, la classification, la détection d'objets et la segmentation d'images.

4. Généralisation et outils opérationnels

Bien que la télédétection optique visible multispectrale soit désormais bien connue des scientifiques et des acteurs en charge de la gestion de l'environnement, rassembler toutes ces pièces et les rendre suffisamment générales pour fonctionner dans différentes régions aux couvertures terrestres et conditions atmosphériques diverses reste un défi pour le domaine des eaux intérieures. En outre, nous sommes confrontés à une augmentation radicale du volume de données (*Big Data*) acquises par des capteurs à plus haute résolution, ce qui pose un nouveau problème pour le traitement des données satellitaires.

Dans le cadre de la détection globale des eaux de surface, un effort vers cette généralisation a été récemment proposé par le projet *World Water* (Tottrup et al., 2022) d'intercomparaison interdisciplinaire qui a été mené dans le but d'évaluer plusieurs méthodologies optiques et radar visant à fournir une détection régionale haute résolution des surfaces en eaux dans différentes régions difficiles du globe.

Une autre difficulté vers la généralisation concerne le développement d'algorithmes de surveillance globale par satellite qui permettent d'évaluer les paramètres de qualité de l'eau indépendamment de la masse d'eau surveillée. Dans la recherche océanographique, la fourniture de cartes synoptiques de la concentration globale de chlorophylle-a (*Chl-a*) à partir de capteurs radiométriques montés sur des satellites est une réalité depuis des décennies. En ce qui concerne

les eaux intérieures, même si de nombreux modèles différents ont été proposés dans la littérature pour différents paramètres, la complexité optique de ces eaux limite souvent leur utilisation opérationnelle à des échelles localisées (Neil et al., 2019).

À côté de cela, ces dernières années, des quantités massives de données satellitaires ont été acquises, augmentant considérablement notre capacité à comprendre notre environnement d'un point de vue spatial et temporel. Cependant, ce volume de données sans précédent (le volume d'images de télédétection recueillies par des satellites ou des drones peut facilement atteindre des niveaux de téraoctets ou de pétaoctets) a mis en lumière les défis et les goulots d'étranglement liés au stockage, au traitement, à l'analyse et finalement à la visualisation de toutes ces informations (Guo et al., 2020).

Cela signifie que les méthodologies développées localement ne peuvent pas être appliquées à plus grande échelle, car les contraintes de stockage, de mémoire et de traitement les empêcheront de fonctionner correctement ou ne produiront pas les résultats escomptés dans les délais prévus, ce qui les rendra obsolètes. C'est une autre raison pour laquelle la plupart des produits et publications sur la qualité des eaux intérieures sont généralement limités à des zones plus petites ou à des masses d'eau spécifiques, ce qui les rend difficiles à appliquer à des échelles régionales ou mondiales pour nous aider à comprendre le contexte spatial de ces interactions complexes.

Dans ce contexte, de nouveaux paradigmes informatiques développés dans d'autres disciplines telles que *Data Mining*, *Grid Computing*, *Cloud Computing* and *High Performance Computing* (HPC) sont incorporés à l'ensemble des outils de télédétection optique afin de fournir un support de traitement pour relever ces défis (Guo et al., 2020).

Ces approches sont nécessaires en fonction des tâches effectuées, car elles ouvrent la possibilité de traiter des quantités de données inimaginables par une seule instance de traitement.

Cependant, être capable de traiter ces énormes quantités de données n'est qu'une étape. L'image satellite reste un produit complexe, il est donc fondamental de développer des méthodologies permettant d'étendre l'analyse à une échelle spatiale et temporelle supérieure.

5. Objectifs

La télédétection peut être un outil efficace pour permettre une surveillance à grande échelle afin de traiter les problèmes de quantité et de qualité de l'eau. Cependant, la création d'un cadre opérationnel permettant d'évaluer la qualité des eaux intérieures à partir de la télédétection par satellite implique de nombreuses étapes et la sélection minutieuse de plusieurs paramètres, afin que le tout puisse fonctionner de manière intégrée. De la sélection de la plateforme satellitaire à l'identification du modèle de correction atmosphérique et au développement d'un algorithme d'inversion robuste, toutes ces étapes doivent être orchestrées de manière pertinente, comme le montre la Figure 12.

Dans ce contexte, cette thèse propose de travailler sur différents aspects de cette chaîne, en utilisant des méthodologies existantes et en proposant de nouveaux avancements vers la robustesse, l'échelle et l'automatisation, par l'application intensive de techniques de science des données.

Par conséquent, l'objectif global est de fournir un cadre complet qui peut être reproduit dans différentes régions du globe pour les études de la qualité des eaux intérieures grâce à l'imagerie de télédétection. Parmi une diversité de sujets impliqués, les aspects à explorer en détail ont été délibérément sélectionnés au cours du développement de la thèse et ont été séparés en trois axes principaux énumérés ci-dessous :

1. Détection automatique et non supervisée des plans d'eau ;
2. Amélioration de la classification des types d'eau optique pour la récupération des paramètres des eaux intérieures ; et

3. Surveillance à grande échelle par satellite et analyse des tendances spatio-temporelles des surfaces en eaux et de leur qualité.

Les objectifs de chaque axe défini sont détaillés dans les sections suivantes.

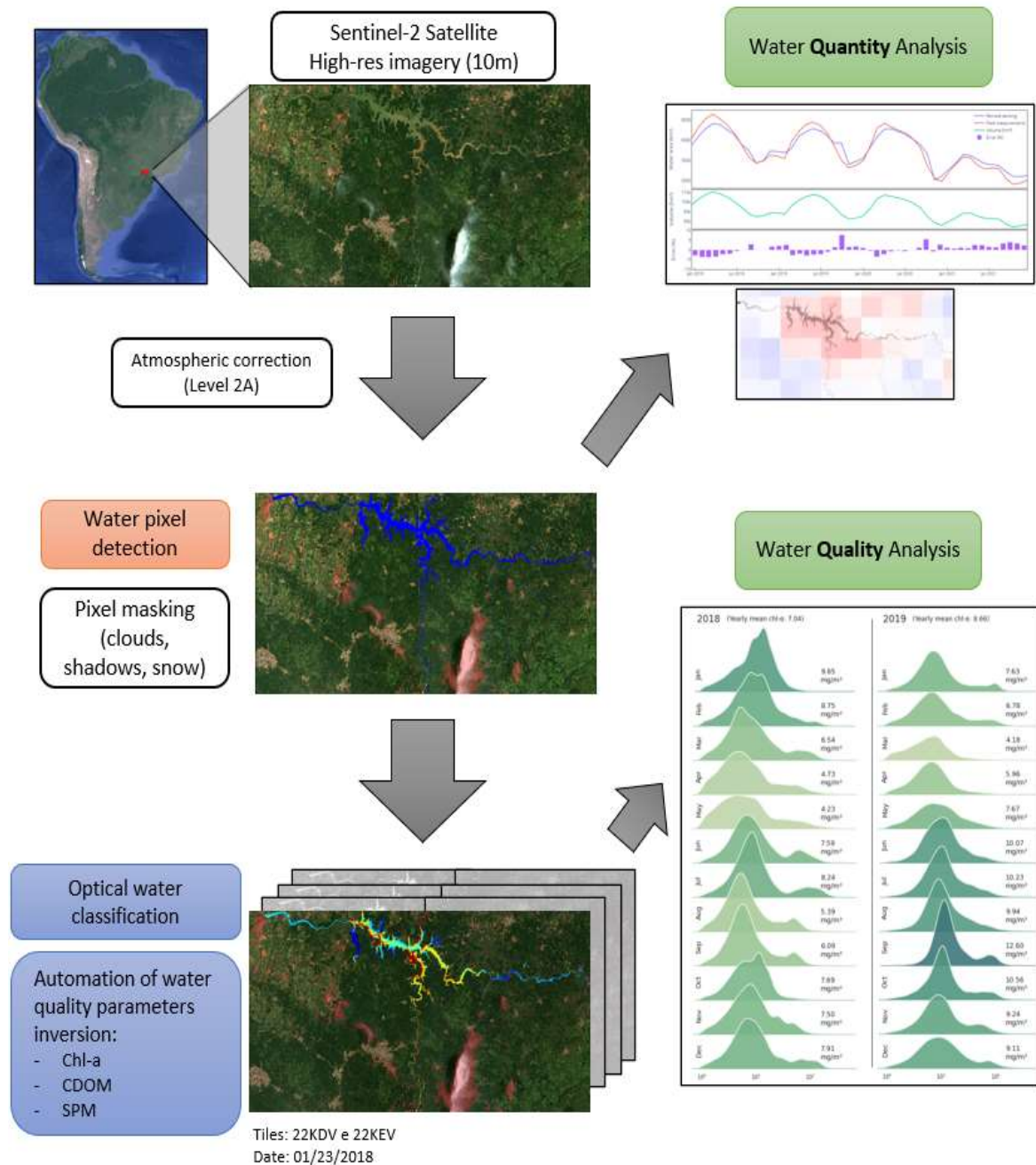


Figure 10: Vue schématique des étapes de l'évaluation de la qualité des eaux intérieures à partir d'images de télédétection par satellite qui ont été développées dans cette thèse. Les cases en orange représentent les étapes couvertes par l'axe 1 (détection de l'eau), les cases en bleu représentent l'axe 2 (classification de l'eau) et les cases vertes représentent l'axe 3 (généralisation et analyse).

Axis 1 – Détection automatique et non supervisée des plans d'eau

Le point de départ de toute évaluation de l'eau par le biais d'un cadre de télédétection est de séparer correctement les pixels d'eau des autres cibles dans la scène. Au cours de la correction atmosphérique des images satellitaires qui les convertit du niveau 1C (réflectance au sommet de l'atmosphère) au niveau 2A (réflectance de surface), un masque géophysique est généralement produit contenant la classification des pixels de la couverture terrestre. Les processeurs de niveau 2A les plus courants pour Sentinel-2 (MAJA, Sen2Cor et F-mask) fournissent le masque d'eau comme sous-produit de la correction atmosphérique. Cependant, comme les eaux intérieures ne sont pas leur priorité, les masques d'eau fournis par ces processeurs sont généralement imprécis, en particulier pour les eaux intérieures complexes et les petits plans d'eau, et ne devraient être utilisés qu'en dernier recours (Cordeiro et al., 2021).

En dehors des processeurs standard, de nombreuses études ont été publiées dans le domaine, mais les algorithmes actuels nécessitent généralement une intervention humaine, dans le cas des méthodes d'apprentissage supervisées qui ont besoin de données d'entraînement pour la calibration, ou bien manquent de robustesse pour être appliqués dans différentes conditions en raison de la diversité des eaux optiquement complexes (en raison de ses constituants) et des types de couverture des sols (végétation, agriculture, etc.).

Dans ce contexte, le premier axe se concentre sur les défis liés à la détection de l'eau à partir d'images satellites optiques. L'objectif est de fournir une méthodologie qui puisse être appliquée automatiquement dans différentes régions présentant des couvertures terrestres et des conditions atmosphériques variées.

Outre la robustesse, la faisabilité d'une détection de l'eau à large échelle est une autre question scientifique à traiter dans l'axe 1, car elle peut représenter une contrainte importante en

fonction de la zone surveillée. À titre de référence, une seule scène Sentinel-2, qui couvre une zone de 110 km x 110 km, comporte plus de 100 millions de pixels, avec 13 bandes radiométriques, et la classification automatisée des pixels en eaux représente une étape importante dans le cadre d'une chaîne dédiée à l'étude de la qualité des eaux.

D'autres facteurs intermédiaires qui peuvent affecter l'identification correcte des pixels d'eau sont également évalués dans cet axe. Par exemple, la correction atmosphérique à appliquer et les méthodes de masquage pour éliminer les pixels indésirables tels que les nuages, les ombres ou la neige sont également discutées dans le contexte de l'axe 1.

Axis 2 – Classification de la couleur de l'eau

L'évaluation de la qualité de l'eau et notamment des matières en suspension (MES) par télédétection dépend de l'interaction qui se produit entre les constituants optiquement actifs (OAC en anglais) et la radiation électromagnétique dans les régions du visible et du proche infrarouge (Barbosa et al., 2019). Des modèles d'inversion mathématiques sont ensuite développés, mais il est généralement admis qu'un seul modèle n'est applicable à tous les types d'eau présents sur la planète (Kirk, 1994). En outre, les résultats sont fortement couplés aux types d'eau et aux mesures de terrain utilisées pour l'étalonnage, et il n'est pas toujours simple de définir le domaine d'applicabilité de ces algorithmes (Xue et al., 2019). Ces problèmes affectent directement l'évolutivité de la télédétection pour les évaluations des ressources en eau, car ils imposent des contraintes lors de la livraison de produits de qualité de l'eau à grande échelle.

Pour surmonter cette difficulté, certaines études ont montré que les algorithmes peuvent être plus performants lorsqu'ils sont calibrés sur des gammes de concentration spécifiques de Chl-*a* (Matsushita et al., 2015 ; Pahlevan et al., 2020 ; Smith et al., 2018) ou des particules en suspension (SPM en anglais), (Condé et al., 2019 ; Yopez et al., 2018). Cependant, pour une évaluation

régionale ou globale automatisée, une difficulté supplémentaire doit être prise en compte. Même si plusieurs algorithmes sont calibrés pour différents types d'eau, il peut être difficile de s'assurer de la plage d'applicabilité et des limites de chaque modèle, de sorte qu'on ne sait pas *a priori* lequel est le mieux adapté à un lieu ou un moment particulier. En outre, une même masse d'eau peut contenir simultanément différents types d'eau (Moore et al., 2014 ; Neil et al., 2019).

À cet égard, des études ont montré que l'utilisation de schémas de classification des eaux et l'étalonnage de modèles pour chaque type d'eau optique (OWT, en anglais) défini peuvent contribuer à améliorer la précision des algorithmes de récupération (Lubac et Loisel, 2007 ; Moore et al., 2014 ; Vantrepotte et al., 2012). Cependant, bien que ces articles puissent démontrer une amélioration de la récupération des paramètres si les OWTs sont pris en compte, la précision de l'inversion n'est pas l'objectif final du processus de classification, c'est un objectif secondaire.

Compte tenu de ce qui précède, l'objectif du deuxième axe est de proposer un cadre permettant de classer les OWTs afin de maximiser la précision d'une inversion en termes de niveau de concentration de MES. En utilisant des stratégies d'apprentissage automatique telles que celles employées dans l'axe 1, comme le regroupement agglomératif non supervisé, l'axe 2 propose de séparer les courbes spectrales en types d'eau, et non par distribution spatiale ou bassin fluvial. L'idée principale est d'explorer la réponse hyperspectrale de différentes rivières et de les combiner automatiquement en OWTs distincts, afin que la récupération des paramètres de qualité des eaux soit optimale.

Une base de données complète de plus de 1 000 mesures radiométriques *in situ*, couplée à d'autres mesures de qualité de l'eau en laboratoire, et préalable à la thèse sera utilisée. En raison des contraintes de données, la méthodologie sera développée pour récupérer les SPM, mais le cadre

méthodologique développé peut être étendu à d'autres paramètres de qualité de l'eau, sous réserve de la disponibilité des données de calibration.

Axis 3 – Généralisation et Analyses

Les défis envers la robustesse et la généralisation ne sont pas limités à ce troisième axe. Comme déjà présenté, les deux premiers axes ont également été motivés par le manque d'outils permettant une utilisation opérationnelle des images de télédétection par satellite pour l'évaluation des eaux intérieures.

Cependant, le développement de méthodes pouvant être appliquées dans différentes régions n'est qu'une première étape. Pouvoir appliquer ces méthodes à l'échelle régionale et analyser ces données ajoute une autre dimension de complexité en raison des contraintes de volume.

Dans cet axe, les concepts de *Big Data* tels que le *cloud computing* et le *parallel processing* sont explorés, afin de permettre le traitement de milliers d'images qui ne serait pas réalisable avec les outils conventionnels.

Une fois ces données correctement traitées, une autre question scientifique se pose et doit être abordée. Comment analyser et extraire des informations significatives de cette énorme quantité de données et comment présenter ses résultats ?

Dans ce contexte, ce troisième axe est consacré à l'exploration des méthodologies qui peuvent permettre de faire évoluer les méthodes actuelles d'évaluation des eaux intérieures à l'échelle régionale, de discuter de leurs lacunes et de proposer des alternatives. Pour atteindre cet objectif, une étude de cas a été proposée pour surveiller par télédétection une grande région de 320.000 km² dans le bassin du Paraná au Brésil. Cette région a été choisie en raison de son importance nationale pour des usages multiples des ressources en eaux et parce qu'elle a été confrontée à une sécheresse majeure en 2021. L'étendue de la surface de l'eau, ainsi que la turbidité

et le Chl-*a* ont été évalués de 2018 à 2021 à l'aide de l'imagerie Sentinel-2 et les résultats ont été analysés en conséquence.

6. Structure de la thèse

Compte tenu des objectifs mentionnés ci-dessus, cette thèse est organisée en trois chapitres principaux, comme détaillé ci-dessous :

Chapitre 1 - Détection de l'eau

Ce chapitre se concentre sur le défi de distinguer les pixels d'eau des autres objets dans les images satellites. Ce chapitre présente l'article « *Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors* » (Cordeiro et al., 2021) publié dans la revue *Remote Sensing of Environment*. L'article propose une nouvelle méthode non supervisée qui utilise des méthodes d'apprentissage automatique, notamment le clustering et la classification de Naïve Bayes, pour améliorer la robustesse et la performance de la séparation des pixels d'eau dans les scènes optiques à haute résolution.

Un deuxième article (Annexe I), « *Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level* » (Peña-Luque et al., 2021), auquel j'ai participé en tant que troisième auteur, s'est concentré sur une comparaison plus poussée de la méthodologie de clustering proposée avec celles existantes et a comparé les précisions des plateformes optiques et radar séparées. Ces comparaisons ont été effectuées au niveau régional et au niveau des réservoirs et l'effet des estimations à fenêtre temporelle a également été évalué.

D'autres développements sur le même thème vers une généralisation globale et des améliorations de résolution ont produit respectivement les articles « *Surface water dynamics from space : a round robin inter-comparison of using Optical and SAR high-resolution (10-m) satellite*

observations for regional surface water detection » (Tottrup et al., 2022), publié dans *Remote Sensing - MDPI journal* et « *Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, Planetscope and Sharpened imagery* », soumis pour publication dans le *GIScience & Remote Sensing journal*. J'ai participé à l'élaboration de ces deux articles qui portent sur le même thème mais dans des environnements et des capteurs différents. Ils sont présentés respectivement en Annexe II et Annexe III.

Chapitre 2 – Classification optique de l'eau

Ce chapitre présente un cadre permettant de classer les types d'eau, en fonction de leurs spectres de réflectance, afin de maximiser la précision de l'extraction des SPM. Il s'intitule « *Optical Water Classification for Suspended Particle Matter (SPM) retrieval* » et utilise des méthodes de regroupement non supervisées pour séparer les spectres de l'eau en groupes. Plusieurs caractéristiques d'entrée, méthodes de normalisation et algorithmes de regroupement sont testés.

Ces dernières années, plusieurs articles ont traité de la classification des classes optiques de l'eau, mais ils étaient principalement axés sur la cartographie de la chlorophylle-a ou sur les eaux côtières. Il est nécessaire d'aborder la question des particules en suspension qui sont le premier facteur de contrôle de la radiance montante de l'eau dans les eaux continentales en fonction des processus d'érosion et de transport des sédiments en suspension dans les rivières, les lacs et les réservoirs. En particulier, nous chercherons le nombre optimal de classes permettant d'atteindre la meilleure précision de récupération et pour différentes résolutions spectrales. Pour cela, nous avons bénéficié d'une base de données sans précédent de mesures hyperspectrales de terrain sur différentes rivières permettant de représenter des bassins versants contrastés en termes de climat, de géologie et de processus hydrologiques.

Le troisième chapitre présente les méthodologies développées pour la généralisation et l'analyse des paramètres de qualité des eaux intérieures. La première partie est une présentation approfondie des méthodologies développées dans l'étude actuelle pour extrapoler la détection de l'eau et l'inversion des paramètres de qualité à une échelle régionale et sur une longue période. Dans ce chapitre, l'automatisation et la robustesse sont exploitées à une échelle supérieure (supérieure à 300 000 km²) pour permettre le traitement de dizaines de milliers d'images satellites sans intervention manuelle.

La présentation des méthodologies est divisée en deux parties. La première partie, *Water Detection Methodology*, entre dans les détails des améliorations apportées au paquet *waterdetect* qui n'étaient pas couvertes dans l'article original « *Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors* » (chapitre 1) pour une exploitation à large échelle. La deuxième partie, *Water Assessment Methodology*, concerne tous les post-traitements qui ont été effectués sur les masques d'eau résultants et les étapes d'analyse correspondantes.

Après cette présentation détaillée des méthodologies qui vont être utilisées dans ce chapitre, une étude de cas est présentée sous la forme d'un article, qui utilise les méthodes proposées pour évaluer les effets d'une sécheresse majeure qui s'est produite dans le bassin du fleuve *Paraná* au Brésil en 2021, sur le quantité et de qualité de l'eau de milliers de lacs artificiels et naturels.

En ce qui concerne l'extension de l'eau, toute la région qui avait été considérée comme étant en situation critique par les autorités brésiliennes lors de la sécheresse de 2021 a été évaluée. Une zone de 320 000km² a été analysée de janvier 2018 à décembre 2021. Un total de 12 000 images Sentinel-2 avec une résolution spatiale de 10 m a été utilisé pour évaluer les effets de la sécheresse

sur plus de 20 000 réservoirs d'une superficie supérieure à 1 ha. Outre l'analyse stratifiée par classes de taille des réservoirs, qui ne serait pas réalisable avec des mesures sur le terrain, une analyse des anomalies spatiales est également effectuée, mettant en évidence la façon dont la sécheresse a affecté différemment les régions du bassin.

Pour l'évaluation de la qualité de l'eau, un sous-bassin du bassin du *Paraná*, appelé bassin du *Parapanema*, a été sélectionné en raison de la disponibilité de données corrigées de l'effet de sunglint (voir chapitre 3). Les variations saisonnières et spatiales des paramètres de qualité de l'eau turbidité et le Chl-a ont été estimés et ensuite analysés pour mieux comprendre l'impact de la sécheresse et de la variabilité hydrologique.

7. Contexte du développement de la thèse

Le présent travail a été développé dans le cadre de l'accord de coopération HIDROSAT entre l'Institut Français de Recherche pour le Développement (IRD), le laboratoire Géosciences Environnement Toulouse (GET), et l'*Agência Nacional de Águas* (ANA), agence chargée des eaux et de l'assainissement du Brésil, qui a financé mes études de doctorat, afin de renforcer les capacités en hydrologie spatiale. Plus particulièrement, mon travail sur l'utilisation des images du capteur Sentinel-2/MSI a été développé dans le contexte du projet OBS2CO (acronyme pour OBServation de la Couleur des eaux Continentales) et du centre de données français pour les surfaces continentales THEIA qui implique plusieurs institutions pour fournir à la communauté des sciences de l'eau une évaluation des produits de haute résolution sur la qualité de l'eau des lacs, réservoirs et rivières.

En plus de l'IRD et l'ANA, d'autres institutions telles que le Centre national d'études spatiales (CNES), l'Office français de la biodiversité (OFB), l'*Universidade de Brasilia* (UNB) et l'*Universidade Federal do Amazonas* (UFAM) participent à ces initiatives. L'objectif final est de

développer des flux de travail opérationnels pour traiter les séries temporelles télédéetectées et de les mettre à la disposition de la communauté. Parmi ses différents fronts de travail, nous pouvons souligner : 1) le développement de bases de données optiques et de qualité de l'eau ; 2) les algorithmes d'inversion ; 3) la validation des produits ; et 4) les plateformes de publication. Les résultats obtenus par ce pipeline permettent le développement de la recherche et des applications pratiques. L'eutrophisation, le flux de sédiments et les catastrophes environnementales sont quelques exemples de sujets abordés dans le cadre du projet.

Dans cette thèse, j'ai travaillé en étroite collaboration avec cette équipe multidisciplinaire et j'ai apporté des améliorations importantes en termes de méthodologie vers le traitement et l'analyse à grande échelle.

Chapter I - Water Detection

Correctly identifying water pixels from multispectral images is the first step towards inland water assessment from remote sensed data but it is not a trivial task, and its operational application remains a challenge. Atmospheric correction algorithms usually deliver land cover masks, but the accuracy is not satisfactory for inland water bodies (Cordeiro et al., 2021). This behavior can be explained because the zero leaving radiance on infrared wavelengths assumption, usually applied to identify clear water, is not valid for inland waters due to the complex interactions that occur among its constituents.

Besides that, other surface targets can have spectral signatures very similar to water, such as very dark vegetation or cloud shadows.

Although many methods have been proposed in the literature, most of them rely on supervised learning, that requires training samples on the region to be studied, that involves cost and time. Another group of automated methods rely on thresholds that should be applied to reflectance or spectral indices. This second group is usually more straightforward, but the thresholds should be selected carefully, and the results can be tricky. This happens because the optimal threshold can vary in space and time making it difficult to be applied automatically to a large number of images. Sometimes, even within the same scene, the best threshold can vary depending on the water body, making it difficult to deliver optimal results.

To address this issue, a new detection method has been proposed in the context of this study, that makes use of unsupervised machine learning clustering method, intitled “**Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors**”. The results are compared to traditional methods and to

the operational masks delivered by the main Level 2A processors and shows an improvement especially for smaller water bodies. This article has been published in the Remote Sensing of Environment (RSE) journal and the resulting piece of software, called Waterdetect was made available as open-source code to the research community through the GitHub platform (<https://github.com/cordmaur/WaterDetect>).

Still in the context of the water detection, a second article intitled “**Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level**” (Peña-Luque et al., 2021), was published on Remote Sensing MDPI journal (Annex-I). This article, in which I participated as third author, focused on further comparison of the proposed clustering method to existing ones in the optical image domain and compared the results to those obtained by radar platforms. In addition to regional scale assessment, this study also compared the performance of the methods to monitor water surface of specific reservoirs over the time. The utilization of time windows has been proposed to minimize inaccuracies due to cloud persistence and other punctual errors.

This second paper led to improvements in the **waterdetect** software, especially concerning the robustness of the method. A new treatment for negative reflectance has been proposed, as well as the possibility to perform the clustering with different combinations of reflectances and spectral indices and to average the final results.

Further development on the same theme has been conducted in a third article “**Surface water dynamics from space: a round robin inter-comparison of using Optical and SAR high-resolution (10-m) satellite observations for regional surface water detection**” (Tottrup et al., 2022), published in Remote Sensing – MDPI journal (Annex II). This round robin was realized in the context of the World Water project, proposed by the European Space Agency (ESA), which

goal is to develop robust and scalable Earth Observation solutions for inland surface water monitoring. The intercomparison has been performed among several peer reviewed algorithms that used either optical or Synthetic Aperture Radar (SAR) or a combination of both do deliver accurate water mapping. The algorithms were put to test in 5 sites covering various eco and climatic regions as well as including a diversity of waterbodies and some of the major challenges for water detection such as clouds, canopy shading, fire scars, urban areas, and regions with permanent low backscatter. In general, the models that synergically combine optical and SAR data outperformed single approach models, being cloud cover the major limiting factor affecting the performance using optical imagery.

The possibility to use the methodology with data from diverse satellites, with different resolution and bands configuration, has been object of another article that I participated in, entitled **“Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, Planetscope and Sharpened imagery”** (Tayer et al., 2022, in review), and presented in (Annex III). In this study, beyond Sentinel-2 imagery, the **waterdetect** algorithm was applied to Planetscope scenes, with 3m of spatial resolution but lacking a SWIR band. Additionally, a sharpened version of Sentinel-2 imagery, with and without considering SWIR band, was also assessed. The objective was to detect an intermittent river in a semi-arid environment, distinct from those used so far. A sensibility test has also been conducted to assess the impact of the main hyperparameters of the model to the final performance. As main outcomes, it is shown that the sharpening procedure had little impact on the overall accuracy and that the NIR band and visible bands are enough to deliver good results.

Beyond those, the methodology has also been used in other side projects conducted during the thesis such as a water surface assessment performed in lake Chad, located in the sub Saharan Africa in the context of a UNESCO project, where the algorithm has been used to retrieve the overall surface area for 7 distinct tiles from 2017 to 2020 (Figure I-1).

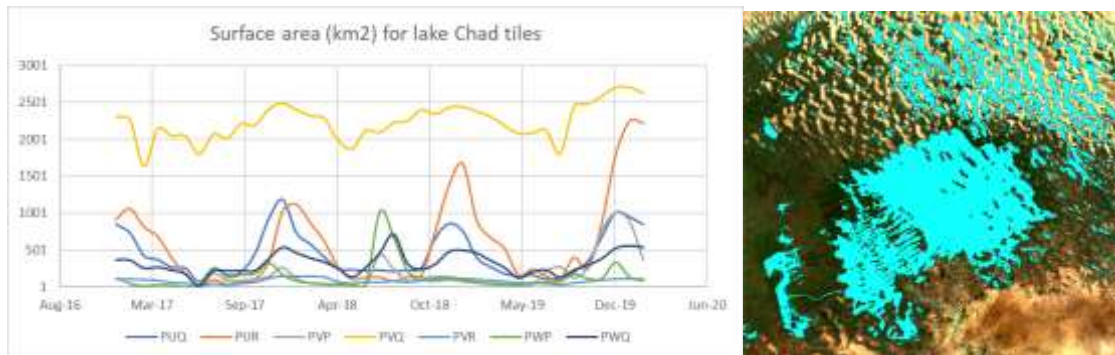


Figure I-1: Example of study conducted using the *waterdetect* algorithm on the lake Chad region. (a) monthly water surface in km² for each considered tile individually; (b) lake Chad water mask.

The participation on these studies following the original paper were important to validate the methodology on different conditions, compare the results to other state of the art methods and improve performance and robustness of the code. The improvements not covered in the articles are presented in more details in Chapter III – Generalization and Analysis: A case study of Brazil’s 2021 drought.

Considering the aforementioned studies, the **waterdetect** methodology could be tested in many different geographically distributed sites, in different seasons and with diverse climatological conditions. In total, more than 20,000 Sentinel-2 images have been successfully processed with a global coverage, as shown in Figure I-2.



Figure I-2: Overview of the main sites where the *waterdetect* algorithm has been successfully applied. Background image: "World Map Blue" licensed under Creative Commons license.

1. Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors



Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors

Maurício C.R. Cordeiro^{a,b,*}, Jean-Michel Martinez^b, Santiago Peña-Luque^c

^a Agência Nacional de Águas (ANA), Setor Policial Sul, Área 5, Quadra 3, Brasília (DF) 70610-200, Brazil

^b Géosciences Environnement Toulouse (GET), Unité Mixte de Recherche 5563, IRD/CNRS/Université, Toulouse 31400, France

^c Centre National d'Etudes Spatiales (CNES), Toulouse 31401, France

ARTICLE INFO

Keywords:

Water detection
Water mask
Sentinel-2
Unsupervised clustering
Machine learning
naïve bayes classifier

ABSTRACT

Continuous monitoring of water surfaces is essential for water resource management. This study presents a nonparametric unsupervised automatic algorithm for the identification of inland water pixels from multispectral satellite data using multidimensional clustering and a high-performance subsampling approach for large scenes. Clustering analysis is a technique that is used to identify similar samples in a multidimensional data space. The spectral information and derived indices were used to characterize each scene pixel individually. A machine learning approach with random subsampling and generalization through a Naïve Bayes classifier was also proposed to make the application of complex algorithms to large scenes feasible. Accuracy was evaluated using an independent dataset that provides water bodies in 15 Sentinel-2 images over France acquired in different seasons and that covers a large range of water bodies and water colour types. The validation dataset covers a water surface of more than 1200 km² (approximately 12 million pixels) including over 80,000 water bodies outlined using a semiautomatic active learning method, which were manually revised. The classification results were compared to the water pixel classification using three of the major Level 2A processors (MAJA, Sen2Cor and FMask) and two of the most common thresholding techniques: Otsu and Canny-edge. An input mask was used to remove coastal waters, clouds, shadows and snow pixels. Water pixels were identified automatically from the clustering process without the need for ancillary or pretrained data. Combinations using up to three water indices (Modified Normalized Difference Water Index-MNDWI, Normalized Difference Water Index-NDWI) and Multi-band Water Index-MBWI) and two reflectance bands (B8 and B12) were tested in the algorithm, and the best combination was NDWI-B12. Of all the methods, our method achieved the highest mean kappa score, 0.874, across all tested scenes, with a per-scene kappa ranging from 0.608 to 0.980, and the lowest mean standard deviation of 0.091. Standard Otsu's thresholding had the worst performance due to the lack of a bimodal histogram, and the Canny-edge variation achieved an overall kappa of 0.718 when used with the MNDWI. For water masks provided by generic processors, FMask outperformed MAJA and Sen2Cor and obtained an overall kappa of 0.764. In-depth analysis shows a quick drop in performance for all of the methods in identifying water bodies with a surface area below 0.5 ha, but the proposed approach outperformed the second best method by 34% in this size class.

1. Introduction

Continuous monitoring of water bodies is essential for water resources management, but for many countries, especially those with continental extensions, the lack of field monitoring is an issue (Barbosa, 2019). In this regard, earth observations from spaceborne sensors have shown the potential to complement field surveys and provide scientific information in various water-related domains (Bukata, 2013).

Recently, there has been increasing capacity of remote sensing earth observation satellites, such as in the Sentinel missions, which offer high spatial, spectral and temporal resolutions at a global scale without cost. This capacity enables scientists to explore different possibilities, including surface water coverage (Buma et al., 2018; Feyisa et al., 2014; Markert et al., 2018; Pelel et al., 2016; Souza et al., 2019), flood and inundation mapping (Kordeias et al., 2019, 2018; Martins et al., 2011; Wieland and Martins, 2019), water quality monitoring (Anspér and

* Corresponding author at: Agência Nacional de Águas (ANA), Setor Policial Sul, Área 5, Quadra 3, Brasília (DF) 70610-200, Brazil.

E-mail addresses: mauricio@ana.gov.br (M.C.R. Cordeiro), martinez@ird.fr (J.-M. Martinez), santiago.penaluque@cnes.fr (S. Peña-Luque).

<https://doi.org/10.1016/j.rse.2020.112209>

Received 19 June 2020; Received in revised form 19 November 2020; Accepted 22 November 2020

Available online 8 December 2020

0034-4257/© 2020 Elsevier Inc. All rights reserved.

Alikas, 2016; Delegido et al., 2014; Frampton et al., 2013; Lins et al., 2017; Toming et al., 2016; Yadav et al., 2019), suspended sediment assessment (Condé et al., 2019; Martínez et al., 2009; Yepes et al., 2016), among others.

In this context, the correct identification of water pixels within an image is the first process required for many subsequent applications. The proposed solutions in the literature include a number of approaches that are suitable for different objectives, scales (local, regional or global) and sensor characteristics (Feng et al., 2016; Kordelas et al., 2018; Pekel et al., 2016; Souza et al., 2019).

Supervised (Höllstein et al., 2016) and unsupervised classification (Yousefi et al., 2018), hue saturation and value (HSV) transformation (Dinh Ngoc et al., 2019; Pekel et al., 2016, 2014), spectral mixture analysis (SMA) (Feng et al., 2015; Souza et al., 2019), water index thresholding (Donchyts et al., 2016; Du et al., 2016; Kordelas et al., 2018; Zhang et al., 2018), object segmentation (Kaplan and Avidan, 2017) and, more recently, deep learning (Wieland and Martinis, 2019) are examples of techniques that have been used to separate water bodies from land cover in radar-based and optical images.

Although radar-based water mapping has the advantage that it is not affected by cloud or weather conditions and detects water through a thin canopy, optical images, when available, are more straightforward for detecting water (Shen et al., 2019) and have the advantage of making it possible to use the spectral bands to perform other analyzes beyond water mapping using an increased number of parameters (Mackert et al., 2016).

Several optical water masking approaches rely on a stack of images acquired on different dates to avoid clouds, shadows or undesirable complex water conditions (Pekel et al., 2016), or they use ancillary data, such as Digital Elevation Models (DEM) or pre-existing cartography (Donchyts et al., 2016; Mackert et al., 2018). Although these approaches can achieve high accuracy scores, some of them are complex, time consuming or require human intervention to prepare and process data (Zhang et al., 2018).

Because this work is focused on providing a water pixel baseline for the development of subsequent products within single scenes, especially for the production of water quality products, this article focuses on automated methods for processing optical images without the use of any auxiliary data or time-series mosaicking.

For automatic water detection, the most commonly used methodologies include thresholding in one or more spectral bands or indices and pretrained machine-learning algorithms (Wieland and Martinis, 2019). The literature on supervised machine learning models for water mapping includes commonly used methods, such as support vector machines (Nandi et al., 2017), decision trees (Acharya et al., 2016), random forests (Fang et al., 2015; Ku et al., 2015), multilayer perceptron (Jiang et al., 2018; Mishra and Prasad, 2015) and convolutional neural networks (Pu et al., 2019; Wang et al., 2020).

The use of supervised machine learning models is preferable over thresholding for accurate results, but due to having different water and atmospheric conditions at different sites and their dependency on the training data, supervised machine learning might not be the best option when multiple sites at a global scale are addressed (Bangira et al., 2019).

According to Bangira et al. (2019), varying concentrations of suspended sediments (turbidity), photosynthetic pigments in algae (e.g., chlorophylls, carotenoids), dissolved organic matter and aquatic plants make the implementation of supervised optical remote sensing-based water extraction methods difficult because training data must be universally applicable and frequently updated, especially in the case of water bodies that present highly variable reflectance over space and time.

With regard to unsupervised water mapping from optical images, to date, the majority of developed methods rely on water indices, such as the Normalized Difference Water Index (NDWI) (McFeeters, 1996), Modified Normalized Difference Water Index (MNDWI) (Xu, 2006), Multiband Water Index (MBWI) (Wang et al., 2018) and Automated

Water Extraction Index (AWEI) (Feyisa et al., 2014). For automated and operational purposes, these indices are usually employed with an automated thresholding procedure, such as Otsu's thresholding (Otsu, 1979), which minimizes within-class variance in bimodal histograms. This procedure can be combined with other bands or ancillary data in complex workflows to overcome some of the indices' limitations (Dinh Ngoc et al., 2019; Donchyts et al., 2016; Kordelas et al., 2018; Mackert et al., 2018; Yang et al., 2018).

The threshold values can be obtained locally or globally, however a unique global threshold is ineffective because it can vary with satellite altitude, illumination, angle, atmospheric conditions (Ji et al., 2009), and water constituents (sediments, organic dissolved matter and chlorophyll). Working over one river catchment, but within boxes of $20 \text{ km} \times 20 \text{ km}$, Donchyts et al. (2016) found that the MNDWI optimal threshold varied from -0.25 to 0.4 , which highlights the need to adjust the threshold dynamically over space and time.

In this context, little attention has been paid to large areas, such as an entire Sentinel 2 scene, in which all the complex variabilities of water can lead to erroneous assessments due to a diversity of spectra occurring at the same time. One exception that is worth mentioning is Global Surface Water (Pekel et al., 2016), which was developed to assess water surfaces worldwide and their long-term changes. The equations for its inference system were obtained by manually drawing hulls in a multi-dimensional feature-space and interactively adjusting them to pixel samples. Ambiguous classification and spectral overlaps were solved by evidential reasoning, followed by a final visual inspection.

Bangira et al. (2019) compared the thresholding method with machine learning classifiers and concluded that the use of only one feature (i.e., one water index) produced relatively poor and unstable results compared to using a basic threshold combination and that more work is needed to investigate the efficacy of other combinations. To overcome some of the drawbacks of using water indices, their combination with other bands in a fully automated method has been an object of study (Dinh Ngoc et al., 2019; Kordelas et al., 2018, 2019), but in a rule-based system, their use in optically complex water remains a challenge (Bangira et al., 2019).

In addition, it appears that certain multidimensional unsupervised methods, such as clustering, have not been as popular as other methods for automatic water mapping based on the number of publications to date. Thus, clustering could be used as a solution for combining multiple features (reflectance bands and water indices) in a single automated procedure.

For this reason, the purpose of this study is to provide a robust method for water extraction from single scenes using optical high-resolution sensors that can be applied at a large scale. Different from other thresholding approaches that use only one dimension, usually a water index, this article aims to analyze the applicability of combining different water indices and spectral bands in unsupervised multidimensional hierarchical clustering, optimized for large-scale processing.

The obtained results are compared to water index thresholding, the most common method to separate water, and other major identification algorithms, such as Sen2Cor (Mueller-Wilm et al., 2019), FMask (Zhu et al., 2015; Zhu and Woodcock, 2012) and MAJA (Hagolle et al., 2010).

2. Materials and methods

2.1. Reference maps and study area

To test and validate the new water extraction methodology, it is necessary to have a good reference dataset. However, obtaining a good reference dataset can be a challenge for large areas because in-situ reference databases are scarcely available at a broad scale and remote sensing databases require costly, time-intensive investment (Boetens et al., 2019). Although previous studies have been performed to provide global water datasets, such as Global Surface Water data, provided by the European Commission (Pekel et al., 2016), or Global Water Bodies –

GLOWABO data (Verpoorter et al., 2014), our goal is to provide a per-scene pixel classification that is affected by instant climate and water conditions.

To overcome the difficulty of reference mask generation with a limited amount of manual work, Baetens et al. (2019) developed an active machine learning algorithm called Active Learning Cloud Detection (ALCD), which can be used for any classification purpose.

The approach of ALCD consists of an iterative process in which the operator manually selects the reference points, trains the machine learning model based on random forest and classifies the scene. After the classification, the operator visually determines the possible imperfections and labels new pixels where the classification went wrong or was uncertain and then repeats the cycle until a desirable result is achieved. The operator performs a manual correction of persistent errors if necessary.

For the continental water domain, the ALCD algorithm was used to produce the CNES ALCD Open Water Masks dataset (Santiago, 2019), which is available online. This open dataset comprises open water masks for 16 Sentinel-2 scenes over 9 different regions in France that were selected to minimize the presence of clouds while including areas with diverse land coverage (Fig. 1). Some areas have reference masks in different seasons, such as the end of summer and the end of winter periods. The masks were generated at a 10 m resolution, with each tile covering $110 \text{ km} \times 110 \text{ km}$. Among the available scenes, tile T31TGG, from the Alpes region during winter, was discarded because of the unavailability of ancillary data related to this scene that are required to complete the study.

The selected dataset has an inland area coverage of 96.315 km^2 with approximately 1.239 km^2 of surface water and water quality characteristics that would match most European countries. The complete list of scenes used in this study, with basic characteristics and water coverages and percentages, is described in Table 1.

2.2. Data collection and preprocessing

The Copernicus Sentinel 2 mission consists of two polar-orbiting satellites (Sentinel 2-A and Sentinel 2-B) that were developed to monitor variability in land surface conditions with a revisit time from 2

Table 1

Description of the scenes used for reference with the corresponding continental water surface in km^2 and as a percentage.

Region	Tile	Dates	Continental Water Area, km^2 (%) ¹	Scene content
Alpes	T31TGL	28 August 2018	125.04 (1.03)	Mountains
Alsace	T32ULU	12 September 2018 21 March 2019	83.55 (0.69) 100.12 (0.83)	Lowlands and gentle slopes
Ariège	T31TCH	23 October 2018 22 March 2019	39.94 (0.34) 31.42 (0.26)	Mountains, small water bodies
Bordeaux	T30TXQ	11 September 2018 23 February 2019	184.40 (2.47) 111.36 (2.51)	Large water bodies, very turbid and clear water
Bretagne	T30UXU	08 July 2018 23 February 2019	50.81 (0.42) 54.93 (0.46)	Lowlands, Wetlands, small water bodies
Camargue	T31TFJ	27 September 2018 31 March 2019	433.83 (4.18) 489.10 (4.71)	Large water bodies, floodplain/delta
Chateauroux	T31TCM	19 August 2018 25 February 2019	133.49 (1.10) 121.42 (1.00)	Lowlands
Gironde	T30TXR	23 February 2019	92.63 (1.37)	Large water bodies, very turbid and clear water
Marmande	T30TYQ	22 February 2019	95.33 (0.79)	Lowlands, small water bodies

¹ The continental water area and percentage were calculated while accounting for only the valid pixels within each scene.

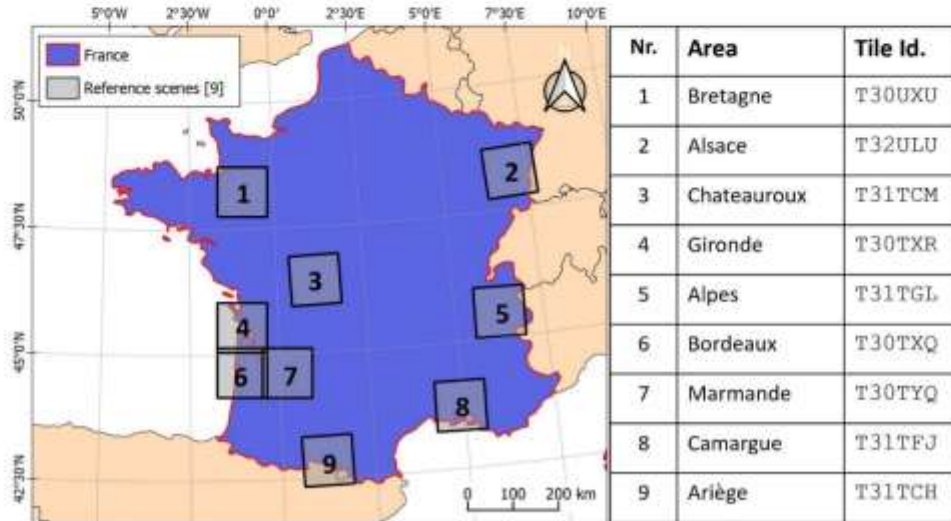


Fig. 1. Spatial distribution of the reference scenes over France used for validation in this work.

to 5 days according to latitude. Both satellites are equipped with an optical Multi-spectral Instrument (MSI) that can monitor 13 spectral bands with wavelengths ranging from 440 to 2200 nm at high resolution, from 10 to 60 m depending on the spectral band (Suhet, 2015).

Different processors have been developed for Sentinel 2 to correct images for atmospheric conditions and to provide the surface reflectance (Level 2A). One important subproduct from these processes is the pixel classification, which provides Clouds, Shadows, Land, and Water masks among other ancillary data. Sentinel 2 - Level 2A products distributed by ESA are generated by the Sen2Cor processor (Mueller-Wilm et al., 2019), while the French Land Data Center (named Theia) uses the MAJA processor (Gascoin et al., 2019; Hagolle et al., 2010). Another important provider, the United States Geological Survey (USGS), generates the final pixel classification using the Function of Mask (FMask) algorithm (Qiu et al., 2019; Zhu et al., 2015).

To compare the performance of the proposed algorithm with the water mask provided by the FMask, MAJA and Sen2Cor processors, all 15 scenes selected within the reference dataset were downloaded from two different sources. The level-2A MAJA corrected images were downloaded from CNES's Theia Land Data Center portal (<http://theia.cnes.fr/>). For the FMask mask, we downloaded level-1C images from ESA's Scientific Data Hub, and for the Sen2Cor masks, we downloaded Level-2A products from the same source. Then, version 4.0 of the FMask processor, which is available online (<https://github.com/GEOSL/FMask>), was used to produce the FMask categorical map.

Considering the importance of the atmospheric correction to better characterize the spectral signature of the desired targets (water and nonwater) and obtain a uniform and consistent dataset (Dinh Ngoc et al., 2019; Jiang et al., 2010; Wang et al., 2010), all of the images used to run the proposed algorithm were Level 2-A. MAJA was selected, considering the quality of its atmospheric correction and cloud mask in comparison to other processors (Baetens et al., 2019).

To avoid undesired pixels that are not objects of this study and that could bias the results from the algorithm, one external reference mask was produced for each scene. Five categories of pixels were masked using different specialized sources, as shown in Fig. 2. The undesired categories are snow, cloud, cloud shadow, geographic shadow and ocean/coastal waters, and the respective sources for each category are listed below:

- Snow: snow cover was downloaded from the Theia Snow collection (Gascoin et al., 2019), which is an operational snow detector with a 20 m resolution, developed by the CRSEBIO laboratory;
- Ocean/coastal waters: CNES ALGD Open Water Masks (Santiago, 2019) has a version called "inland masks" that delineates the oceans/coastal waters. This dataset was produced considering the coastal lines available at the Global Self-consistent, Hierarchical, High-resolution Geography Database - GSHHG (<http://www.soest.hawaii.edu/pwestel/gshhg/>) (Wessel and Smith, 1996) at level H and using an erosion filter of 400 m toward the continent. The shoreline was produced at a working scale of 1:100,000 according to the World Vector Shoreline database;
- Clouds and shadows: For the treatment of clouds, cloud-shadows and geographic shadows, the geophysical - MG2 mask from MAJA was selected because its multitemporal approach has been proven to be superior to other cloud detection processors, as shown in previous studies (Baetens et al., 2019). Although the MG2 layer is provided at 10 m and 20 m resolutions, these masks were computed at 240 m for the Sentinel-2 imagery.

2.3. Computational framework and infrastructure

Considering the amount of data and the many experiments required to complete this study, as described in section 3, all the processing was performed on a high-performance cluster within the CNES's computational infrastructure.

The main algorithm was developed in Python 3.7. The clustering, qualification indices and machine learning algorithms were implemented using the SciKit Learning library (Pedregosa et al., 2011), and the Otsu's thresholding algorithm (Otsu, 1979) was from the SciKit Image library (van der Walt et al., 2014). For the geospatial data manipulation, the GDAL library (GDAL/OGR contributors, 2020) was employed.

The comparison task was performed using the validation module of SurfWater software developed by CNES's team, which computes the confusion matrix with more relevant parameters, including Kappa, Precision, Recall, F1 Score, among others.

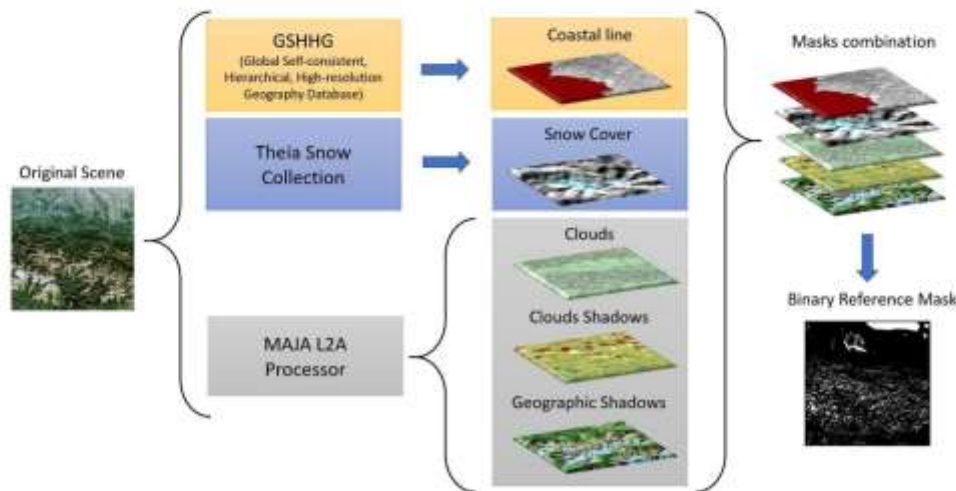


Fig. 2. Schematic view of the sources and categories of the pixels selected to produce the reference mask for each scene.

3. Methodology

3.1. Concept and rationale

Despite their simplicity, simple water extraction methods, such as two-band water indices and single-band thresholding, are not able to accurately distinguish water in complex environments that include build-up areas, dark surfaces or mountain shadows (Freyta et al., 2014; Wieland and Martinis, 2019). In addition, when addressing large areas (i.e., 110 km × 110 km when considering a full Sentinel-2 scene), it is common to have different environments at the same time. The combination of water indices results with other reflectance bands has been tested in some rule-based water extraction approaches (Dinh Ngoc et al., 2019; Koutelas et al., 2018), but the optimal combination of different thresholding results remains uncertain and often requires tuning to local conditions (Bangira et al., 2019).

In addition, one of the major criticisms of the Otsu thresholding method is the instability that occurs when the histogram of the pixel values is not bimodal (Douchyts et al., 2016). That circumstance can occur in different situations, depending on the relations of the land and water pixels in the selected scene or study area, for example, in areas with very few water pixels compared to land pixels.

The method proposed in this study combines different water indices and reflectance bands into a single unsupervised multidimensional clustering approach for large areas; we compare its results with those from the most commonly used threshold methods and ready-to-use processors. The proposed method is applied on individual Sentinel-2 scenes and is intended to be used as an input water mask for different water resources applications, such as inland water quality monitoring, water temperature assessment, surface water storage or geomorphological analysis of river streams.

3.2. Feature selection

An important step in machine learning is called feature engineering, in which features are the input information provided to the algorithm. The primary features considered are the reflectance bands.

Pure water is known to have very strong light absorption for wavelengths above 800 nm, covering the near-infrared (NIR) and short-wave infrared (SWIR) Sentinel-2 bands, compared to the visible bands. This behavior serves as the rationale for the most commonly used water indices (NDWI and MNDWI) (McFeeters, 1996; Xu, 2006). In addition, Bangira et al. (2019) assessed the utilization of each of the Sentinel 2 spectral bands, among other indices, for water discrimination through Otsu-based thresholding, and the single bands that achieved the highest accuracies were B8 (NIR) and B11 (SWIR).

Because Sentinel-2 has two bands in the short-wave infrared spectra and because the B11 band is already used in the MNDWI and the MBWI indices, B12 was selected considering its lowest response over water surfaces. In addition to the selected B8 and B12 Sentinel-2 bands, the most commonly used indices for water segmentation, NDWI and MNDWI, were generated as input features for each scene. Additionally, the recently proposed MBWI (Wang et al., 2018), which was developed to maximize the spectral difference between water and nonwater surfaces in Landsat 8 scenes, especially for confused backgrounds (i.e., mountainous shadows and dark built-up areas), was adapted to work with Sentinel 2 bands and was also included in the current analysis (Table 2).

Lower resolution bands (B11 and B12) were upsampled to 10 m using the simple average algorithm. The two reflectance bands, B8 and B12, and three indices were then combined in different clustering experiments with up to five dimensions when all of the features were considered together. Considering that the order of the features is not important for the clustering method, the number of necessary experiments can be assessed by the combination formula (Eq. (1)):

Table 2

Features and calculation methods considered in the clustering experiment.

Feature	Resolution	Notation	Equation
Visible green (537-582 nm)	10 m	B3	–
Visible red (646-685 nm)	10 m	B4	–
Near-infrared (767-908 nm)	10 m	B8	–
Short-wave infrared (1539-1681)	20 m	B11	Average up-sampling
Short wave infrared (2072-2312)	20 m	B12	Average up-sampling
Modified normalized difference water index	10 m	MNDWI	$MNDWI = \frac{B3 - B11}{B3 + B11}$
Normalized difference water index	10 m	NDWI	$NDWI = \frac{B3 - B8}{B3 + B8}$
Multi band water index	10 m	MBWI	$MBWI = 3 * B3 - B4 - B8 - B11 - B12$

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad (1)$$

where n stands for the number of features and k is the number of dimensions. To test the applicability of B11 instead of B12, we ran extra tests by changing the input band to B11, but the results were systematically less robust than the same experiment with B12.

Additionally, as part of the experiment, for comparison purposes, Otsu-based thresholding and its modified version using the Canny-edge filter were also adopted for each index, achieving a total of 29 experiments for each scene. Table 3 summarizes the water mapping experiments and features used in this study.

3.3. Clustering method

One of the most commonly used clustering algorithms in remote sensing applications is k-means because of its high performance, low complexity and the fact that it is implemented in most image packages. The k-means method partitions data into K given clusters, in which each observation belongs to the cluster with the nearest centroid (the mean value in each dimension). This process partitions the data space into Voronoi polygons and results in clusters of similar size.

This characteristic has strong implications when attempting to separate water from land pixels, because depending on the scene, the bands included in the analysis and the number of water pixels compared to land pixels can imply very different cluster sizes.

Agglomerative clustering, which was selected for this study, can address the difference in cluster sizes because it does not have any constraint with regard to the sizes of the resulting clusters. Agglomerative clustering is a subtype of hierarchical clustering that follows a bottom-up approach, in which each observation starts in its own cluster and is then merged iteratively until the desired number of clusters (K) is reached (Nielsen, 2016).

The key parameters for the algorithm to decide whether to merge two clusters are the metric (Eq. (2)) and the linkage (Eq. (3)). The metric specifies the measure of the distance between pairs of observations, and we used the simple Euclidean distance - ED, described in (Eq. (2)):

Table 3

Number of water mapping experiments performed in this study for each method.

Method	Dimensions	Number of experiments
Otsu Thresholding	1	3
Canny edge Thresholding	1	3
Clustering	2	10
Clustering	3	10
Clustering	4	5
Clustering	5	1

$$ED_{ab} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}, \quad (2)$$

where ED_{ab} is the distance between points a and b in an n -dimensional space and i is the considered dimension.

In addition to the metric, the linkage criteria determine how to compute the distance between clusters as a function of the pairwise distances between observations. The Average Linkage - AL, used in this study, considers the mean distance considering all the points in each cluster and can be described as in (Eq. (3)):

$$AL = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} ED_{ab}, \quad (3)$$

where a and b are the coordinates of points in a n -dimensional space and $|A|$ and $|B|$ are the total number of observations in each cluster.

During each iteration, the algorithm merges the two clusters, among all of the clusters, that are closest to each other considering the criteria described above. The iteration continues until the specified number K is reached.

3.4. Selecting the ideal number of clusters

Agglomerative clustering continues to merge clusters until a targeted number of clusters is achieved. Because the final objective is to develop a nonparametric fully automated method, one important step is to identify the best value of K . One way to evaluate the performance of a clustering process when the true classes are not known is to employ a coefficient that measures whether the clusters obtained are dense and well separated.

The coefficient used in this study is the Variance Ratio Criterion, which is also called the Calinski-Harabasz Index (Calinski and JA, 1974). The Variance Ratio Criterion considers intraclass (Eq. (4)) and inter-cluster (Eq. (5)) variances, represented by W_k and V , respectively, which are defined as S_{ch} in (Eq. (6)):

$$W_k = \frac{1}{|k|} \sum_{x \in k} ED_{x\mu_k}, \quad (4)$$

$$V = \sum_{k=1}^K k |ED_{\mu_k \mu}|, \quad (5)$$

$$S_{ch} = \frac{(N - K)V}{(K - 1) \sum_{k=1}^K W_k}, \quad (6)$$

where x is a point in cluster k , $|k|$ is the number of samples in cluster k , μ_k is the centroid of cluster k , μ is the centroid of all clusters, N is the total number of points, and K is the total number of clusters.

To achieve the best value of K , the Calinski-Harabasz Index is evaluated for multiple experiments with K ranging from 2 to 10. The superior limit of 10 was defined empirically in the context of this study. More information on the selection of the maximum number of clusters as well as the number of sampled pixels is proved in section 4.4.

3.5. Identifying the water cluster

Once the best number of clusters is known, the next step is to identify which cluster, among all possibilities, includes the water points. To achieve this goal, the MBWI is calculated for the centroids of each cluster. The cluster that presents the highest MBWI value is selected as the water cluster, and the others are labeled as nonwater clusters. The MBWI was selected due to its robustness over confused backgrounds, which could result in false positives with the other indices (Wang et al., 2016).

3.6. Pixel sampling and generalization

One disadvantage of agglomerative clustering compared to k -means is that it has a space complexity (the amount of memory needed to compute) of $O(n^3)$ and a time complexity of $O(n^3)$ (Firdaus and Uddin, 2015), which make it inefficient for even medium-sized datasets.

In addition, the clustering procedure is performed many times for each different K value until the method can obtain the best number of clusters. Considering that each Sentinel-2 scene includes approximately 120 million pixels (10,980 by 10,980 pixels) and that our objective is to provide an operational method for monitoring large areas, it would not be feasible to apply the process to the whole scene at once.

The proposed method, therefore, randomly selects a subset of pixels in the scene to apply the clustering until the best K is found and the water cluster is identified. Once the best clustering solution is achieved and the subset pixels are labeled for water and nonwater, we generalize the solution to all the pixels in the scene through a supervised machine learning classifier. After several tests, including Naïve Bayes, support vector machines and multilayer perceptron models, Naïve Bayes was selected because it offered the best results within a short training time. The schematic flow-chart of the processing chain is shown in Fig. 3.

Even though most of the processing is performed using a high-performance cluster, as mentioned in section 2.3, a whole Sentinel-2 scene can be classified in 3 to 6 min using an intel i7 computer with 32 Gb of memory running the Windows 10 Pro operating system.

3.7. Image thresholding

The most straightforward method for water body mapping is to apply a threshold to an index and select the pixels whose values are greater than a selected threshold value. A dynamically selected threshold is used as a standard because the values can vary both temporally and spatially among different regions, depending on the different image and water characteristics. Among the variety of methods for automating the threshold selection proposed in the literature (Al-Bayati and El-Zaart, 2013; Yang et al., 2017), the Otsu algorithm, developed initially to separate background from foreground in computer vision applications, is one of the most commonly used approaches. This method aims to maximize interclass variance and minimize intraclass variance based on the histogram of a variable. Considering σ to be the interclass variance between classes separated by a threshold t (Eqs. (7)–(9)), the algorithm exhaustively searches for the value of t that maximizes σ , as follows:

$$\sigma^2(t) = P_0(\mu_0 - \mu_T)^2 + P_1(\mu_1 - \mu_T)^2, \quad (7)$$

$$P = \frac{\sum_{\text{Pixels in class}}}{\text{Total Pixels}}, \quad (8)$$

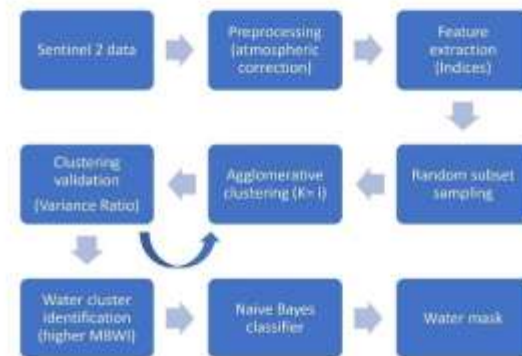


Fig. 3. Schematic flow diagram of the proposed method.

$$\text{Threshold} = \underset{\min(t) \leq t \leq \max(t)}{\text{ArgMax}} [\sigma^2(t)], \quad (9)$$

where P is the probability of a pixel belonging to one of the classes, μ_0 and μ_1 are the mean values of the variable for both classes, μ_T is the mean value of the whole histogram, and the threshold value is the value of t that maximizes σ .

To analyze the proposed multidimensional clustering method, we compare its results with those obtained by the plain Otsu's method applied to the three different indices, MNDWI, NDWI and MBWI. It has been shown, however, that this method yields unstable results if the amounts of water and land pixels are not equivalent and do not form a bimodal histogram (Donchyts et al., 2016).

To overcome these limitations, another set of experiments is performed using the improved Canny-edge method introduced by Donchyts et al. (2016), which first identifies the edges and then applies the Otsu algorithm on the pixels around those edges.

Both methods are implemented in Python, while taking advantage of the algorithm implementations available in the Scikit Image package (van der Walt et al., 2014). The implementation uses a mask for invalid pixels (clouds, no-data, shadows and snow). An iterative process is introduced to select the strongest edges. To provide similar conditions to the clustering technique, the final threshold is limited by preconfigured min/max values, differing, in this regard, from the original implementation.

4. Results

4.1. Multidimensional clustering

The kappa coefficients obtained from the best-performing combinations of indices/bands are presented in Table 4, as well as the average and standard deviation for all of the scenes. Within this subset, all the combinations achieved a mean kappa (calculated as the average kappa among the scenes) higher than 0.8 for all of the scenes.

There is variation of accuracy for clustering combinations with two, three or four dimensions, and therefore, no clear benefit is obtained by increasing the number of channels. In addition, all of the top combinations, including NDWI, MNDWI or both, corroborate with a previous study (Bangiza et al., 2019) regarding the capacity of these indices to correctly separate water and land pixels.

The most successful combination comprises NDWI and the B12 (SWIR) reflectance band; it achieved a mean kappa of 0.874, with a high score of 0.980 in the Bordeaux summer scene and good results in most areas. Even with a considerably high mean kappa and low standard deviation (std), the best combinations failed to produce reliable results

in Ariège, especially for the summer scene. In addition, all of the combinations worked comparably well in most scenes, and the major differences appeared on more challenging scenes, such as Ariège or Alpes, which include snow, a very mountainous area with topographic shadows and the lowest percentage of water surface. In such complex cases, the use of a higher-dimension feature space can increase performance, as is the case of the combinations (MNDWI, NDWI, B12, MBWI) and (NDWI, MBWI, B8, B12), which scored the best in Ariège winter (kappa = 0.780) and Ariège summer (kappa = 0.740), respectively. We also noted that the inclusion of the MBWI index improved the results, especially in the presence of very low reflectance waters.

The accuracies among the scenes varied substantially. Fig. 4 presents the mean kappa coefficient and the standard deviation for each area calculated over the best combinations. The water pixel percentage for each scene is shown in the right Y-axis (red line); it was calculated considering all of the unmasked pixels. There appears to be an inverse relationship between accuracy and water percentage. Clearly, the best accuracies were obtained from scenes with a water percentage above 1.5%. For the most challenging scenes (Ariège, Alsace and Alpes), only a few combinations were able to produce accurate results, which can be seen from the low mean and high standard deviation.

Examples of the clustering results, with the combination of NDWI and B12 channels, are shown in Fig. 5, as well as a comparison with the labels from the reference masks. The X-axis represents the reflectance, in sr^{-1} , of each pixel in the SWIR band (B12), and the Y-axis represents its NDWI value. The left column graphs show the pixels of the reference masks, where blue represents water pixels and green nonwater pixels. As expected, the water pixels are grouped in the higher NDWI and lower SWIR values due to the higher absorption of water in this wavelength.

The right column graphs show the results of the clustering algorithm for the same scenes. The pixel colors represent different clusters found throughout the clustering process. The best solution is achieved by testing a different fixed number of clusters (K), from two up to a predefined maximum, and selecting the value of K that maximizes the Variance Ratio index. It can be observed that the quantity of clusters varies depending on the overall characteristics of the scene, such as different coverage types, the presence of build-up areas and other features. Once the clusters are well-defined, the algorithm automatically identifies which cluster is the water cluster based on the higher MBWI value of the cluster centroids. The influence of the maximum number of clusters on the overall model accuracy is assessed in Section 5.1.

From the examples shown in Fig. 5, it can be observed that the NDWI (Y-axis) discriminates between land and water well, but the threshold value (lower boundary of the water pixels) can vary from almost 0, as in Alsace, to as low as -0.2, in Bordeaux. This finding agrees with those of other studies that found significant thresholding differences among

Table 4
Mean and standard deviation values of the kappa coefficients for the six best-performing clustering experiments. Best values per scene are highlighted in Bold.

Scene	NDWI B12	MNDWI NDWI B12	MNDWI NDWI B12 MBWI	MNDWI NDWI MBWI	NDWI MBWI B8 B12	MNDWI NDWI B8
Bordeaux Summer	0.980	0.984	0.984	0.953	0.894	0.921
Camargue Winter	0.970	0.966	0.960	0.966	0.969	0.893
Camargue Summer	0.977	0.969	0.912	0.912	0.976	0.924
Bordeaux Winter	0.925	0.968	0.879	0.957	0.960	0.935
Chateauxoux Summer	0.869	0.832	0.936	0.821	0.942	0.854
Alsace Summer	0.857	0.924	0.938	0.886	0.757	0.878
Alpes Summer	0.947	0.937	0.942	0.912	0.560	0.943
Alsace Winter	0.917	0.906	0.777	0.768	0.905	0.868
Chateauxoux Winter	0.887	0.843	0.901	0.876	0.753	0.733
Beauce summer	0.884	0.756	0.859	0.789	0.894	0.799
Normandie	0.866	0.915	0.919	0.815	0.733	0.662
Gironde	0.811	0.846	0.798	0.736	0.778	0.848
Beauce winter	0.856	0.813	0.794	0.685	0.847	0.792
Ariège winter	0.749	0.721	0.780	0.779	0.625	0.613
Ariège summer	0.608	0.545	0.405	0.512	0.740	0.508
Mean	0.874	0.862	0.852	0.823	0.821	0.812
Std	0.091	0.111	0.132	0.112	0.120	0.122

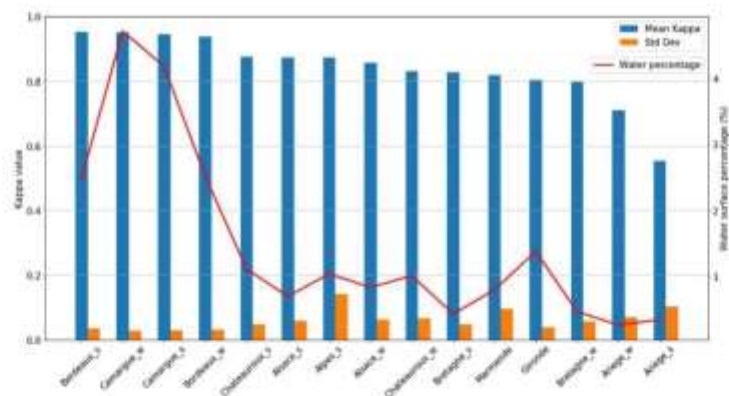


Fig. 4. Mean kappa score and standard deviation (left Y-axis) for each scene and comparison with the mean Water Area / Water Perimeter ratio (right Y-axis). In the scenes' labels, "x" stands for summer, and "w" stands for winter.

regions (Donchyts et al., 2016), which demonstrates the need for a dynamic method.

The fine separation ability provided by the NDWI, however, is not sufficient to correctly extract water bodies in some scenes. In scenes such as Bordeaux summer and Bretagne winter, for example, a single NDWI threshold would include many nonwater pixels (commission error) or miss many water bodies (omission error). The high variability of the NDWI in the Bordeaux region is due to the presence of water with a high suspended sediment matter (SPM) concentration (i.e., turbid waters), as is the case for the Garonne River. The higher the SPM concentration, the lower the NDWI index due to the increased reflectance response on the near-infrared band (B8).

Another issue is the presence of land pixels with a high NDWI response, which can occur in high albedo or shadow areas (Bangira et al., 2019; Feyisa et al., 2014; Wang et al., 2016). In such cases, the addition of complementing information, such as the SWIR band (B12), improves the results, as can be visually observed from the clustering results (Fig. 5).

4.2. Otsu and canny-edge Thresholding

To compare our results with those of previously published methods, we applied several water masking algorithms to the same scenes that we analyzed with the clustering technique. A simple thresholding experiment was conducted using the Otsu's algorithm in each of the three selected indices (MNDWI, NDWI and MBWI) with all of the pixels (no subsampling is necessary) while considering the reference mask to treat the clouds, shadows, snow and coastal waters. The results obtained from Otsu's thresholding algorithm, presented in Table 5, were rather disappointing because the model failed to produce reliable results in most scenes. These results could be explained, as previous studies have already noted (Donchyts et al., 2016), by the fact that in Otsu's algorithm, it is necessary to have a bimodal histogram for the independent variable to be able to produce reliable results.

To overcome this Otsu thresholding issue with a very low percentage of water pixels (not a bimodal histogram), we tested the same scenes using the method proposed by Donchyts et al. (2016), in which a bimodal histogram is obtained by selecting pixels around the contour between water and land through the use of the Canny-edge filter. With this approach, the results were greatly improved, but they are still not reliable for the most challenging scenes (Ariège and Alsace).

The Canny-edge algorithm using the MNDWI performed better in comparison with other thresholding experiments, with a mean kappa of 0.718 when considering all scenes. Plain Otsu's algorithm, without

adjustments for the bimodal histogram, is not a feasible option when processing large scenes. A similar problem can be observed with the MBWI index because it fails to provide a good separation between water and land in all of the scenes, regardless of whether the Canny-edge algorithm or Otsu's algorithm is used.

To further compare the different methods and to reduce the errors caused by the small number of water pixels, we tested a modification of the Otsu's algorithm, which is named here Modified Otsu (M-Otsu). M-Otsu takes advantage of using the same sampling method implemented for clustering, with minor adjustments. During the sampling phase, instead of selecting a random sample of pixels, the algorithm is modified to select equivalent amounts of pixels with high and low MNDWIs, defined as $MNDWI < 0$ (low) and $MNDWI > 0$ (high), and the threshold value is defined using this new set of pixels. The results were much improved, and the accuracy surpassed that of Canny-edge in all of the tested scenes. Table 6 presents the new results, marked as M-Otsu, and compares them with the results obtained by the best clustering and Canny-edge approaches.

With the improved method, Modified Otsu thresholding achieved better results than those for the simple thresholding, but the final mean accuracy of M-Otsu was lower than the clustering, with mean kappas of 0.807 and 0.874, respectively. When comparing the overall behavior of the methods (Fig. 6), the results obtained from the best clustering combinations presented better accuracies than those from thresholding, with a lower standard deviation. In addition, one can note that recall (also known as sensibility) is significantly higher than precision in all of the thresholding methods, especially for Canny-edge. This behavior can be explained because certain types of objects, such as roads, shadows, snow and some built-up areas, can have similar values to water bodies after index calculation due to similarities in their reflectance patterns (Feyisa et al., 2014; Wang et al., 2018; Yang et al., 2018). This characteristic leads to an overestimation of water surfaces with single indices, regardless of the selected threshold. This phenomenon can be more pronounced depending on the percentage of these confusion elements in the scene, and a single index thresholding might not be sufficient to discriminate between water and nonwater surfaces (Bangira et al., 2019; Feyisa et al., 2014).

For some scenes, especially those with larger water surfaces, the results are comparable, with the advantage that thresholding is faster and more straightforward. For more difficult scenes (i.e., a small number of water pixels, small water bodies and increased presence of confusing targets), increasing the number of channels in the clustering approach gives more stable and accurate results, such as the cases in which snow and topographic shadows are presented (mountainous scenes).

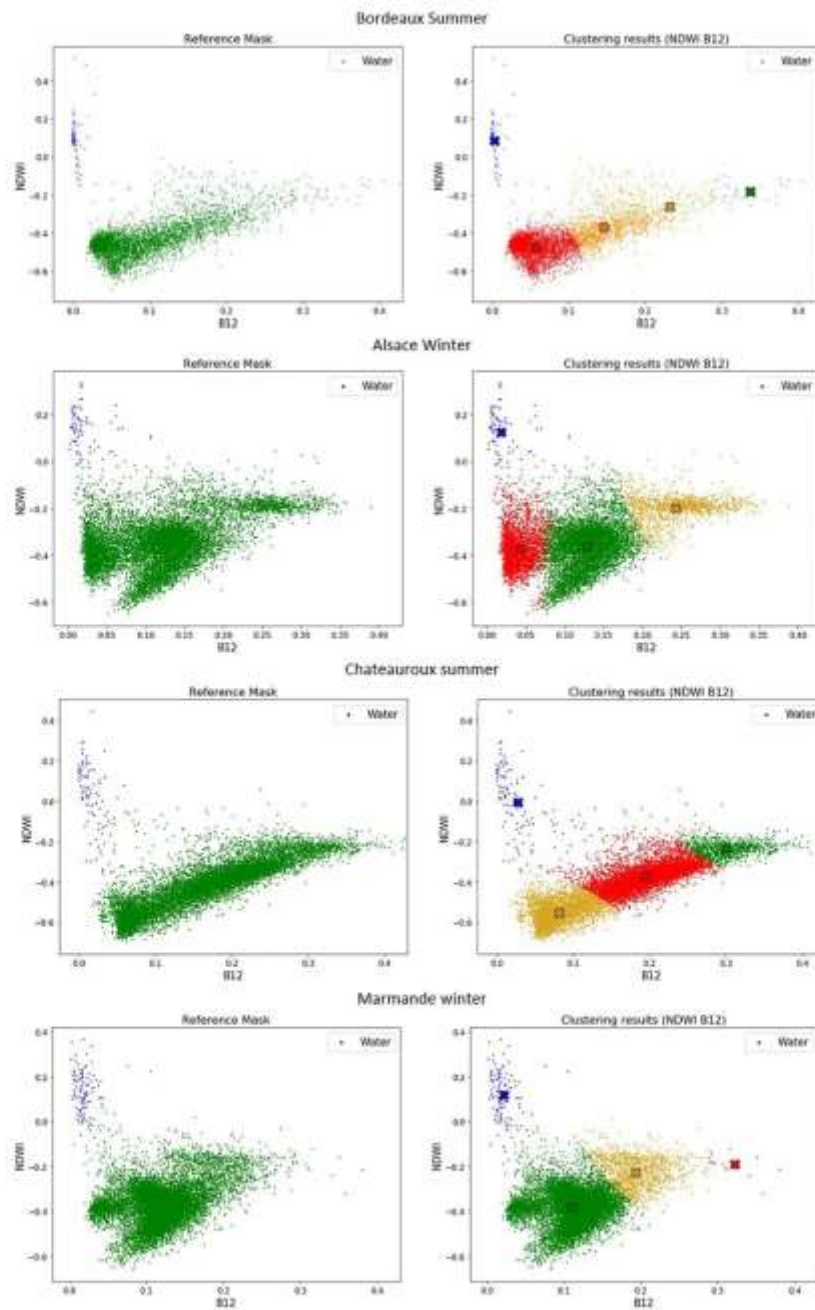


Fig. 5. Scatter plots of the NDWI (Y-axis) and B12 reflectance (X-axis) in sr^{-1} for four different scenes. The left column graphs represent the reference masks, in which the water pixels are plotted in blue and the nonwater pixels are in green. In the right column graphs, each colour represents a different cluster, and the colored crosses show the cluster centroid. The cluster identified as the water group is plotted in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Results (kappas) from Otsu's thresholding and modified Canny-edge thresholding on the MNDWI, NDWI and MBWI indices. Best values per scene are highlighted in Bold.

Scene	Canny MNDWI	Canny NDWI	Otsu MNDWI	Canny MBWI	Otsu NDWI	Otsu MBWI
Bordeaux summer	0.960	0.675	0.089	0.024	0.338	0.046
Camargue winter	0.938	0.883	0.904	0.179	0.467	0.121
Camargue summer	0.909	0.794	0.878	0.104	0.481	0.098
Bordeaux winter	0.962	0.834	0.091	0.029	0.234	0.041
Chateauxoux summer	0.829	0.535	0.045	0.504	0.017	0.023
Alsace summer	0.800	0.364	0.032	0.035	0.023	0.011
Alpes summer	0.503	0.163	0.061	0.043	0.085	0.030
Alsace winter	0.515	0.661	0.043	0.054	0.026	0.024
Chateauxoux winter	0.827	0.776	0.063	0.282	0.018	0.036
Bretagne summer	0.724	0.291	0.016	0.491	0.017	0.008
Marnande	0.853	0.714	0.038	0.128	0.020	0.025
Gironde	0.852	0.672	0.142	0.118	0.054	0.044
Bretagne winter	0.702	0.362	0.014	0.131	0.009	0.019
Ariège winter	0.136	0.177	0.023	0.009	0.004	0.009
Ariège summer	0.245	0.113	0.016	0.007	0.018	0.008
Mean	0.718	0.534	0.164	0.143	0.123	0.036
Std	0.248	0.256	0.287	0.137	0.166	0.031

4.3. Results for standard processors

As already mentioned, one of the objectives of this study is to provide a robust, nonparametric and fully automated process for identifying water pixels in large scenes. In this context, large-scale water masks for Sentinel 2 can be produced using ready-to-use multipurpose classification processors, such as Sen2Cor, MAJA and FMask. Sen2Cor and MAJA have been used as operational ground segments for Sentinel 2 images by ESA and CNES, respectively. In this case, the classified products were downloaded directly from the sites mentioned in section 2.2. FMask classification maps were produced running version 4.0, which is available online.

To compare the results from these processors to the reference masks, the multiclass classification outputs were transformed into a binary classification, with the water pixels being assigned a value of 1 and the nonwater pixels a value of 0. Because these processors are meant to

correctly classify clouds, shadows and snow, the only external no-data mask considered was the coastal line, which was obtained from CSSHG with a 400 m dilation toward the continent.

FMask and Sen2Cor compute the classification at a 20-m resolution, while MAJA, although it outputs the EDG layer at a 10-m resolution, produces the water mask at a 240-m resolution. To compare the results, the reference data are resampled to the resolution of the input mask at the moment of the comparison, and no-data pixels are excluded from the statistics. This procedure is performed automatically by the validation module of the surf-water system developed by CNES.

The results presented in Table 7 show a clear advantage for FMask and Sen2Cor over the MAJA processor, with mean kappas of 0.764, 0.726 and 0.355, respectively. These results could be explained by the lower spatial resolution used by MAJA to produce the water masks and the relatively high proportion of small water body areas found in most France scenes. This comparison shows that the clustering technique outperforms the other water masks on average and, more specifically, on the scenes that have the lowest accuracy levels. Among the four scenes with the lowest kappa scores, Ariège summer, Ariège winter, Bretagne summer and Chateauxoux summer only Ariège winter processed by Sen2Cor obtained a kappa value higher than 0.6 (kappa = 0.673). All of these scenes share the characteristics of having a small number of water pixels and small water body areas. The same scenes obtained 0.608, 0.780, 0.684 and 0.936, respectively, with the clustering analysis. The

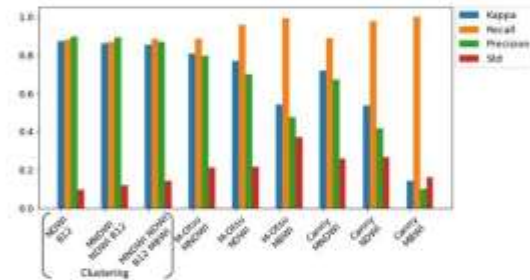


Fig. 6. Values of kappa, recall, precision and standard deviation for the clustering method (identified in red squares) considering different input channels, and for other methods tested, calculated for all scenes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Results (kappas) from Modified Otsu's thresholding and a comparison with Canny-edge thresholding on the MNDWI (Table 5) and the three best clustering combinations (Table 4). Best values per scene are highlighted in Bold.

Scene	M-Otsu MNDWI	M-Otsu NDWI	M-Otsu MBWI	Canny MNDWI	NDWI B12	MNDWI NDWI B12	MNDWI NDWI B12 MBWI
Bordeaux summer	0.984	0.967	0.982	0.968	0.980	0.984	0.984
Camargue winter	0.971	0.958	0.897	0.938	0.970	0.966	0.960
Camargue summer	0.966	0.947	0.888	0.909	0.977	0.969	0.912
Bordeaux winter	0.988	0.980	0.977	0.962	0.925	0.968	0.879
Chateauxoux summer	0.825	0.789	0.693	0.829	0.869	0.832	0.936
Alsace summer	0.870	0.713	0.188	0.808	0.857	0.924	0.938
Alpes summer	0.589	0.196	0.155	0.503	0.947	0.937	0.942
Alsace winter	0.642	0.815	0.116	0.515	0.917	0.906	0.777
Chateauxoux winter	0.821	0.838	0.508	0.827	0.887	0.843	0.901
Bretagne summer	0.841	0.746	0.843	0.724	0.884	0.756	0.859
Marnande	0.883	0.798	0.251	0.853	0.866	0.915	0.919
Gironde	0.963	0.905	0.847	0.852	0.811	0.846	0.798
Bretagne winter	0.909	0.750	0.700	0.702	0.856	0.813	0.794
Ariège winter	0.208	0.750	0.039	0.136	0.749	0.721	0.780
Ariège summer	0.652	0.388	0.043	0.245	0.608	0.545	0.405
Mean	0.807	0.769	0.542	0.718	0.874	0.862	0.852
Std	0.203	0.209	0.356	0.248	0.094	0.111	0.132

Table 7

Validation results (kappa) for the MAJA, FMask and Sen2Cor water masks compared with the results obtained from the three best clustering combinations (Table 4). Best values per scene are highlighted in Bold.

SCENE	FMask	Sen2Cor	MAJA	NDWI B12	MNDWI NDWI B12	MNDWI NDWI B12 MNDWI
Bordeaux summer	0.982	0.951	0.773	0.980	0.984	0.984
Camargue winter	0.968	0.919	0.661	0.970	0.966	0.960
Camargue summer	0.977	0.960	0.738	0.977	0.969	0.912
Bordeaux winter	0.977	0.965	0.797	0.925	0.968	0.879
Chateauroux summer	0.573	0.329	0.069	0.869	0.832	0.936
Alsace summer	0.908	0.815	0.123	0.857	0.924	0.930
Alpes summer	0.672	0.886	0.764	0.947	0.937	0.942
Alsace winter	0.903	0.752	0.280	0.917	0.906	0.777
Chateauroux winter	0.837	0.705	0.025	0.887	0.843	0.901
Bretagne summer	0.512	0.352	0.124	0.884	0.796	0.859
Marnande	0.772	0.678	0.014	0.866	0.915	0.919
Gironde	0.960	0.892	0.738	0.811	0.846	0.798
Bretagne winter	0.751	0.657	0.042	0.856	0.813	0.794
Ariège winter	0.443	0.673	0.149	0.749	0.721	0.780
Ariège summer	0.239	0.352	0.088	0.608	0.545	0.405
Mean	0.764	0.726	0.355	0.874	0.862	0.852
Std	0.223	0.217	0.317	0.094	0.115	0.137

largest differences were observed in Ariège summer and Chateauroux summer, which had 72.2% and 63.3% higher accuracies, respectively, with clustering compared to the best processor in each scene.

5. Discussion

5.1. Overall comparison

A visual comparison among the tested methods in the Camargue summer scene, near Avignon city, is shown in Fig. 7. This area comprises rivers that vary from 50 m to 450 m wide and have dense cities on their margins. We note that the Canny-edge MNDWI (Fig. 7-c) appears flooded compared to the reference mask, with many false-positives, especially in cities and other built-up areas. On the contrary, the three processors (Fig. 7-d, e, f) miss smaller water bodies and narrower river stretches. In this regard, the clustering technique represents a good compromise between precision and recall.

The results show that the overall performance of all of the tested methods is greatly influenced by the scene characteristics, such as the quantity and size of the water bodies and the presence of snow, mountains, shadows and other features. Despite being one of the most commonly used methods, Otsu thresholding achieved the poorest results; it only provided satisfactory results in the Camargue area, which has the highest fraction of water pixels (approximately 4–5%). Consequently, the Otsu thresholding method should be seen as a viable option for smaller study areas, with a well-balanced amount of water and land pixels, but it should not be employed as a standard when large scenes are considered as a whole.

Utilization of the Canny-edge filter to identify the edges before the use of Otsu thresholding can considerably improve results because it surpasses the limitation of nonbimodal histograms, but there are still areas where it fails to identify a threshold, such as the Ariège summer and winter, Alpes summer and Alsace winter scenes, in which the kappa

values were 0.245, 0.136, 0.503 and 0.515, respectively. The proposed M-Otsu method achieved kappa values of 0.652, 0.208, 0.589 and 0.642 in the same set of scenes, which shows its clear advantage over Canny-edge in these complex environments. In addition, the mean overall kappa was also higher (0.807) than that for Canny-edge (0.716), with a lower standard deviation (0.203 vs. 0.248).

For the three thresholding methods (M-Otsu, Canny-edge and Otsu), the MNDWI has been shown to produce better results, with greater discrimination of the classes. These results are comparable with those from Xu (2006) and Zhai et al. (2015), who found that the MNDWI performed substantially better than the other indices in extracting water bodies under different conditions, even considering that those studies used bands from Landsat satellites instead of the Sentinel 2 results presented here.

Among the processors, MAJA is the worst performer for water detection, most likely because of the lower spatial resolution of the water mask (i.e., 240 m) and the important areas of small water bodies in the selected scenes. These results contrast with those observed for cloud identification, where MAJA outperformed the other two processors (Baetens et al., 2019). The best overall processor performance was obtained for FMask. It delivered better kappa accuracies than Sen2Cor in twelve scenes, and its performance was only worse for the Alpes and Ariège scenes. To the best of our knowledge, no other study has been conducted comparing the performance of these processors with regard to water pixel extraction.

The results obtained using the proposed automated clustering were overall better than those obtained from the Level-2 processors or from the three thresholding approaches. The best configuration was obtained by combining the NDWI and the B12 band into a single bidimensional clustering. This combination achieved a mean kappa score of 0.874, which surpassed all the processors and thresholding procedures.

Comparing the results of the present study to those of previous studies, Zhai et al. (2015) evaluated their results from water body extraction using different indices based on Landsat-8 and Sentinel-2 imagery at two different sites, a city and village. In the city, the best kappa (0.68) was obtained using the AWEI, and for the village, the best kappa (0.89) was from the MNDWI. In both cases, the thresholds were selected manually. Acharya et al. (2018) evaluated the performances of the indices for water extraction in a 37,127.3 km² area in eastern Nepal using the Landsat 8 sensor. The area's elevation ranges from 60 m to 8848 m and contains flat lands and mid-hilly regions as well as rugged mountains with snow and glacial lakes. The optimal threshold was searched to provide a higher accuracy. The best kappa score was 0.596, which was obtained from the NDWI. Bangira et al. (2019) also compared the performance of the thresholding method using Sentinel-2 images across eight sites located in South African with a Mediterranean climate and areas ranging from 0.4 to 4.88 km². Different from the previously mentioned studies, the optimal threshold was obtained automatically using the standard Otsu algorithm. After analyzing more than two hundred different bands in combination, the best solution was achieved with the NDWI band, with kappa scores ranging from 0.78 to 0.92, and a mean kappa of 0.82 considering the eight sites. In the wetlands of Doñana, in Southwest Spain, Kordelas et al. (2018) proposed an automatic thresholding method for inundation mapping, using the MNDWI and SWIR bands, and compared it to a supervised approach. The unsupervised thresholding method resulted in a kappa score of 0.892 for the whole area, with 0.90 for recall and 0.88 for precision.

5.2. Influence of the quantity of sampling points and maximum clusters

Considering that the algorithm works initially with a subsample of the entire scene for unsupervised clustering and then generalizes results using a supervised machine learning technique (in this study, Naive Bayes), the results can be influenced by the number of random points selected initially. In addition, another important parameter for the algorithm is the maximum number of clusters to limit the unsupervised

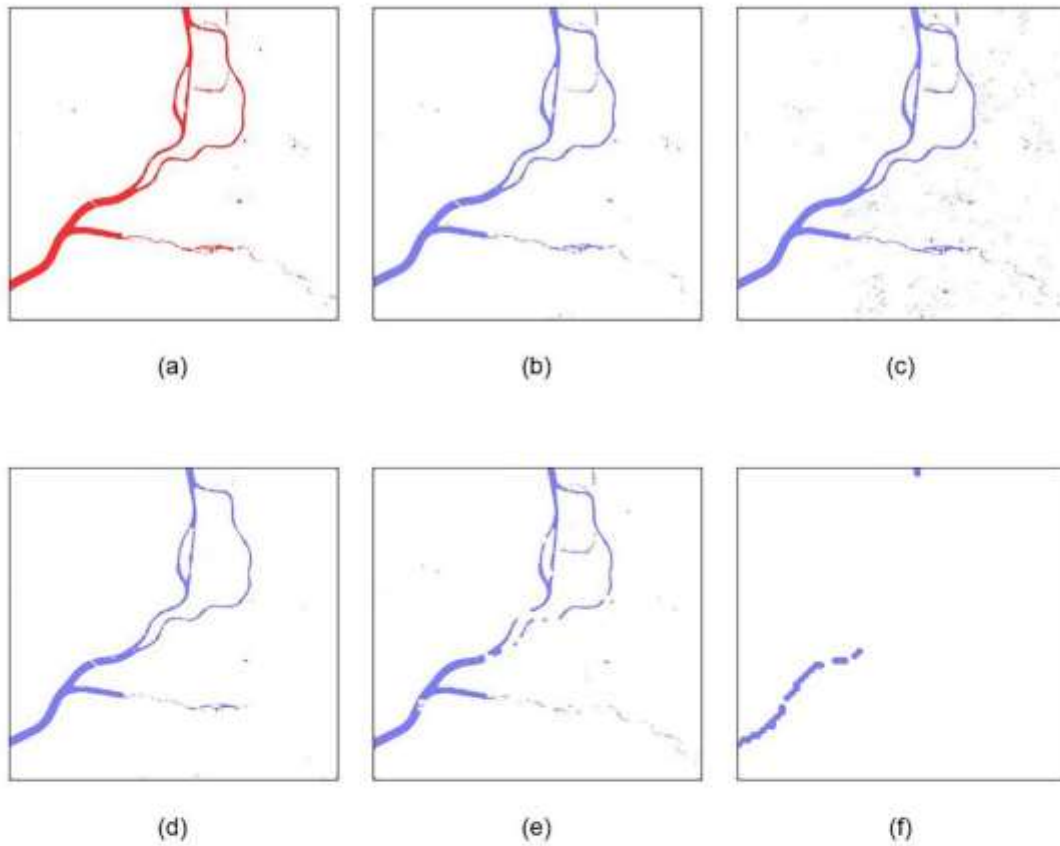


Fig. 7. Water segmentation comparison in a Camargue summer scene using the following methods: (a) reference mask; (b) clustering (NDWI x B12) result; (c) Canny-edge MNDWI thresholding; (d) Sen2Cor mask; (e) FMask mask; (f) MAJA mask.

step.

To assess the effects of the number of sampling pixels and the maximum limit on the number of clusters in the overall results, an additional experiment was conducted with the six best clustering combinations.

The mean kappa over all scenes and over the six best clustering combinations was evaluated for the number of sampling points varying from 1×10^3 to 25×10^3 , with steps of 2.5×10^3 , 5×10^3 , and 10×10^3 . The maximum number of allowed clusters was also fixed at $K = 5$ and $K = 15$ for testing purposes. Additionally, the time for processing a single

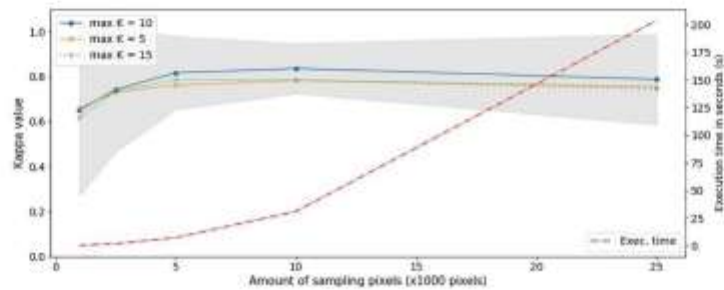


Fig. 8. Variation of the mean kappa according to number of sampling pixels and the maximum allowed number of clusters (K). The shaded area represents the standard deviation of the accuracy among the scenes when considering the max $K = 10$. The red line indicates the time for processing one single scene. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

scene was also measured according to the number of sampling points. The results were obtained running the algorithm on a node of CNES's cluster with 16 cores and 60 Gb of RAM memory. The results are presented in Fig. 8.

The figure shows that increasing the number of sampling pixels does not translate into a better overall accuracy and that the computation cost increases exponentially, as expected from the time complexity of agglomerative clustering, which is $O(n^2)$ (Pirdous and Uddin, 2015). On the other hand, a very small number of pixels is also prejudicial for the model to work because the sampling might not include all of the different pixel signatures presented in the scene. A maximum number of sampling pixels between 5×10^3 and 10×10^3 appears to provide a better trade-off between accuracy and computational cost.

With regard to the maximum number of clusters K , we note that there is an optimal number of clusters for the classification accuracy. Not enough clusters mean that the classes can be defined too roughly because each cluster can aggregate pixels with very different spectral behaviors. Too many clusters can lead to the creation of different groups of water pixels, resulting in significant errors because the algorithm looks for one unique water pixel cluster.

5.3. Influence of water body sizes on the clustering performance

To understand how the different methods are compared to each other according to the sizes of the water bodies in the scenes, we stratified the results by eight different classes, 0.0, 0.5, 1, 10, 50, 100, 500 and 1000 ha of water surface.

The water surface was estimated through a polygonization procedure

of the reference dataset, and each water body was accounted for separately to make an accuracy assessment. The water bodies from all of the scenes were grouped together in the analysis to obtain an overall mean for the size range across all scenes.

Another important remark is that no further adjustments were made to differentiate rivers from lakes or reservoirs. In such a scenario, it is possible to have a large water surface in a thin water body, which would affect the accuracy of the algorithms by the introduction of mixed pixels, for example.

The results from the stratified analysis are shown in Fig. 9. The mean kappa score is presented in panel (a). Panels (b) and (c) present the average precision and recall indices for the methods, respectively. It can be seen that there is a clear advantage for the clustering approach in every range. The Canny-edge method has the closest performance, followed by M-Otsu, FMask, Sen2Cor and MAJA, in that order. Considering water bodies of sizes between 1 and 10 ha, only clustering (0.826) and Canny-edge (0.809) were able to provide a kappa score above 0.8. In the size ranges from 0.5 to 1 ha and below 0.5 ha, clustering outperformed the second best method, Canny-edge, by 7% and 34%, respectively. All of the parameters that describe accuracy presented lower values for small water bodies, especially for a water surface area beneath 0.5 ha, which represents a 50-pixel area in a Sentinel-2 10-m resolution image. For that class, the clustering technique presented much better results than other techniques, with a mean kappa of 0.47, a mean recall of 0.56 and a mean precision of 0.5. These results are much better than the second-best technique, canny edge based on MNDWI images, with a mean kappa of 0.35, mean recall of 0.42 and mean precision of 0.37. Very small water bodies are especially challenging because all of the

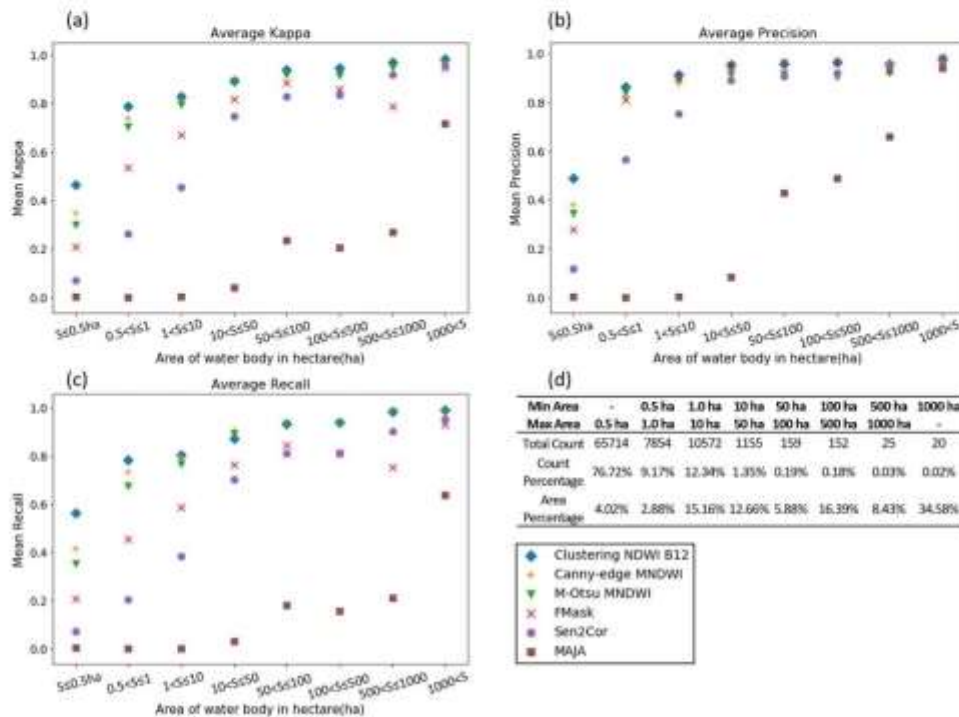


Fig. 9. Stratified results according to the area of the water body. Average kappa (a), precision (b) and recall (c) for clustering (NDWI B12 combination), Canny-edge (MNDWI), FMask, Sen2Cor and MAJA. Number of water bodies stratified by size and their respective percentages related to the total count and total area, in panel (d).

pixels are located at or near a shore. Consequently, water reflectance is affected more severely, proportionally, by adjacency effects compared with larger water bodies. Other artifact effects can also lead to errors in the mapping accuracy for such small water bodies: landscape shadows projected onto the water surface (terrain, vegetation), lake bottom visibility or reduced water/land edge visibility. It appears that clustering techniques make it possible to soften the impact of these artifacts because this technique provides much better results for the smallest water bodies than any other methods considered in this work. For the 0.5–1 ha water body size class, the clustering technique leads to an accurate mapping assessment ($\kappa = 0.79$, precision = 0.86, recall = 0.78), followed by Canny edge based on the MNDWI and the M-Otsu technique. Considering water bodies of sizes between 1 and 10 ha, only clustering and Canny-edge were able to provide a kappa score above 0.8. For recall (Fig. 9-c), clustering and Canny-edge performed very similar above 0.5 ha, while F-mask, Sen2cor and MAJA missed many water bodies below 50 ha. MAJA was the worst performer for identifying water bodies in all of the aspects analyzed due to its lower resolution.

5.4. Performance on challenging scenes

To understand the best performance achieved by multidimensional clustering over single band thresholding, Fig. 10 shows scatter plots with a comparison of the pixel separation of water and land pixels for these

methods and the reference mask in the Alsace winter scene. The red horizontal lines represent the best thresholding value for the index on the Y-axis, according to the M-Otsu method (the threshold for MNDWI is -0.033 , and the threshold for NDWI is -0.044).

It can be observed that regardless of whether the index is used for thresholding, there are nonwater pixels above the threshold (commission errors) and water pixels below the threshold (omission errors), which a single band cannot fully resolve. In such complex scenes, the use of clustering with the addition of more bands can improve the results.

The Ariège winter scene performed poorly with most algorithms and clustering combinations. A careful inspection shows that there is a large difference between the water signature of the water bodies in the mountainous region and the water bodies in the lowland area. The water reflectance in the green wavelength is of the order of 0.01 sr^{-1} in the mountainous lakes, which is mostly due to the very clear water with a low concentration of SPM in this region, whereas it reaches more than 0.06 sr^{-1} in flat areas as a result of higher SPM concentrations. Because the indices used for water detection are mostly a relationship between visible and near or middle infrared wavelengths, this large difference in the reflectance at the green band can result in two or more different water classes for the NDWI or MNDWI. Using the clustering method, the best accuracy for this scene was obtained using four dimensions including combination of the three indices, NDWI, MNDWI and MBWI, and the B12 band, and it achieved a kappa of 0.790. Although the NDWI

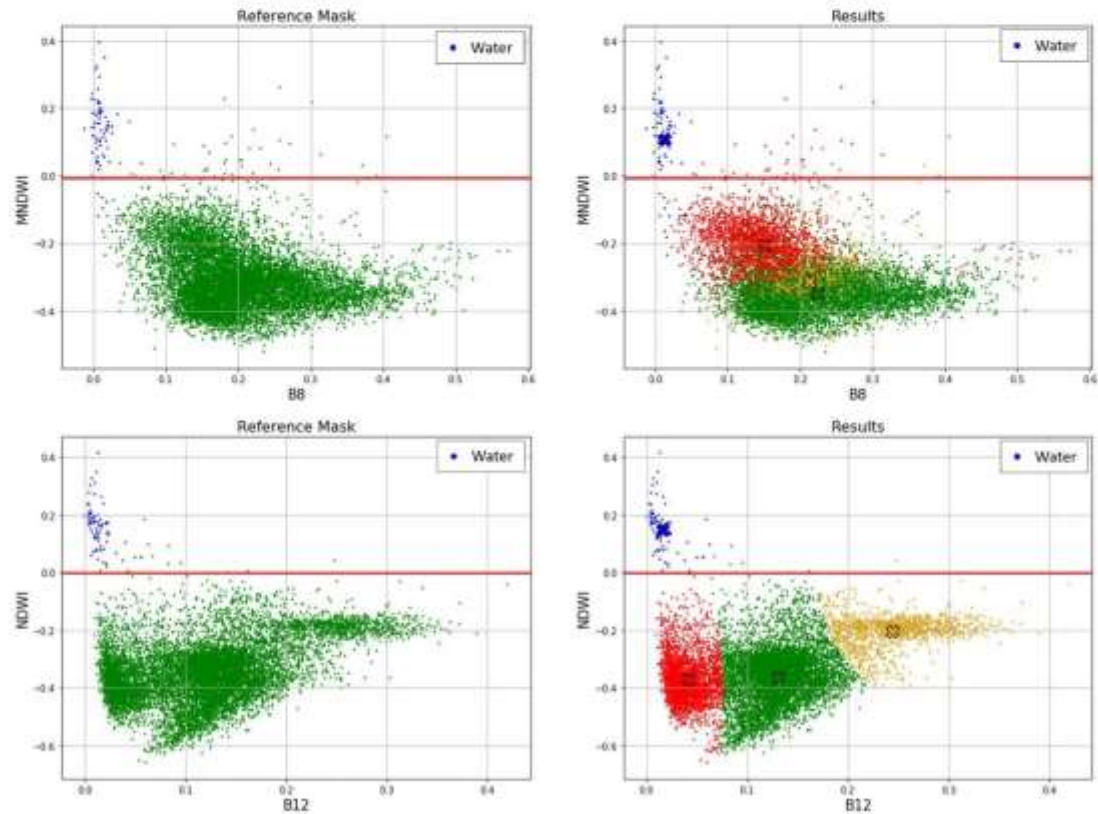


Fig. 10. Water segmentation comparison in the Alsace winter scene. The red horizontal lines represent the thresholding solution obtained by the M-Otsu method, and the pixel colors on the right panes are the clustering solution. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

B12 clustering combination can provide good results for most areas, this finding shows that no single combination fits the complexities of every scene and that sometimes it is necessary to use different combinations and even increase the number of clustering dimensions. Moreover, the clustering approach performs better in these mountainous areas and gives more flexibility by allowing the possibility of multiple combinations. A next step would be to develop an adaptive selection technique that can decide the ideal number of clustering dimensions based on some a priori parameters retrieved from the scene itself or from external knowledge. Information on the presence of snow, on solar zenith angle for shadow prediction or on the expected water coverage would allow to automatically assess the best clustering dimension and sampling size.

Although a direct comparison to previous studies is not straightforward, considering the differences in the study areas or satellite data, most unsupervised approaches were applied and validated over smaller regions or depend on subjectively assigned thresholds and ancillary data. In contrast, our method shows robustness (with a kappa standard deviation of 0.094 across all scenes) while considering the diversity of land coverage and environmental conditions (winter and summer scenes) in the present study. Consequently, our method can be easily applied to other regions without adaptation. The clustering technique coupled with a machine learning approach with random subsampling and generalization through a Naïve Bayes classifier presented an efficient running time (approximately 4 min for a Sentinel-2 tile). With regard to its disadvantages, the clustering analysis does not account for any special relationships among water pixels, such as spatial connectivity. Further testing against confusing targets, such as snow, clouds or landscape shadows, could allow us to fully assess the sensitivity of the method to extremely adverse conditions, such as lakes located in mountainous ranges covered by snow or ice or overcast by shadows.

6. Conclusions

Continuous monitoring of water surfaces is essential in many applications for water resource management, and the use of satellite images has been increasing with a higher availability of finer spatial and temporal resolution data.

This study proposed and tested a new method for the automation of water pixel classification of Sentinel-2 images that combines reflectance and water indices in a multidimensional clustering approach. Additionally, a machine learning method was proposed to subsample the whole image into a smaller classification set to make the process feasible for an entire Sentinel-2 scene. The method was tested on a set of 15 scenes from France, for which the reference masks are available for download (Santiago, 2019). The reference dataset had diverse coverage, including turbid water and some snow, which matched most European countries but did not consider extreme weather or complex canopy conditions, as in Sahelian areas, tundra plains or tropical forests. The validation process was performed at a 10-m spatial resolution, covering an inland area of more than 96,315 km², with an approximately 1239 km² water surface and more than 80,000 water bodies of different sizes and with a high percentage (76%) of water bodies smaller than 0.5 ha. Undesired pixels were removed according to established methodologies, as explained in section 2.2.

Comparing our results to the water classification performance of the three major Level-2 processors, namely, MAJA, Sen2Cor and FMask, the most common thresholding approaches showed that the proposed clustering method achieved higher accuracy results with a lower standard deviation, which indicate the better reliability of the proposed method across different scenes. Stratified analysis showed that the clustering method achieved the best mapping accuracy (i.e., kappa) for all water body size ranges from 0.5 ha to 100 ha. However, there was a quick drop in performance for water bodies that were smaller than 0.5 ha for all of the methods, and there were good results (kappa greater than 0.8) for water bodies above 1 ha with the clustering and thresholding methods. In addition, the newly proposed Modified Otsu

thresholding method produced good results with very little complexity and should be further evaluated to test its applicability in a wider range of situations. All of the tested methods, however, presented a lower accuracy rate in complex scenes with mountainous terrains, where the correct water identification remains a challenge, especially without the use of ancillary data.

This study demonstrates that adding more spectral information, as new dimensions, increases the water detection accuracy, especially for smaller water bodies and more challenging environments. From a total of 32 different combinations of spectral indices and bands over 15 scenes (a total of 480 experiments), the combination of the NDWI and the B12 band produced the best overall accuracy across all of the considered scenes, and a four-dimensional combination including NDWI, MNDWI, MBWI, and B12 improved the most difficult scenes.

In summary, the proposed clustering method shows promise in automating water pixel extraction in large scenes from optical images without need for ancillary or pretrained data. Moreover, the subsampling approach leads to a gain in performance and makes it feasible to apply complex algorithms to large amounts of data without a penalty in accuracy. Currently, the use of good input masks for clouds, shadows, snow and coastal water is still critical for current methods to avoid misclassification, and future studies should be conducted to evaluate the methods on images while considering these elements.

The code necessary to run the algorithm is available for download at <https://github.com/corineaur/WaterDetect>.

Funding

This research received no external funding.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to acknowledge the technical and financial support from the French Spatial Agency CNES (OBS2CO and SWOT-AVAL projects) for the realization of this study.

References

- GDAL/CGI contributors, 2020. GDAL/GRASS Geospatial Data Abstraction Software Library (Open Source Geospatial Foundation).
- Acharya, T.D., Lee, D.H., Yang, I.T., Lee, J.K., 2016. Identification of water bodies in a Landsat 8 OLI image using a J48 decision tree. *Sensors* 16, 1075. <https://doi.org/10.3390/s16071075>.
- Acharya, T.D., Subedi, A., Lee, D.H., 2018. Evaluation of water indices for surface water extraction in a Landsat 8 scene of Nepal. *Sensors (Basel)* 18. <https://doi.org/10.3390/s18092580>.
- Al Bayati, M., El Zant, A., 2013. Automatic thresholding techniques for SAR images. In: *Computer Science & Information Technology*. In: Presented at the First International Conference on Computational Science and Engineering, Academy & Industry Research Collaboration Center (AIRCC), pp. 75–84. <https://doi.org/10.5121/1081-2013.33006>.
- Anser, A., Alikov, K., 2018. Retrieval of chlorophyll a from Sentinel 2 MSI data for the European Union water framework directive reporting purposes. *Remote Sens.* 11, 64. <https://doi.org/10.3390/rs11010064>.
- Baeteis, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel 2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sens.* 11, 433. <https://doi.org/10.3390/rs11040433>.
- Bangira, T., Alfieri, S.M., Menenti, M., van Nieuwen, A., 2019. Comparing thresholding with machine learning classifiers for mapping complex water. *Remote Sens.* 11, 1351. <https://doi.org/10.3390/rs11111351>.
- Barbosa, C.C.F., Netto, F.M.L. de M., Martins, V.S., 2019. *Introdução Ao Sensoriamento Remoto De Sistemas Aquáticos: Princípios E Aplicações* (Instituto Nacional de Pesquisas Espaciais).

- Bakata, R.P., 2013. Retrospection and introspection on remote sensing of inland water quality: Lake Deijia Vu All Over Again. *J. Great Lakes Res.* Remote Sensing Great Lakes Other Inland Waters 39, 2–5. <https://doi.org/10.1016/j.jglr.2013.04.001>.
- Bana, W.G., Lee, S.-I., Seo, J.-Y., 2018. Recent surface water extent of Lake Chad from multispectral sensors and grc. *Sensors* 18, 2092. <https://doi.org/10.3390/s18072092>.
- Calliński, T., Ja, H., 1994. A dendrite method for cluster analysis. *Commun. Stat.* 3, 1–27. <https://doi.org/10.1080/0162145927488827101>.
- Camê, R.C., Martínez, J. M., Pesosito, M.A., Villar, R., Cochenneau, G., Henry, R., Lopes, W., Nagata, M., 2019. Indirect assessment of sedimentation in hydroelectric dams using MODIS remote sensing images. *Remote Sens.* 11, 314. <https://doi.org/10.3390/rs11030314>.
- Delegido, J., Terço, C., Ruiz Verdú, A., Peña, R., Moreno, J., 2014. Modelo empírico para la determinación de clorofila *a* en aguas continentales a partir de los datos Sentinel 2 y 3. Validación con imágenes HICO. *Rev. Teledetección* 37. <https://doi.org/10.4995/raei.2014.2295>.
- Dinh Ngoc, D., Loisel, H., Janet, C., Vantrepote, Y., Dufort-Gaurier, L., Minh, C., Marzin, A., 2019. Coastal and inland water pixels extraction algorithm (WPG) from spectral shape analysis and HSV transformation applied to Landsat 8 OLI and Sentinel 2 MSI. *Remote Sens. Environ.* 223, 18. <https://doi.org/10.1016/j.rse.2019.01.024>.
- Douchya, G., Schellekens, J., Wismans, H., Eisenmann, E., Van de Giesen, N., 2016. A 30 m resolution surface water mask including estimation of positional and thematic differences using landsat 8, arctic and OpenStreetMap: a case study in the Murray-Darling basin, Australia. *Remote Sens.* 8, 386. <https://doi.org/10.3390/rs8060386>.
- Du, Y., Zhang, Y., Ling, F., Wang, Q., Li, W., Li, X., 2016. Water bodies' mapping from Sentinel-2 imagery with modified normalized difference water index at 10 m spatial resolution produced by sharpening the SWIR band. *Remote Sens.* 8, 354. <https://doi.org/10.3390/rs8040354>.
- Feng, Q., Gong, J., Liu, J., Li, Y., 2015. Flood mapping based on multiple random forest spectral mixture analysis and Random Forest classifier—the case of Yuzao, China. *Remote Sens.* 7, 12539–12562. <https://doi.org/10.3390/rs70912539>.
- Feng, M., Sexton, J.O., Chinnon, S., Townshend, J.R., 2016. A global, high-resolution (30-m) inland water body dataset for 2000: first results of a topographic-spectral classification algorithm. *Int. J. Digital Earth* 9, 113–133. <https://doi.org/10.1080/17513759.2015.1026420>.
- Feyssa, G.L., Meilly, H., Frensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: a new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* 140, 23–35. <https://doi.org/10.1016/j.rse.2013.08.025>.
- Finkels, S., Urbán, A., 2015. A survey on clustering algorithms and complexity analysis. *Int. J. Comp. Sci. Issues* 12, 24.
- Franpton, W.J., Dadd, J., Witkowski, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS J. Photogramm. Remote Sens.* 82, 83–92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>.
- Gascon, S., Grizonnet, M., Bouchet, M., Salgues, G., Hagolle, O., 2019. The Snow collector: high-resolution operational snow cover maps from Sentinel-2 and Landsat-8 data. *Earth System Science Data* 11, 493–514. <https://doi.org/10.5194/essd-11-493-2019>.
- Hagolle, O., Hue, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENUS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* 114, 1747–1755. <https://doi.org/10.1016/j.rse.2010.03.002>.
- Hollstein, A., Segl, K., Guarnier, L., Brell, M., Enecco, M., 2016. Ready-to-use methods for the detection of clouds, cirrus, snow, shadow, water and clear sky pixels in Sentinel-2 MSI images. *Remote Sens.* 8, 666. <https://doi.org/10.3390/rs8060666>.
- Ji, L., Zhang, L., Wylie, B., 2009. Analysis of dynamic thresholds for the Normalized Difference Water Index. *PEARS* 75, 1307–1317. <https://doi.org/10.14358/PEARS.75.11.1307>.
- Jiang, W., He, G., Long, T., Ni, Y., Liu, H., Peng, Y., Lv, K., Wang, G., 2018. Multilayer perceptron neural network for surface water extraction in Landsat 8 OLI satellite images. *Remote Sens.* 10, 755. <https://doi.org/10.3390/rs10050755>.
- Kaplan, G., Avdan, U., 2017. Object-based water body extraction model using Sentinel-2 satellite imagery. *Environ. J. Remote Sensing* 50, 137–143. <https://doi.org/10.1080/22797254.2017.1297590>.
- Ko, B.C., Kim, H.H., Nam, J.Y., 2015. Classification of potential water bodies using Landsat 8 OLI and a combination of two Boosted Random Forest classifiers. *Sensors* 15, 13763–13777. <https://doi.org/10.3390/s150613763>.
- Kordelas, G.A., Manakos, I., Aragonés, D., Díaz-Delgado, R., Bustamante, J., 2018. Fast and automatic data driven Thresholding for inundation mapping with Sentinel-2 data. *Remote Sens.* 10, 910. <https://doi.org/10.3390/rs10060910>.
- Kordelas, G.A., Manakos, I., Lefebvre, G., Poulin, B., 2019. Automatic inundation mapping using Sentinel-2 data applicable to both Camargue and Doñana biosphere reserves. *Remote Sens.* 11, 2251. <https://doi.org/10.3390/rs11102251>.
- Lima, R.C., Martínez, J. M., Motis Marqués, D.D., Cirilo, J.A., Pragnes, C.R., 2017. Assessment of chlorophyll *a* remote sensing algorithms in a productive tropical estuarine-lagoon system. *Remote Sens.* 9, 516. <https://doi.org/10.3390/rs9060516>.
- Merkert, K.N., Chidiste, F., Anderson, E.B., Saah, D., Griffin, B.E., 2019. On the merging of optical and SAR satellite imagery for surface water mapping applications. *Results Phys.* 9, 275–277. <https://doi.org/10.1016/j.rinp.2018.02.054>.
- Martínez, J.M., Guyot, J.L., Fillola, N., Soudag, F., 2009. Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data. *CATENA* 79, 257–264. <https://doi.org/10.1016/j.catena.2009.05.011>.
- Martins, S., Twele, A., Voigt, S., 2011. Unsupervised extraction of flood induced backscatter changes in SAR data using Markov image modeling on irregular graphs. *IEEE Trans. Geosci. Remote Sens.* 49, 251–263. <https://doi.org/10.1109/TGRS.2010.2072816>.
- McFeeters, S.R., 1996. The use of the normalized difference water index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* 17, 1425–1432. <https://doi.org/10.1080/01431159608994674>.
- Mishra, K., Prasad, P.R.C., 2015. Automatic extraction of water bodies from Landsat imagery using perceptron model. *J. Computat. Environ. Sci.* 2015, 1–9. <https://doi.org/10.1155/2015/903465>.
- Mueller-Wilm, U., Desjardis, O., Pesenti, L., 2019. *Scikit-learn Configuration and User Manual*.
- Nandi, L., Srivastava, P.K., Shit, K., 2017. Floodplain mapping through support vector machine and optical/infrared images from Landsat 8 OLI/TIRS sensors: case study from Yamunotri. *Water Resour. Manag.* 31, 1157–1171. <https://doi.org/10.1007/s11269-017-1569-y>.
- Nielsen, F., 2016. Hierarchical Clustering. in: *Introduction to HPC with MPI for Data Science*. Springer International Publishing, Cham, pp. 195–211. https://doi.org/10.1007/978-3-319-21903-5_8.
- Olivi, N., 1979. A threshold selection method from gray level histograms. *IEEE Trans. Syst. Man Cybernet.* 9, 62–66. <https://doi.org/10.1109/TSMC.1979.4310876>.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, Y., Vanderplas, J., Passos, A., Cournapeau, D., 2011. Scikit-learn: machine learning in Python. *JMLR* 12, 2825–2830.
- Pekel, J.-F., Vancevic, C., Bestin, L., Clerici, M., Vanbogaert, E., Bartholomé, E., Defourny, P., 2014. A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data. *Remote Sens. Environ.* 140, 704–716. <https://doi.org/10.1016/j.rse.2013.10.030>.
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High resolution mapping of global surface water and its long-term changes. *Nature* 540, 418–422. <https://doi.org/10.1038/nature20584>.
- Pi, F., Ding, C., Chao, Z., Yu, Y., Xu, X., 2019. Water quality classification of Inland Lakes using Landsat-8 images by convolutional neural networks. *Remote Sens.* 11, 1674. <https://doi.org/10.3390/rs11111674>.
- Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing Environ.* 231, 111205. <https://doi.org/10.1016/j.rse.2019.05.034>.
- Santiago, P.L., 2019. CMES ALCD Open water masks (Version 1.1) [Data set]. <https://doi.org/10.5281/zenodo.3522960>.
- Shen, X., Wang, D., Mao, K., Anagnostou, E., Hong, Y., 2019. Inundation extent mapping by synthetic aperture radar: a review. *Remote Sens.* 11, 879. <https://doi.org/10.3390/rs11070879>.
- Souza, C., Kirchhoff, T., Oliveira, B., Ribeiro, J., Sales, M., 2019. Long-term annual surface water change in the Brazilian Amazon basin: potential links with deforestation, infrastructure development and climate change. *Water* 11, 566. <https://doi.org/10.3390/w11030566>.
- Schler, Hansch, B., 2015. *Sentinel-2 User Handbook*.
- Tuning, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nöges, T., 2016. First experiences in mapping Lake water quality parameters with Sentinel-2 MSI imagery. *Remote Sens.* 8, 640. <https://doi.org/10.3390/rs8080640>.
- Verpoorter, C., Kutser, T., Soekel, D.A., Tranvik, L.J., 2014. A global inventory of lakes based on high-resolution satellite imagery. *Geophys. Res. Lett.* 41, 6396–6402. <https://doi.org/10.1002/2014GL060661>.
- van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Goullard, E., Yu, T., 2014. Scikit-image: image processing in Python. *PeerJ* 2, e453. <https://doi.org/10.7717/peerj.453>.
- Wang, X., Xie, S., Zhang, X., Chen, C., Guo, H., Du, J., Duan, Z., 2018. A robust multi-band water index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. *Int. J. Appl. Earth Obs. Geoinf.* 68, 73–91. <https://doi.org/10.1016/j.isprsjprs.2018.01.018>.
- Wang, G., Wu, M., Wei, X., Song, H., 2020. Water identification from high-resolution remote sensing images based on multidimensional densely connected convolutional neural networks. *Remote Sens.* 12, 795. <https://doi.org/10.3390/rs12050795>.
- Wesell, P., Smith, W.J.F., 1996. A global, self-consistent, hierarchical, high-resolution shoreline database. *J. Geophys. Res.* 101, 8741–8743. <https://doi.org/10.1029/96JB00104>.
- Wieland, M., Martins, S., 2019. A modular processing chain for automated flood monitoring from multi-spectral satellite data. *Remote Sens.* 11, 2330. <https://doi.org/10.3390/rs11192330>.
- Xu, H., 2006. Modification of normalized difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* 27, 3025–3033. <https://doi.org/10.1080/01431150600689179>.
- Yader, S., Yamashiki, Y., Suzuki, J., Yamashita, Y., Ishikawa, K., 2019. Chlorophyll estimation of lake water and coastal water using Landsat-8 and Sentinel-2A satellite. *Int. Arch. Photogramm. Remote Sens. Spatial. Inf. Sci. XLII-3/W7*, 77–82. <https://doi.org/10.5194/isprs-archives-XLII-3-W7-77-2019>.
- Yang, X., Zhao, S., Qin, X., Zhao, N., Liang, L., 2017. Mapping of urban surface water bodies from Sentinel-2 MSI imagery at 10 m resolution via NDWI based image sharpening. *Remote Sens.* 9, 596. <https://doi.org/10.3390/rs9060596>.
- Yang, X., Qin, Q., Grunsemayr, P., Kuehl, M., 2018. Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery. *Remote Sens. Environ.* 219, 259–270. <https://doi.org/10.1016/j.rse.2018.07.016>.
- Yepez, S., Laraque, A., Martínez, J. M., De Sa, J., Garreta, J.M., Castellanos, B., Galley, M., Lopez, J.L., 2018. Retrieval of suspended sediment concentrations using Landsat 8 OLI satellite images in the Orinoco River (Venezuela). *Complexity: Earth and Environmental Sciences, Rivers of the Andes and the Amazon Basin: Deciphering global change*

- from the hydrodynamic variability in the critical zone 350, 29–30. <https://doi.org/10.1016/j.rse.2017.08.004>.
- Yousef, P., Jalab, H.A., Ibrahim, F.W., Noor, N.F.M., Ayub, M.N., Gani, A., 2018. Water-body segmentation in satellite imagery applying modified kernel k-means. *Malaysian J. Comput. Sci.* 31, 143–154. <https://doi.org/10.22462/mjcs.vol31no2.4>.
- Zhai, K., Wu, X., Qiu, Y., Du, P., 2015. Comparison of surface water extraction performances of different classic water indices using OLI and TM images in different situations. *Geo spatial Information Science* 18, 32–42. <https://doi.org/10.1080/10095820.2015.1017911>.
- Zhang, F., Li, J., Zhang, B., Shen, Q., Ye, H., Wong, S., Lu, Z., 2018. A simple enhanced dynamic threshold extraction method for the classification of large water bodies from Landsat 8 OLI water index images. *Int. J. Remote Sens.* 39, 3429–3451. <https://doi.org/10.1080/01431161.2018.1444292>.
- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sens. Environ.* 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.026>.
- Zhu, Z., Wang, S., Woodcock, C.E., 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. *Remote Sens. Environ.* 159, 269–277. <https://doi.org/10.1016/j.rse.2014.12.014>.

Chapter II - Optical Water Classification

1. Article Draft - Optical water type classification for suspended Particulate Matter Retrieval Over inland tropical waters

1.1 Abstract

Monitoring suspended sediment in river systems evidences a wide range of environmental issues. However, conventional measurement networks need to improve in most rivers to provide an adequate spatial and temporal sampling frequency. To support the operational use of water color remote sensing data for suspended sediment monitoring over inland waters, there is a lack of systematic studies based on field measurements making possible to fully study the dependency of the remote sensing reflectance to the suspended sediment characteristics in relation to their watershed. In the present study, an exceptional dataset of 1,067 hyperspectral measurements collected in 13 different watersheds in the tropics was collected jointly with water sampling representing a wide range of suspended sediment matter (SPM) concentration levels from 0.1 to 1,800 mg/l. The article analyzed if retrieval models adjusted for generic optical water classes can outperform retrieval models tailored to rivers or catchments. Both techniques for optical water class definition (clustering) and membership functions (that match reflectance samples to pre-defined spectral groups) were evaluated to understand how different methods may result in distinct water classes definition. Different combinations of clustering algorithms (e.g., K-means, Agglomerative, and Fuzzy C-Means), distance metrics (e.g., Mahalanobis, Euclidian) and input features (e.g., reflectance and normalized reflectance) were tested to achieve the best overall SPM concentration retrieval. Multiple optical water types were considered starting with a simple 2-class typology (low and high SPM concentration), catchment-based classes, optical water types previously proposed in the literature, and with a varying number of optical clusters up to 10 different groups. As a result,

a four-cluster classification was a reasonable cutoff, as there was no significant advantage in accuracy by augmenting the number of clusters beyond this limit. This four-cluster classification reaches the performance of the catchment-based classification with $R^2 = 0.93$, RMSE = 86.08 mg/l, and RMSLE = 0.46 mg/l. The influence of the spectral resolution was also analyzed, demonstrating that multispectral data such as Sentinel-2 multispectral instrument (MSI) sensor configuration makes it possible to deliver fine precision without significant accuracy loss in relation to the hyperspectral dataset. The spectral bands in the 700-800 nm range of the Sentinel-2/MSI sensor were the most frequent bands used by the best retrieval models, independently of the optical water type classification. This work paves the way for the global and systematic use of satellite water color data for turbidity and suspended sediment monitoring in rivers, lakes and reservoirs, supporting a large range of environmental studies in catchments.

1.2 Introduction

Besides the continuously increasing demand for water, due to population growth and the increase in the standard of living, half of the inland available water may be polluted (A. K. Makarigakis and Jimenez-Cisneros, 2019). Therefore, surface water monitoring is critical to ensure efficient water resources management, but field monitoring for many countries can be challenging to maintain, considering the high costs involved (Barbosa et al., 2019).

The increasing capacity of remote sensing earth observation satellites, which now offer high spatial, spectral, and temporal resolutions at a global scale without cost, enables researchers to assess surface water quality over inland remotely. There is a large amount of research on the analysis of optical properties for phytoplankton assessment (Ansper and Alikas, 2018; Cairo et al., 2020; Delegido et al., 2014; Toming et al., 2016; Yadav et al., 2019), but the analysis of sediment fluxes using remote sensing has been relatively less studied in fluvial systems (Condé et al., 2019;

Martinez et al., 2009; Yepez et al., 2018). Besides, the high interest in quantifying sediment fluxes from the continent to the oceans, monitoring the fine fraction of the inorganic particulate matter, such as clays, is crucial for many environmental studies. Nutrients, bacteria, or heavy metals are adsorbed on the clays during their transport in the water column in rivers and lakes, which means that assessing suspended sediment concentration and its particle type is a valuable information for predicting pollution diffusion in watersheds.

Various methodologies have been proposed for studies of suspended sediment recovery: simple reflectance spectral relationship, hybrid models, deterministic models, semi-analytical models, bio-optical models and approach to optical water typologies.

Kirk (1994) stated that there is no universal retrieval model for suspended sediment as the complex dependency of the inherent optical properties to the particles' number, size distribution, and type (i.e., mineralogical composition and relative organic fraction) may not be resolved by a simple set of equations. Later, several studies (Doxaran et al., 2002; Espinoza Villar et al., 2013; Moore et al., 1999) demonstrated that using a simple reflectance spectral ratio, mainly between infrared and red wavelengths, is highly efficient in normalizing these multiple dependencies to the sediment size and type with a series of studies on coastal and inland waters. In the last years, hybrid models (Han et al., 2016; Novoa et al., 2017) have been proposed by triggering different models as a function of reflectance level in order to select optimal models using short wavelength for low concentrations (from green to red bands) and models integrating NIR bands for higher concentrations. Although several hybrid algorithms have been developed to retrieve water quality parameters from earth observation satellites, they have limited applicability in space and time. Furthermore, the limitations of these retrieval algorithms across different optical complex systems are often not considered (Neil et al., 2019). Consequently, the concern expressed by Kirk, three

decades ago, remains open and needs to be addressed as there is a lack of systematic studies on rivers, which are the primary vector for suspended sediment transport.

There is a wide range of inversion model types, allowing to retrieve SPM from apparent and/or inherent optical properties. Strictly deterministic models take into account all the processes involving light interaction with the atmosphere and water column but are strongly dependent on the definition of a large number of input parameters that can be unsustainable for calibration in natural water environments. For suspended particles, the main difficulty is to provide realistic values for suspended size distribution or for real and complex refractive index that may reproduce fairly the natural variability of the suspended sediment fluxes in inland waters (Pinet et al., 2017; Woźniak and Stramski, 2004). Semi-analytical models have also been proposed with fine agreement when compared to field surveys, but on local scales or for a limited number of locations, especially for inland waters, making it difficult to define the applicability range of these algorithms (Xue et al., 2019).

Bio-optical models for SPM retrieval vary in terms of mathematical expressions or spectral bands, depending on the water type, SPM range, and geographic region or sensor design (Chen et al., 2015; Long and Pavelsky, 2013). Doxaran et al. (2002) analyzed the spectral signature of highly turbid waters and demonstrated no acceptable correlation for the whole concentration range using a single model. Bands in the visible range (i.e., 500-590 nm and 610-680nm) performed better for measuring low concentrations, with a linear relationship observed in the lowest concentrations ($\text{SPM} < 50 \text{ mg l}^{-1}$), but a saturation effect limited its use for higher concentrations. NIR reflectance (790-890 nm) and the ratios NIR/Visible are usually used for higher concentrations,. Long and Pavelsky (2013) compared the results of 35 empirical models for turbid coastal and inland waters

and concluded that multispectral models are preferable for remotely sensing SPM and suggested the combination of near-infrared and visible bands.

However, the SPM concentration is not the only variable affecting the complex interactions between optically active components, which can also change according to particle size distribution and other constituents, such as organic matter content. In this regard, many studies have shown that Optical Water Typologies (OWTs) can be used to apply tailored algorithms to improve overall inversion accuracy (Bi et al., 2019; Moore et al., 2014; Neil et al., 2019; Xue et al., 2019). Posterior to the commonly used Case 1 or Case 2 twofold classification scheme (Morel and Prieur, 1977), that separated complex water according to the predominance of unique optically active components - OAC (notably photosynthetic pigments) or of a couple of OACs simultaneously, several subsequent studies have proposed OWT to describe waters in optically complex environments. Most of these studies have focused on describing coastal and marine systems (Lubac and Loisel, 2007; Moore et al., 2014, 2009; Vantrepotte et al., 2012), with only a few on inland systems (Spyrakos et al., 2018)

Several clustering techniques have been implemented to classify the reflectance spectra into meaningful group types, such as k-means, agglomerative, fuzzy-c-means, and even artificial neural networks. Spyrakos et al. (2018) moved further and presented an integrated classification schema with 13 inland and coastal waters types while providing a comprehensive analysis of the physical basis for each type (cluster) in terms of Inherent Optical Properties (IOPs) variability.

Overall, these studies have demonstrated how these classification schemas can contribute to an improvement of retrievals algorithms accuracy. In this regard, Mélin et al. (2011) proposed a band ratio empirical Chl-*a* blending algorithm for 2 distinct water classes. Vantrepotte et al. (2012) demonstrated the potential for optical 4 class-based approach to increase the performance of SPM

concentration retrieval in coastal waters. Moore et al. (2014) showed that when a blending approach is used together with OWT classification for Chl-*a* retrieval, the root mean squared error (RMS) is considerably reduced. More recently, Neil et al. (2019) calibrated empirical Chl-*a* algorithms for 13 inland OWTs and achieved a dynamic ensemble algorithm with a correlation coefficient of 0.89 and a mean absolute error of 0.18mg/m³.

Although these papers can demonstrate an improvement in parameters retrieval if OWTs are considered, inversion accuracy is not the final objective of the classification process; it is a secondary goal. The optimal number of clusters, and their homogeneity are normally determined by cluster validity measurements such as the Silhouette index (Vantrepotte et al., 2012), gap statistics (Spyrakos et al., 2018; Xue et al., 2019), or a combination of validity functions as compactness and separation (Moore et al., 2009).

By focusing on these validity metrics, the final clustering schema is a good descriptor of different water types, but it is not necessarily optimized for retrieval algorithm performances and robustness. Moreover, to the best of our knowledge, little attention has been paid to proposing a framework for global SPM retrieval over inland water.

For this reason, this study aims to assess different classification methodologies in terms of algorithms, input features, and spectra normalization, analyze their impact on the final SPM inversion accuracy and, ultimately, provide a robust framework intended for global SPM retrieval from remote sensed images. In particular, we aspire to determine which combination of clustering algorithms (e.g., K-means, Agglomerative, or Fuzzy C-means) and input features (i.e., normal versus normalized reflectance) provide the best optical classes for SPM retrieval. In particular, we would like to assess if models adjusted for specific OWTs may outperform models tailored for each catchment.

1.3 Materials and methods

Radiometric and Water Quality Field Dataset

For this study, a large dataset of above reflectance measurements obtained from field campaigns since 2008 has been assembled. Figure II-1 displays the location of the 60 rivers and reservoirs in tropical regions located in Brazil, Peru, French Guyana, Burkina Faso, and Laos that were sampled using a homogeneous protocol for both radiometric measurements and water sampling are included (Figure II-1).

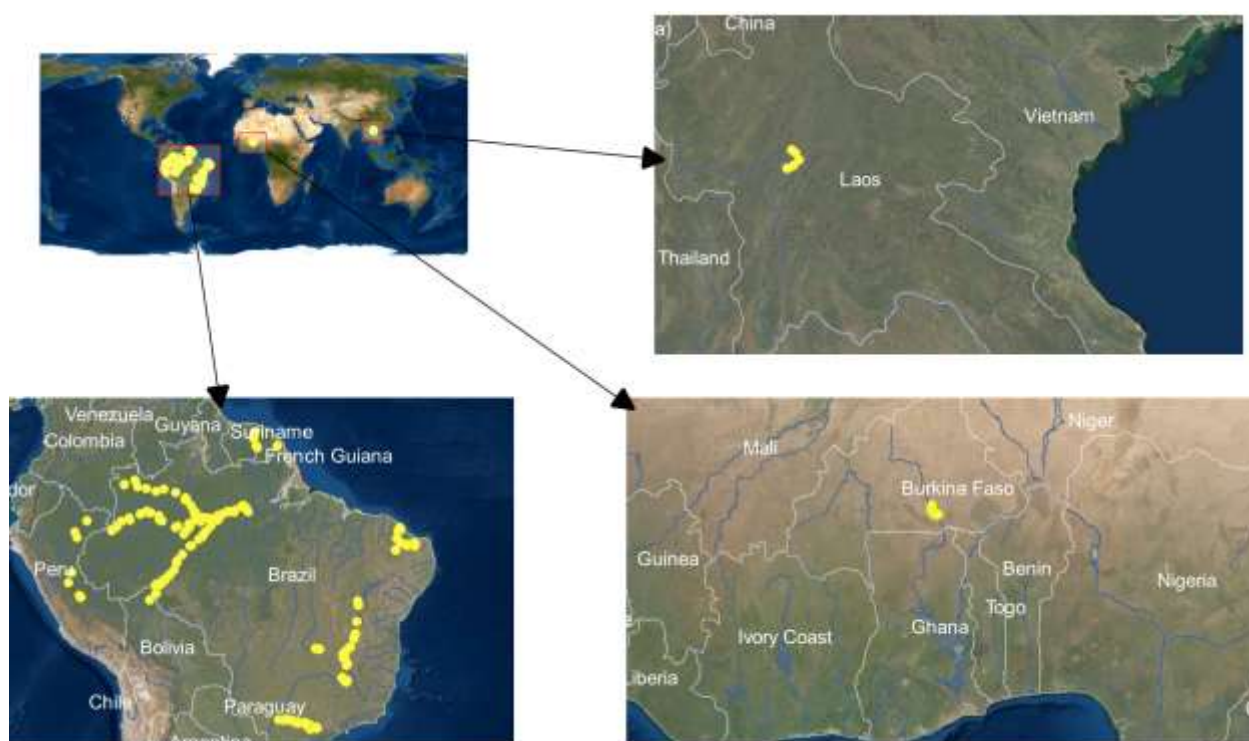


Figure II-1: Map of the locations that were sampled for both radiometric above water measurements and surface water collection.

This dataset comprises 1067 in-situ above-water hyperspectral radiometric measurements paired with SPM laboratory analysis. In addition to the radiometric measurements and the SPM concentration, Particle Size Distribution - PSD (182 values) is also available for some sites, as detailed in Table II-1.

Table II-1: Summary of data stored in the database and its water quality parameters.

Area	Radiometry	SPM	Particle Size Distribution
Madeira	185	185	66
Negro	177	166	1
Solimões and Amazonas	121	121	29
Paranoá	119	119	0
São Francisco	99	98	0
Maroni	82	82	0
Ucayali	62	59	5
Jaguaribe and Piranhas-Açu	58	55	0
Amazon Floodplains	46	46	0
Parapanema	41	41	0
Purus	34	34	17
Volta	29	21	0
Mekong	14	14	0
Total	1067	1041	182

The measurements were grouped in 13 more extensive areas, according to their watershed or sub-watersheds. Figure II-2 displays the SPM concentration range for each catchment, sorted by mean SPM concentration, denoting a very large overall range from 0.1 mg/l to more than 1,800 mg/l.

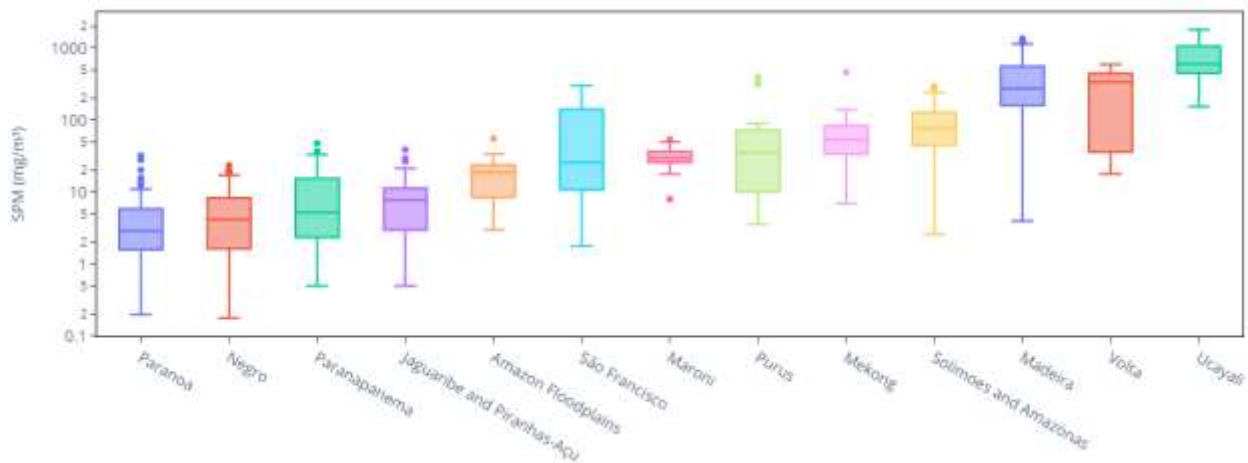


Figure II-2: SPM concentration range box plots displayed as a function of the watershed where radiometric and water samples were collected.

During the sampling field campaigns, above-water radiometric measurements were performed using TriOs RAMSES radiometers operating in the 350-950nm spectral range. A total of 3 hyperspectral radiometers were mounted on a boat to measure: a) the downwelling irradiance (E_d); b) the total upwelling radiance (L_u); and the downwelling sky radiance (L_d). The schematic representation is presented in Figure II-3. The radiometers setup follows the schema proposed by Mobley (1999), with a viewing direction θ of 40 degrees from the nadir and azimuthal angle ϕ of 135 degrees from the Sun, in order to minimize the effect of reflected light from the Sun (i.e., sunglint).

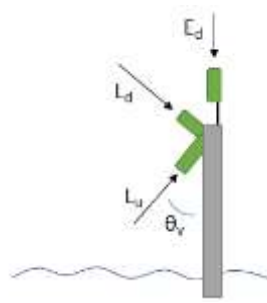


Figure II-3: Schematic representation of the above-water reflectance measurements that was used for all the sampling points analyzed in this study.

The estimation of the remote-sensing (R_{rs}) reflectance from above-surface measurements was performed according to the methodology proposed by Mobley (1999), defined as:

$$R_{rs}(\theta, \phi, \lambda) = \frac{L_w(\theta, \phi, \lambda)}{E_d(\lambda)} \quad \text{Eq. II-1}$$

where L_w is the water-leaving radiance, θ and ϕ are the polar (from nadir) and azimuthal (from the sun) directions, respectively, and λ the spectral wavelength.

This methodology introduces the factor ρ that represents the conversion factor between the measured downwelling sky radiance and the water surface radiance, i.e., the signal component originating from the skylight reflection of the rough air-water interface. This correction is necessary

because the L_u radiance contains sky radiance (L_d) reflected by the water surface (Eq. II-2). In this scenario, ρ is a function of θ, ϕ and λ , as well as of the water roughness modeled as a function of wind speed.

$$L_w(\theta, \phi, \lambda) = L_u(\theta, \phi) - \rho L_d(\theta, \phi) \quad \text{Eq. II-2}$$

For each measurement, the factor ρ has been adjusted according to the method presented in Harmel et al. (2012) based on full vector radiative transfer computations using the OSOAA model (Chami et al., 2015). This approach properly handles the directional and spectral variations of the water-surface radiance contribution considering all the related variables, such as the polarization features, aerosol impacts, and air-water interface roughness. For this processing, the open-source Python package (<https://github.com/Tristanovsk/trios>) has been used. Radiometry curves and mean reflectances for each group are shown in Figure II-4. Data collected from diverse campaigns were gathered into a unified format and then visually corrected and filtered for the occurrence of outliers and other artifacts. The original spectrum for each measurement was interpolated in 1 nanometer resolution and compiled into a database including, among other variables, the Greenwich Mean Time (GMT), geographic coordinates, and altitude. A comprehensive framework for managing and manipulating these data, called RadiometryTrios was developed in the context of the current thesis. The resulting package is available to the research community on GitHub through the following link: <https://github.com/cordmaur/RadiometryTrios>.

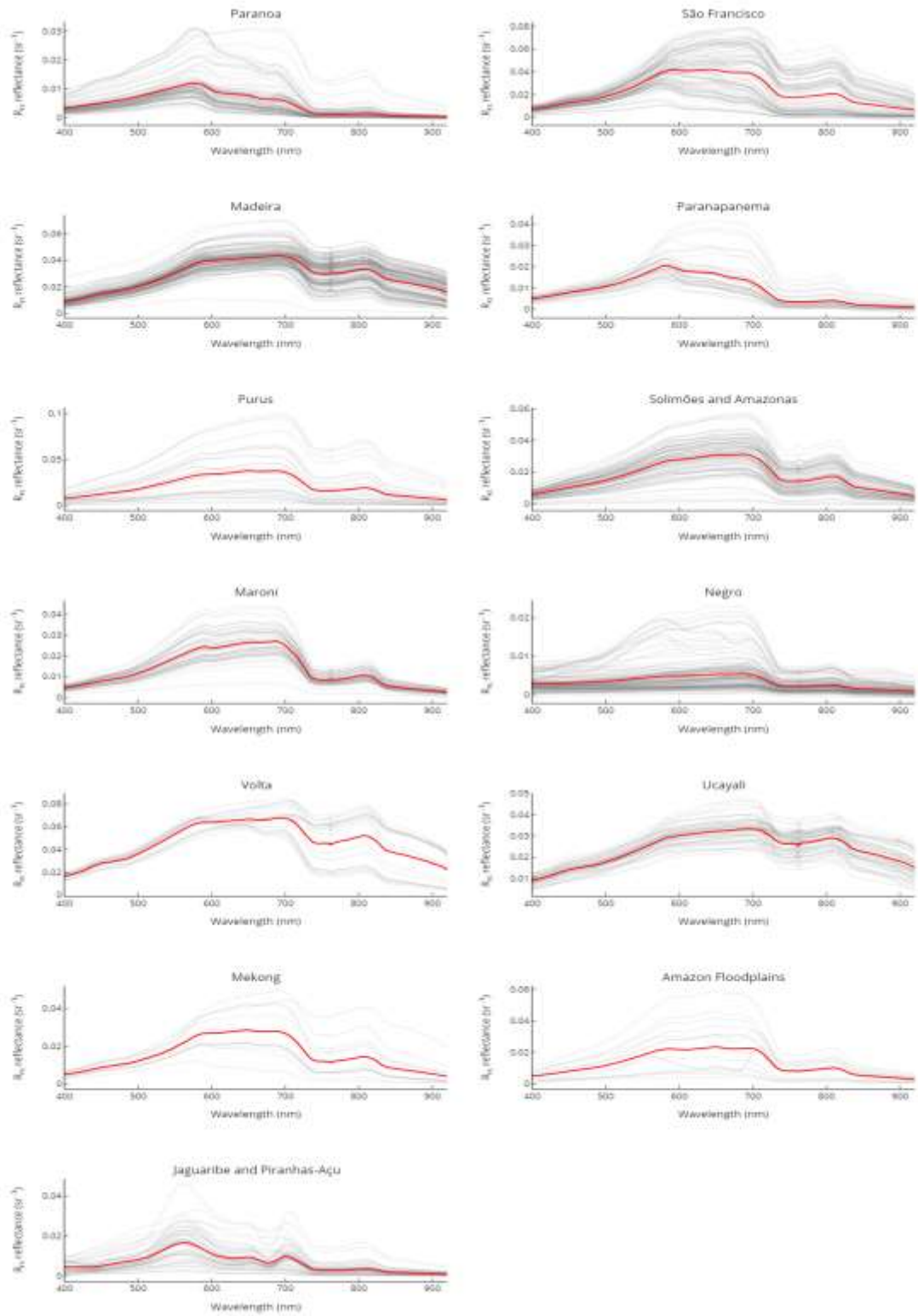


Figure II-4: Spectra curves and mean reflectance for each Area within the dataset. Red lines represent the mean reflectance in sr^{-1} .

Data Normalization

Besides the clustering algorithm itself, it is known that normalization or non-normalization of input features can lead to different clustering results, as already reported by previous studies (Jackson et al., 2017; Vantrepotte et al., 2012). While the non-normalized spectra classification is dominated by the gradient in the concentrations of SPM, the normalized spectra are better for focusing on shape variations.

The normalized field-measured remote sensing reflectance ($NR_{RS}(\lambda)$) was obtained by dividing the reflectance at each wavelength by the integral of the spectral curve (area) between 400nm and 920nm, as follows (Eq. II-3):

$$NR_{RS}(\lambda) = \frac{R_{RS}(\lambda)}{\int_{400}^{920} R_{RS}(\lambda) d\lambda} \quad \text{Eq. II-3}$$

After the normalization process, all spectra will have an area equals 1, and the spectral shape will be privileged in detriment to overall magnitudes. The effect of the normalization process can be visualized in the following example

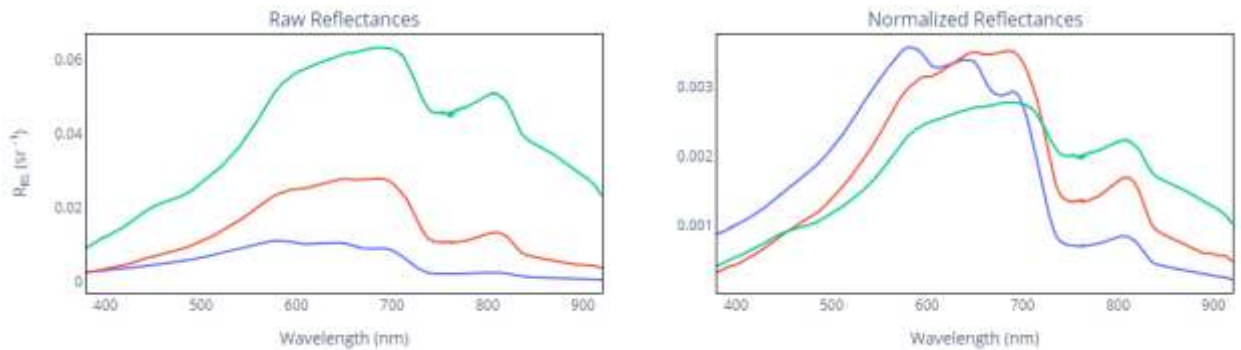


Figure II-5 of 3 spectra extracted from the Madeira River corresponding to 3 different SPM concentration levels of 4mg/l (blue), 75.8mg/l (red) and 286mg/l (green).

One issue that may arise from the normalization process occurs when the original reflectances are too low and then the normalized spectra become “noisy”. This may happen for spectra acquired over waters presenting extremely low albedos for all the spectrum, such as black water typical of tropical dissolved organic carbon-rich waters. To account for this, a second data cleaning process has been performed on the normalized spectra to identify possible outliers: a Principal Component Analysis (PCA) on the normalized spectra has been performed to reduce the data dimensionality to 2. The outliers were then identified visually and removed from the dataset; otherwise, they could disturb the clustering algorithm.

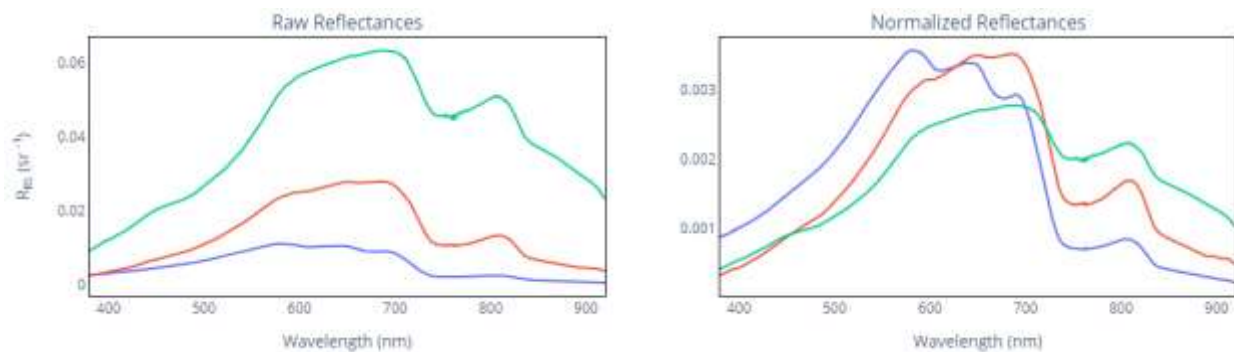


Figure II-5: Example of spectra normalization in 3 reflectance spectra from the Madeira river. Original spectra on the left and normalized spectra on the right. SPM concentration were of 4mg/l (blue), 75.8mg/l (red) and 286mg/l (green).

Water type classification algorithms

Different unsupervised classifications have been used to obtain the classes of OWTs, considering clustering algorithms with input resources (input information provided to the algorithm) and cluster validity measures (notably gap statistics). The K-means method remains the most used clustering algorithm to partition waters into different groups (Spyrakos et al., 2018; Xue et al., 2019), despite its inefficiency when the clusters present different sizes, as discussed in Cordeiro et al. (2021). Other algorithms have also been tested, such as Fuzzy C-means (FCM) (Bi et al., 2019; Jackson et al., 2017; Moore et al., 2014), Agglomerative (Vantrepotte et al., 2012), and Spherical K-means (Jia et al., 2021) . For this work, we analyzed K-means, FCM, and

Agglomerative algorithms. For the primary features used as inputs to the OWTs classification, we considered the R_{RS} data and normalized reflectance (see next section) obtained by dividing the reflectance at each wavelength by the integral of the spectral curve (area). Spectral resolution resampling has also been tested to assess its impact on the final clustering. The following wavelength steps were tested: 1nm, 10nm, 20nm, and 30nm. Additionally, Sentinel-2 and Sentinel 3 bands were also tested for clustering purposes. All tests have been done from the number of clusters (K), ranging from 2 to 15, to find the optimal K value regarding the overall SPM retrieval error.

SPM concentration models

To test the performance of each final classification with regard to SPM retrieval accuracy, we calibrated three simple mathematical models (linear, power and exponential) and an analytical single-band model developed for multisensor SPM retrieval (Nechad et al., 2010). All four models were calibrated for different spectral reflectance $R_{RS}(\lambda)$, normalized reflectance $NR_{RS}(\lambda)$ and reflectance ratios as $\frac{R_{RS}(\lambda_1)}{R_{RS}(\lambda_2)}$. These input features are denoted in the models as independent variable x , being a, b and c the parameters to be calibrated:

- Linear: $SPM = ax + b$
- Power: $SPM = ax^b + c$
- Exponential: $SPM = a^x + b$
- Nechad: $SPM = \frac{ax}{1-\frac{x}{b}} + c$

Considering the final goal to obtain a robust model that could be used to estimate SPM concentration from high-resolution remotely sensed images, the wavelengths (λ) used for the R_{RS} to be tested in the calibration phase, were the central wavelengths for the 10m and 20m spatial

resolution Sentinel-2 MSI spectral bands. When the central wavelength differs for Sentinel-2A and Sentinel-2B sensors, the average is considered.

For the reflectance ratios, all the possible combinations using the Sentinel-2 MSI bands have been tested during the fitting process. As we considered only 10m and 20m bands (10 bands in total), 45 different ratios (combination $\binom{10}{2} = \frac{10!}{2!(10-2)!}$) have been tested in each mathematical model, besides the R_{RS} and normalized bands.

The final wavelength values considered were rounded in nanometers, as described in Table II-2. As our main objective was to provide a robust SPM framework to be applied in different regions at a global scale and test its generalization performance for new, unseen conditions, the main dataset was randomly split into calibration (80%) and validation (20%) sets.

Table II-2: Sentinel-2 spectral bands used in models' calibration.

Sentinel-2 band	Rounded Central Wavelength (nm)	Bandwidth (nm)	Spatial Resolution
Band 2 – Blue	492	66	10
Band 3 – Green	560	36	10
Band 4 – Red	665	31	10
Band 5 – Red edge 1	704	15	20
Band 6 – Red edge 2	740	15	20
Band 7 – Red edge 3	781	20	20
Band 8 – Near-infrared (NIR)	842	106	10
Band 8A – Narrow NIR	865	21	20
Band 11 – Short wave infrared	1612	92	20
Band 12 – Short wave infrared	2194	180	20

An important factor that must be considered when using satellite reflectance to invert water quality parameters is the spectral resolution at each band. From Table II-2 **Error! Reference source not found.** one can note that the Sentinel-2 MSI sensor has bandwidths ranging from 15 to 180

nm, and this can interfere in the reflectance “as seen” by the satellite. To reproduce this behavior simulated Sentinel-2 bands have been derived from the field reflectances by applying the corresponding satellite’s spectral response function (SRF). The spectral response functions are provided by the European Space Agency (ESA) for each platform (Sentinel-2A and Sentinel-2B). As the sensors are almost identical, the SRF from Sentinel-2A has been used.

The SRFs for the bands listed in Table II-2 are presented in Figure II-6.

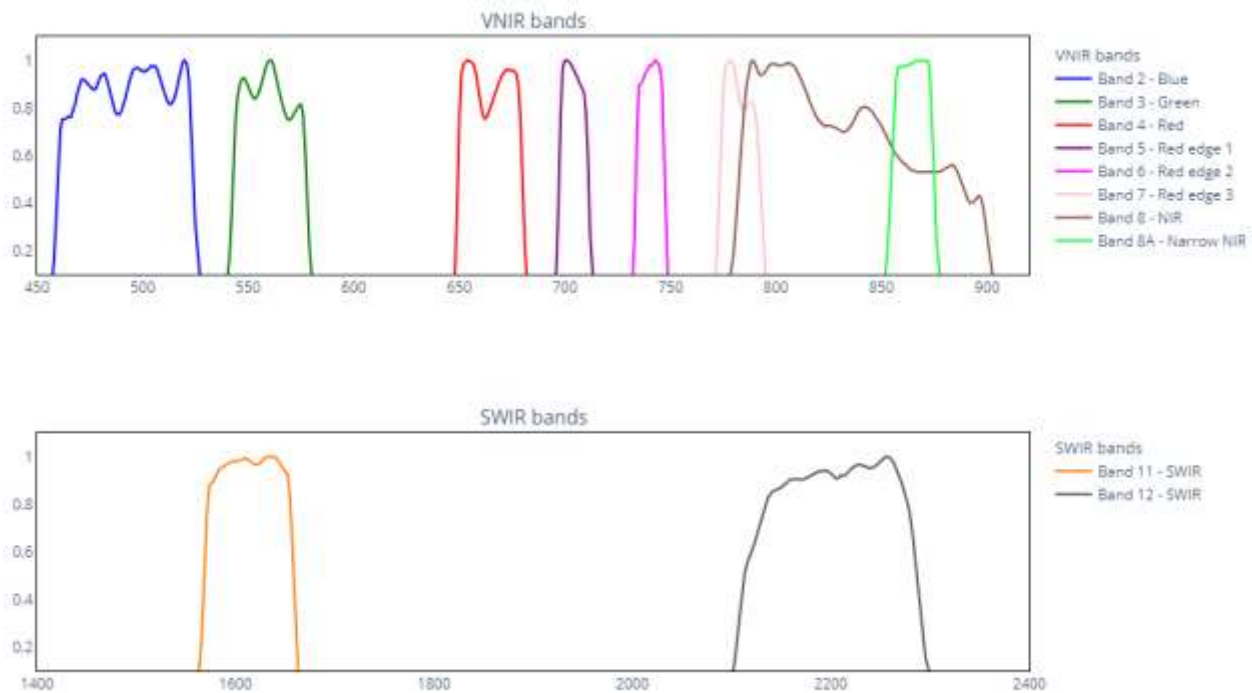


Figure II-6: Spectral response functions for Sentinel-2 Visual-NIR and SWIR bands.

Metrics and Fitting

As depicted from Figure II-2, the SPM values we are predicting can range from less than 1 mg/l to almost 2,000 mg/l and this exponential growth will lead to a non-linear fit. In this scenario, the R^2 (R-squared) measurement can be misleading, as it assumes a linear fit and will privilege some few high-concentration measurements in detriment to lower concentrations. To take account for this, the Root Mean Squared Log Error (RMSLE) is also calculated, according to Eq. II-4.

$$RMSLE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(\hat{y}_i + 1) - \log(y_i + 1))^2} \quad Eq. II-4$$

where y_i is the actual sediment concentration measurement and \hat{y}_i is the value predicted by the model.

To automate the process of model fitting to the calibration set, considering all input features and models and to test the goodness of fit to the validation set, a comprehensive fitting framework has been developed in Python. The SciPy package (Virtanen et al., 2020) has been used in a two-step basis. Initially, the curve fit method is used. It provides a non-linear least square fit of an arbitrary function f to the given data. As the metric being used is RMSLE, the least square optimization does not match the same result expected to minimize the RMSLE, but the parameters obtained can be used as initial guesses.

Afterwards, the minimize method was used, and the loss function was defined to be the RMSLE metric. This two-step fitting assures that the RMSLE metric will be correctly minimized during the calibration phase and the best fit is selected as a combination of input feature (x), mathematical model and corresponding parameters.

Besides the RMSLE metric, which is used to optimize the model, other metrics proposed by Moriasi et al. (2007) were also calculated to assess the goodness of fit performance of the final model and compare it to the baseline. The metrics are the coefficient of determination (R^2), the Root Mean Squared Error (RMSE), the Standard Deviation Ratio (RSR) and the Percent bias (PBIAS) as detailed in the following equations (Eq. II-5 to Eq. II-7):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad \text{Eq. II-5}$$

$$RSR = \frac{RMSE}{STDEV_{obs}} = \frac{\sqrt{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\sqrt{\sum_{i=1}^n (y_i - \bar{y}_i)^2}} \quad \text{Eq. II-6}$$

$$PBIAS = 100 * \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i)} \quad \text{Eq. II-7}$$

To test the coherence between the clustering algorithm and the membership function, as discussed in the Membership Function section, the accuracy score has been computed, according to Eq. II-8.

$$Accuracy\ Score = \frac{1}{n} \sum_{i=1}^n (y_i = \hat{y}_i) \quad \text{Eq. II-8}$$

Membership Function

Once the clusters are defined by using the training dataset, to assign any new sample to its correct cluster, it is necessary to define a distance function to compare the similarity degree of the new spectra to the mean spectra of each cluster. In this study, we've tested the Euclidean and the Mahalanobis distances.

In the simple Euclidean distance, the distance between two points in an n -dimensional space (\mathfrak{R}^n) is defined as the length of a line segment between these points. In our study, we can consider each reflectance spectrum as a point in a n -dimensional space, where n will be the number of wavelengths (or reflectance bands) considered. Therefore, we can define a spectrum as a vector $\vec{x} = (R_{RS}(B_1), R_{RS}(B_2), \dots, R_{RS}(B_n))$, where $B_1 \dots B_n$ are the reflectance bands. Similarly, each cluster can also be represented by a vector $\vec{\mu}$, with the mean values in each wavelength. Using

vectorial notation, the Euclidean distance can be written as the norm of the subtraction between the two vectors (Eq. II-9).

$$\text{Euclidean distance} = \|\vec{x} - \vec{\mu}\| \quad \text{Eq. II-9}$$

Besides the Euclidean distance, the Mahalanobis distance is also proposed in the literature to measure the distance between a point and a distribution (e.g., clusters) because it accounts for two issues: scaling and correlation among variables. The first issue occurs when the variables have distinct magnitudes. In this scenario, the Euclidean distance will privilege the variable with greater magnitude in detriment of the others. This issue can be easily solved by normalizing each variable before applying the Euclidean distance (i.e., Normalized Euclidean distance). The second issue occurs because the Euclidean distance weights each variable equally, so it assumes that the variables represent unrelated and equally important information, which is certainly not the case for adjacent wavelengths. To account for this, the Mahalanobis distance uses the covariance matrix of the distribution. Therefore, the Mahalanobis distance between a vector and a set of observations (cluster) is defined in Eq. II-10.

$$\text{Mahalanobis Distance} = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})} \quad \text{Eq. II-10}$$

where S is the covariance matrix of the cluster, $\vec{\mu}$ is its centroid vector and \vec{x} is the vector to be assessed.

Experimental Methodology

To select which classification schema is better to improve SPM retrieval accuracy, several experiments have been performed to test the significance of the following hyperparameters to the final results:

- Clustering algorithm: K-means, Agglomerative, and Fuzzy C-means;

- Spectra type: original (R_{RS}) and normalized (NR_{RS}) reflectances;
- Spectral resolution: wavelengths in 1, 10, 20, and 30 nm steps and Sentinel-3 bands;
- Number of clusters: K varying from 1 to 10; and
- Membership function: Euclidean or Mahalanobis distance.

The spectra database has been randomly split into training and validation datasets with 80% and 20% of the measurements, respectively. Each experiment follows five steps highlighted in Figure II-7: 1- clustering, 2- model calibration, 3- membership assignment, 4- SPM inversion, and 5- Evaluation of combined metrics.

The process starts with the clustering performed on the training dataset according to its parameters (algorithm, spectra type and resolution). Then, for each cluster, a model is calibrated (i.e., the algorithm searches for the best mathematical model and input bands). To test the performance, the measurements of the validation set are assigned to the respective clusters by the membership function. Besides the membership function to be used (Euclidean or Mahalanobis), it is also necessary to set the input bands that will be used during the assignment.

As the final objective is to create a schema to be used by the Sentinel-2 imagery, the Sentinel-2 MSI bands (Table II-2) were selected. After the cluster assignment, the SPM inversion is performed for all the points in the validation dataset by using the corresponding models and the combined metrics are evaluated. This process is repeated for K values varying from 1 to 10.

1.4 Results

Overall optical water classes

To assess the spectral coverage of the dataset for different inland water types, we used as a reference the optical water types (OWT) described by Spyarakos et al. (2019) for inland water bodies. In Spyarakos' framework, 13 different OWTs were proposed and the below-water remote

sensing reflectance ($R_{RS}(0 -)$) means were provided in wavelengths, ranging from 400 to 800nm. The OWTs were identified for each record on our dataset through Euclidean distance. After membership assignment, we didn't find any measurement from our dataset to pertain to water type 13. This was expected, as type 13 comprises clear "blue-water" that is usually present in coastal and oceanic systems and it is not scope of our study. Figure II-8 shows the reflectance data sorted into the specific OWTs.

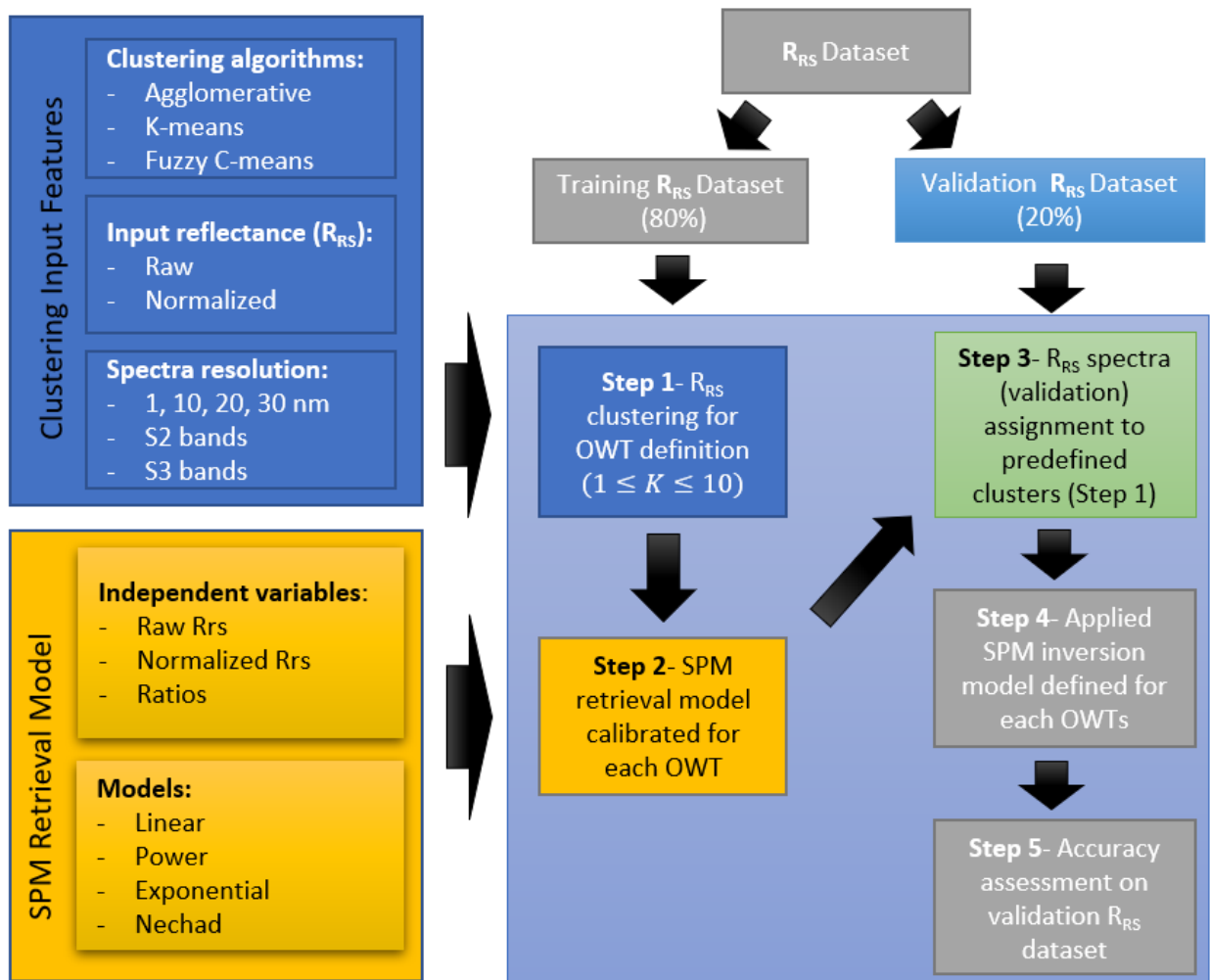


Figure II-7: Schematic view of the steps involved in the OWT definition assessment for different scenarios considering different input features (blue box) and model inversion types (yellow box).

Additionally, OWTs 1 and 7 are also absent from our dataset. According to Spyrakos et al. (2018), they represent hypereutrophic waters with scum cyanobacterial bloom and vegetation-like R_{RS} (i.e., OWT=1) and highly productive waters with high cyanobacteria presence (i.e., OWT=7). These hypereutrophic waters are not considered important targets for our study as we focus mainly on sediment-laden fluvial waters for SPM retrieval.

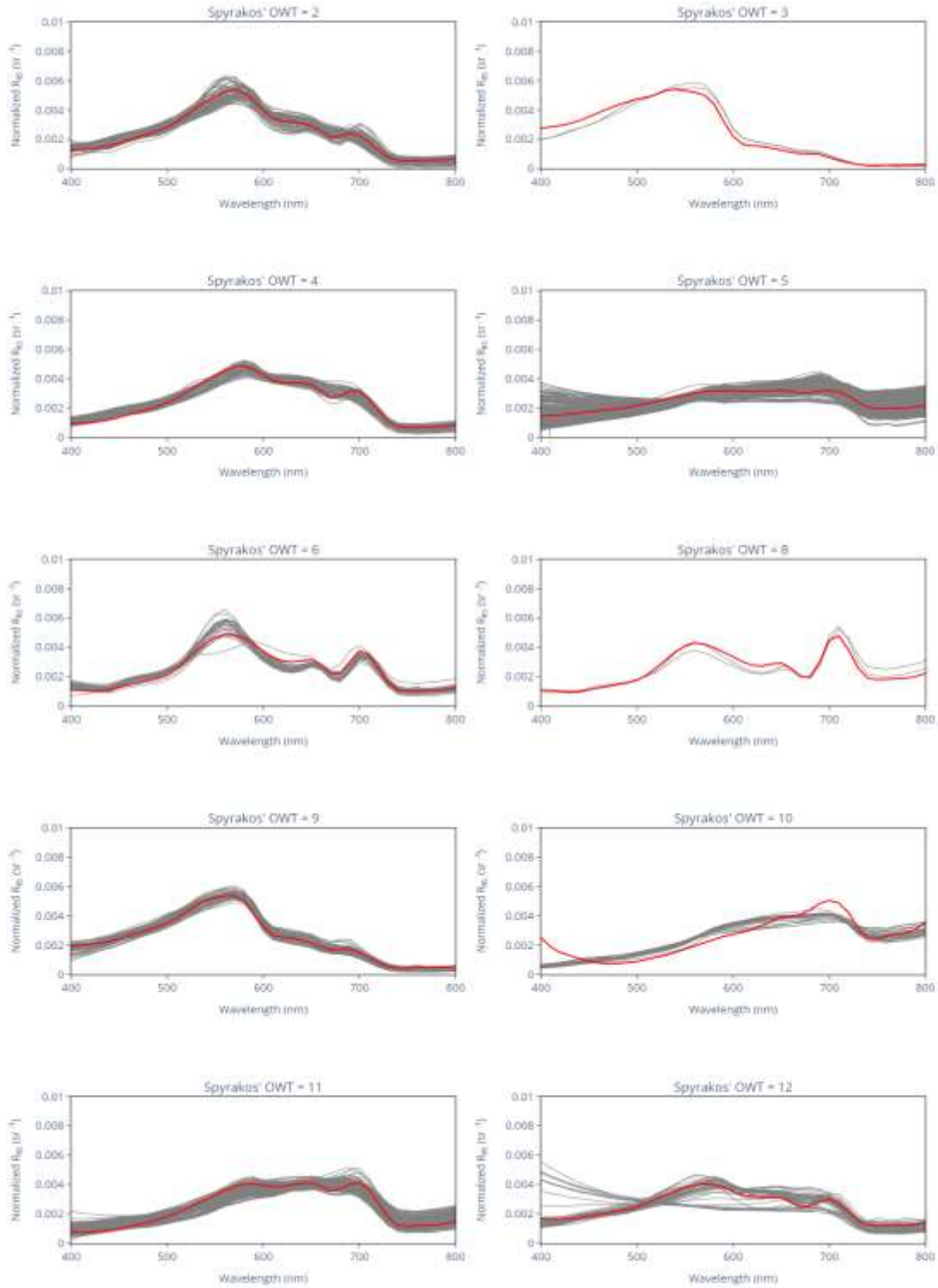


Figure II-8: R_{rs} spectra used in this study classified by Optical Water Type (OWT) as proposed by Spyraeos et al. (2019). Red lines represent cluster means for each OWT.

Baseline retrieval models

To provide a comparison reference for the results obtained from the different water classification methods, three baselines have been produced. First, a simple hybrid SPM model with two algorithms calibrated specifically for high and low SPM concentrations range. A second baseline model was calculated with the calibration of one algorithm for each of the 13 catchments sampled. Additionally, one baseline has been calibrated considering one model for each Spyrakos' OWT. Each calibration was performed using the models and metrics described in sections: SPM concentration models and Metrics and Fitting.

Hybrid Algorithm (High/Low SPM concentration class)

The complex interactions between optically active components and its relationship with SPM concentration depend on the SPM concentration range. Previous studies have concluded that short wavelengths are better for retrieving lower SPM concentrations whereas near-infrared wavelengths are better for higher concentrations (Doxaran et al., 2002; Long and Pavelsky, 2013). In this context, hybrid algorithms, with more than one mathematical model calibrated for different concentration ranges, can be a viable alternative to serve as a baseline. This approach has been used successfully in water quality studies for retrieving specific ranges of chl-*a* (Matsushita et al., 2015; Pahlevan et al., 2020; Smith et al., 2018).

In the present study, we opted to calibrate two mathematical models, for lower and higher SPM concentrations, considering the wavelengths available in the Sentinel-2 MSI sensor.

To develop an operational retrieval model that works for an entire SPM domain using multiple algorithms (i.e., for high and low SPM concentrations), it is crucial to define how the radiometric samples will be categorized in the High and Low SPM groups. As the SPM concentration is our dependent variable and it is not known "*a priori*", it is necessary to use a

screening method on the reflectance to select the best algorithm to be used for each pixel in the scene.

To identify which Sentinel-2 band and which threshold value could be used to split the samples in the two desired groups (high and low SPM), an algorithm has been developed to test various combinations. The algorithm proceeds as follows: 1- For each tested band, the percentile reflectance (from 10th to 90th percentile) values were calculated; 2- for each percentile value (threshold), the dataset is split into two groups, where the samples with band value below the threshold belongs to the low SPM group and samples with band value above the threshold are assigned to the high SPM group; 3- each dataset group (low and high) is then fitted to find the best combination of mathematical model and input reflectances (R_{RS} vs NR_{RS}), as described in section Metrics and Fitting. The advantage of splitting the dataset in the percentiles is that the results obtained for each band can be compared easily regardless of the actual absolute reflectance values used for the split and for each band. Figure II-9 presents the RMSLE metric obtained by splitting the dataset in the corresponding band/percentile. For clarity, only the best-performing bands in the visible range (665nm), in the red edge (740nm) and in the near-infrared range (865nm) are displayed.

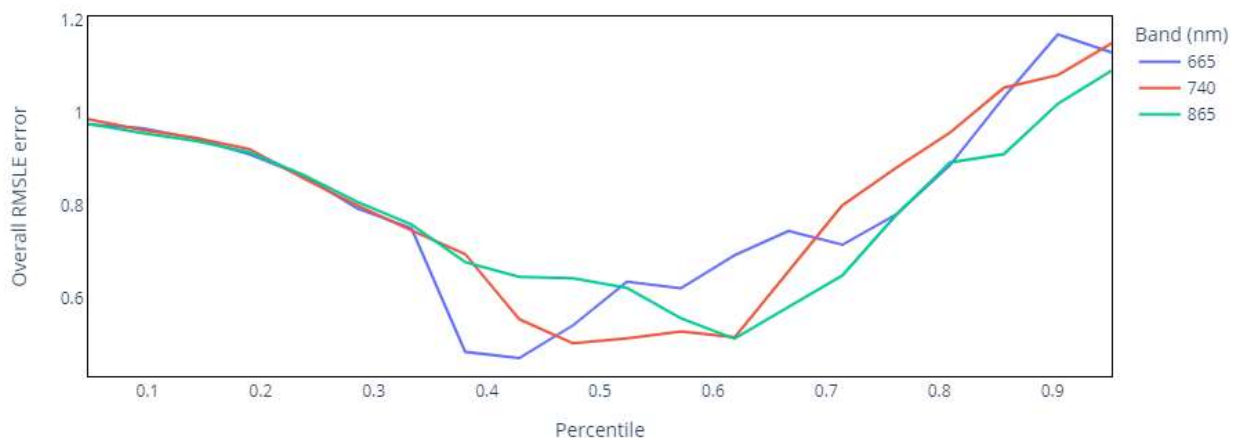


Figure II-9: Overall RMSLE errors for hybrid model calibrated for high and low concentrations, according to different “proxy” bands and threshold values for three different spectral bands: 665, 740 e 865 nm. Results obtained from the calibration dataset.

The best performances on the calibration dataset were achieved using bands 665nm and 740nm with RMSLE values of 0.47 mg/l and 0.50 mg/l, respectively. Adopting band 665nm as a proxy for determining high and low SPM concentrations, the threshold value for the optimal split was $R_{RS}(665nm) = 0.0182 \text{ sr}^{-1}$, which corresponds to the quantile 0.428 obtained from the graph.

Once defined the band and cutoff value to serve as a proxy to split the dataset, each sub-dataset was calibrated accordingly. The best fit for the lower concentration points (defined as $R_{RS}(665nm) < 0.0182$) was a linear relationship using $R_{RS}(705nm)$, with RMSLE=0.55 mg/l measured on the validation dataset. For the high concentration sample (defined as $R_{RS}(665nm) \geq 0.0182$), the best fit was a power relationship using the ratio $\frac{R_{RS}(783nm)}{R_{RS}(705nm)}$, which obtained RMSLE=0.33 mg/l in the validation dataset. The best models (mathematical model and spectral band used as independent variable) for each group of measurements (i.e., low and high SPM concentration), as well as its metrics, are presented in Table II-3.

Applying the best models to the entire validation dataset to calculate the combined metrics, we obtained $R^2 = 0.86$ and $\text{RMSLE} = 0.45 \text{ mg/l}$. The predicted versus actuals graph on the validation set is presented in Figure II-10c, as well as the mean reflectances (Figure II-10a) and concentration range (Figure II-10b). The actual versus prediction plot presents a good agreement for SPM values between 10 mg/l and 1,000 mg/l. Above 1,000 mg/l, there is an underestimation of SPM, and for the lowest SPM concentrations ($\text{SPM} < 2 \text{ mg/l}$), it happens the inverse, with an overestimation behavior.

Table II-3: Best fits for each group of measurements (low and high SPM concentrations). The combined metrics were calculated on the validation dataset.

Group	Model	Spectral Band	R ²	RMSE (mg/l)	RMSLE (mg/l)	RSR	PBIAS (%)
Low SPM concentration	Exponential	705	0.46	4.81	0.55	0.74	17.70
High SPM concentration	Linear	783/705	0.83	128.40	0.33	0.40	7.34
Overall metrics			0.86	123.78	0.45	0.37	12.68

Watershed-based model baseline

The second baseline is calculated by using one explicitly calibrated model for each watershed in the training dataset. The disadvantage of this approach is that a new model calibration should be required for every new region to be remotely monitored. Even so, the results obtained from this watershed-oriented model are essential to assess the maximum expected accuracy of the water classification obtained from this study. The algorithm and metric results calculated for each watershed are presented in Table II-4 and correspond to the fitting results on the training dataset. Two options were considered for validation considering the assignment method: i) using a simple assignment decision based on its geographic / watershed location; ii) using a membership function (Euclidean distance) ignoring the measurement location.

For this first option (simple geographic assignment) and applying these models to the validation dataset, we obtained $R^2 = 0.95$ and $RMSLE = 0.46$ mg/l, which is better than those obtained with the High/Low SPM hybrid model, especially for higher concentrations and considering the R^2 metric. The watershed-based classification performed better than the Low/High SPM classification also in terms of RMSE retrieval achieving 74.4 mg/l and 123.78 mg/l,

respectively. The actual versus prediction plot (

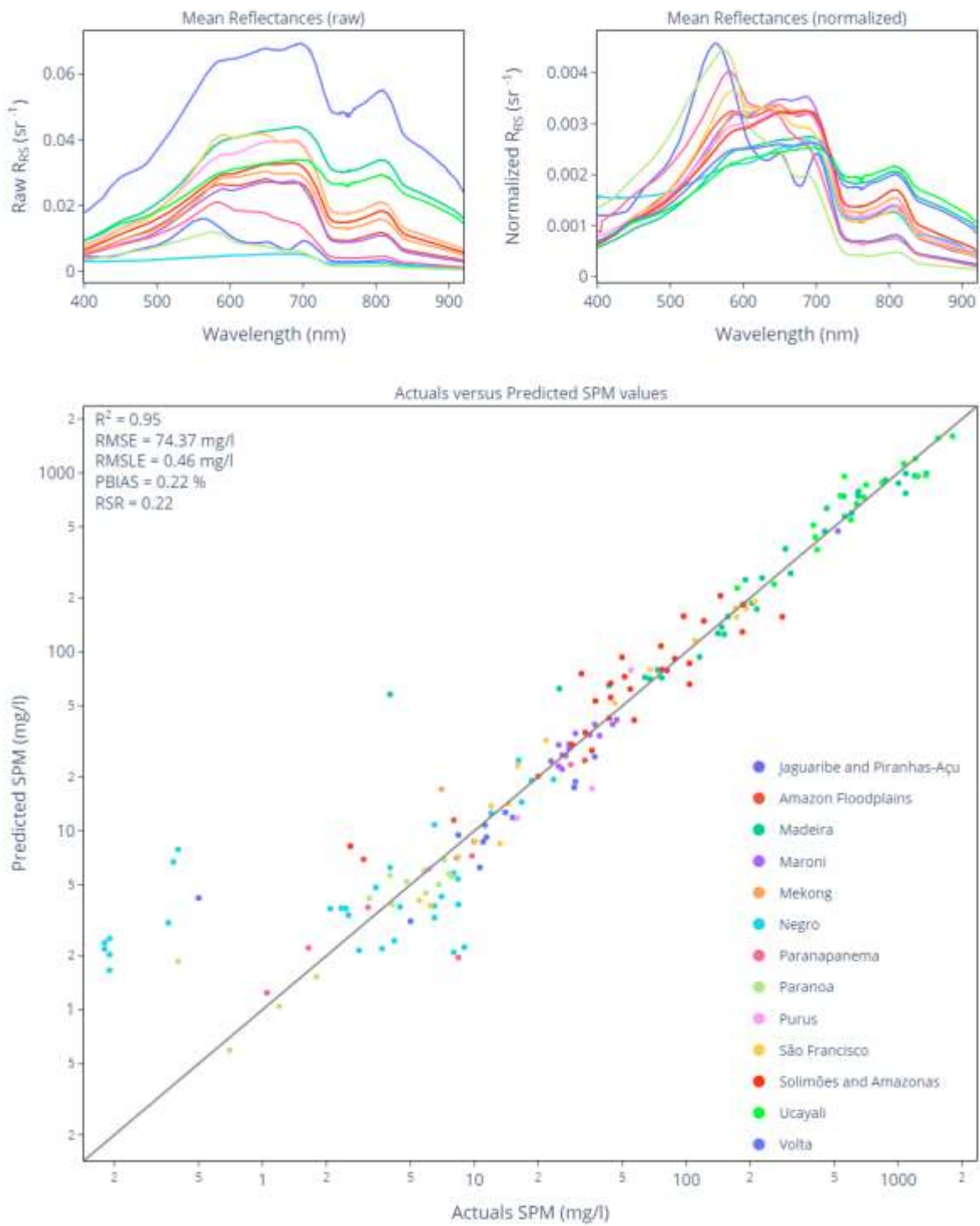


Figure II-11c) presented an excellent agreement for SPM values above 10 mg/l and a relative

dispersion between 1 and 10 mg/l with an overestimation (predictions higher than actuals) for the lowest SPM concentrations ($SPM < 1$ mg/l).

Although we obtained consistent overall metrics by calibrating one model for each basin, these results require an “*a priori*” knowledge about the area to be monitored, which is not desirable for an operational workflow designed to monitor large areas with distinct water types. Considering our final objective is to provide an algorithm robust enough to be applied globally, the second option has been conducted where we didn’t inform the algorithm to be used for the points in the validation dataset. The assignment of the algorithm was achieved using the Euclidean membership function, as described in item 1.3 - Materials and methods/Membership Function. In this new scenario, the obtained results were: $R^2 = 0.93$ and $RMSLE = 0.59$ mg/l. The RMSLE error was greater than the one obtained with the hybrid algorithm, but with a R^2 slightly better. This performance degradation is related to inaccurate automatic assignment based on the spectral shape. Most of this assignment confusion occurred in the lower reflectance range, while high reflectance groups were identified correctly.

With respect to the models calibrated for each watershed (Table II-4), it is clear that regardless the region to be monitored, the power function with input reflectance on the NIR range (between 700 and 800) or a reflectance ratio between NIR and visible (from 490nm to 665nm), seems to adjust better for most cases.

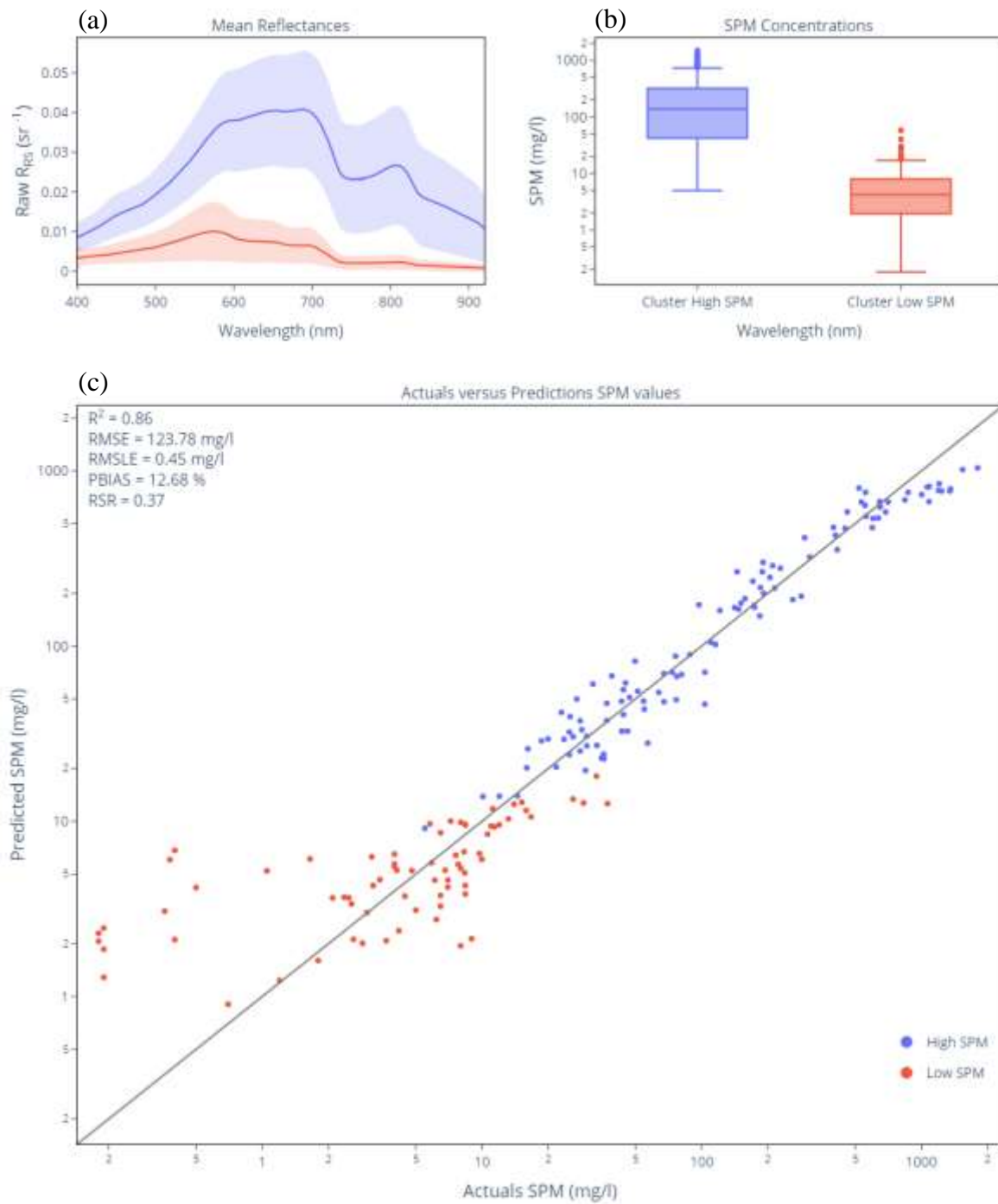


Figure II-10: Baseline retrieval model: 2-Class Optical Water Types (Low/High SPM range) (a) Mean reflectance spectra for each group; (b) SPM range for each group; (c) Actual SPM (X axis) vs Predicted SPM (Y axis) graph in mg/l. The colors for panels a, b, and c represent the corresponding low SPM (Red) or high SPM (Blue) concentration group.

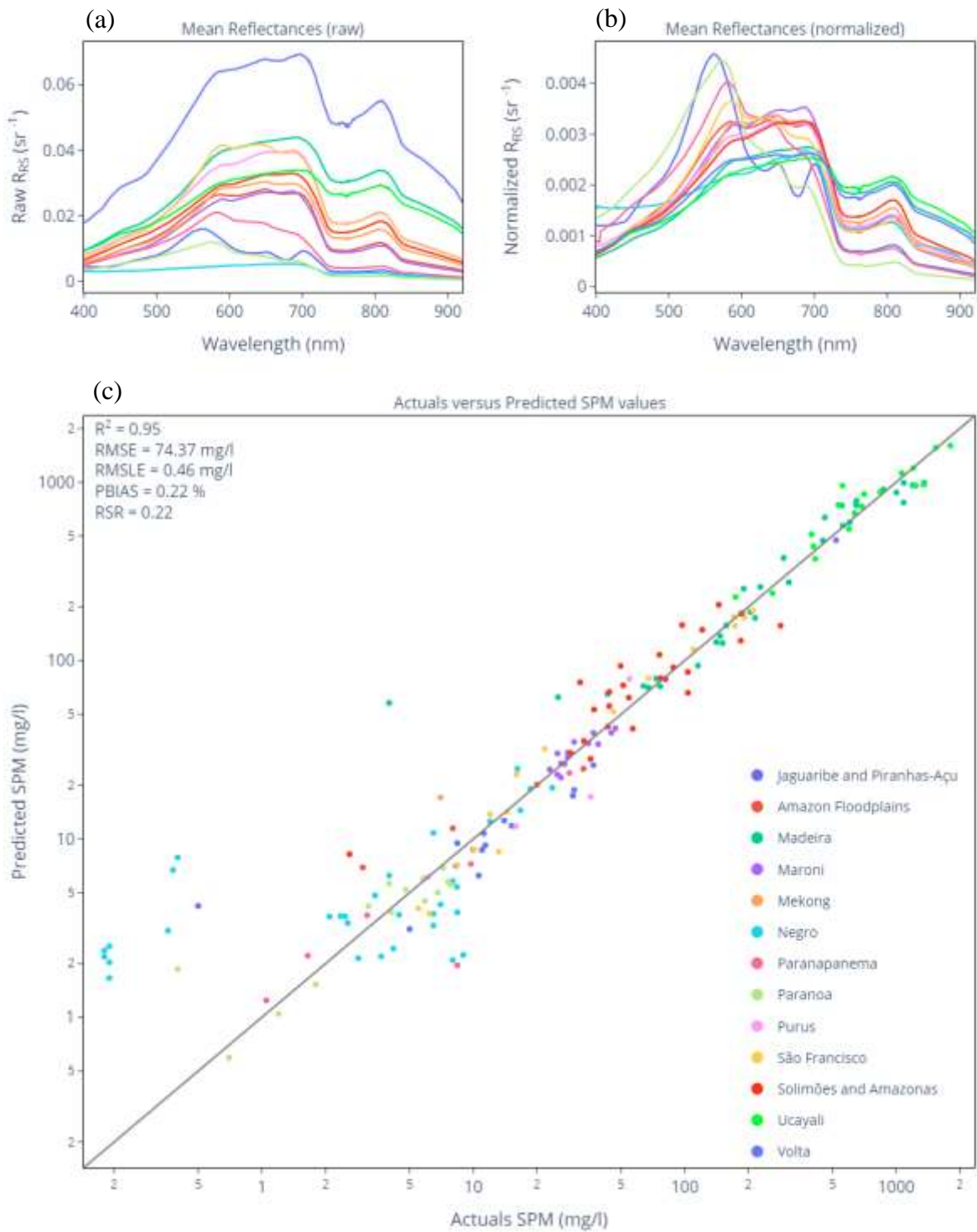


Figure II-11: Baseline retrieval model: Watershed-based OWTs (a) Mean raw reflectance for each considered watershed; (b) Mean normalized reflectance for each considered watershed; (c) Actual vs prediction SPM concentrations in mg/l considering one algorithm for each watershed. The colors represent different watersheds.

Table II-4: Models for the each watershed. The metrics are calculated considering the points in the validation dataset.

Watershed	Model	Spectral Band	R ²	RMSE (mg/l)	RMSLE (mg/l)	RSR	PBIAS (%)
Jaguaribe and Piranhas-Açu	Power	783	0.61	4.36	0.46	0.60	19.49
Amazon Floodplains	Linear	740	0.82	4.49	0.37	0.37	2.92
Madeira	Power	b740/b705	0.91	117.86	0.50	0.30	1.64
Maroni	Exponential	705	0.78	3.48	0.10	0.45	1.19
Mekong	Power	842/443	-0.56	10.09	0.82	1.17	-14.36
Negro	Power	705	0.64	3.52	0.74	0.59	-3.51
Paranapanema	Power	b705/b560	0.86	3.51	0.50	0.34	24.14
Paranoa	Power	783	0.78	1.19	0.26	0.46	6.00
Purus	Power	705	0.33	15.40	0.42	0.71	-1.27
São Francisco	Power	n783	0.99	8.94	0.21	0.11	1.59
Amazon / Solimões	Exponential	b740/b665	0.62	37.70	0.39	0.60	-4.76
Ucayali	Power	b740/b705	0.86	170.06	0.21	0.37	-1.20
Volta	Power	b783/b490	0.98	28.79	0.33	0.10	11.85
Overall metrics			0.95	74.37	0.46	0.22	0.22

Spyrakos' OWT-based model baseline

The last baseline was calculated using the OWTs proposed by Spyarakos et al. (2018) as a starting point in order to create retrieval models for each class. Considering that each class is defined by its mean normalized reflectances, the original reflectances in our dataset were normalized accordingly.

Classes 3 and 8 didn't have results due to the number of points being below the minimum required to fit a model properly. Therefore, a total of 8 models have been created, as presented in

Table II-5.

Table II-5: Models for the Spyrakos' OWT classes. The metrics are calculated considering the points in the validation dataset.

Spyrakos' OWT	Model	Spectral Band	R ²	RMSE (mg/l)	RMSLE (mg/l)	RSR	PBIAS (%)
2	Power	783	-0.02	2.80	0.62	0.96	20.31
4	Linear	842	-113.3	98.84	1.09	10.44	-177.8
5	Power	b740/b443	-0.20	260.55	1.05	1.08	40.66
6	Power	842	-0.23	2.49	0.24	1.04	15.86
9	Nechad	b783/b490	0.03	0.52	0.31	0.85	9.33
10	Nechad	b740/b705	0.26	408.31	0.44	0.84	37.03
11	Linear	783	0.55	17.98	0.39	0.67	3.00
12	Power	665	0.87	3.09	0.73	0.35	0.17
Overall metrics			0.64	200.77	0.71	0.60	33.89

Applying these models to the validation dataset, we obtained $R^2 = 0.64$, $RMSLE = 0.73$ mg/l, and $RMSE = 200.7$ mg/l, which was worse than those obtained with the previous baselines (i.e., High/Low SPM classes and watershed-based OWTs).

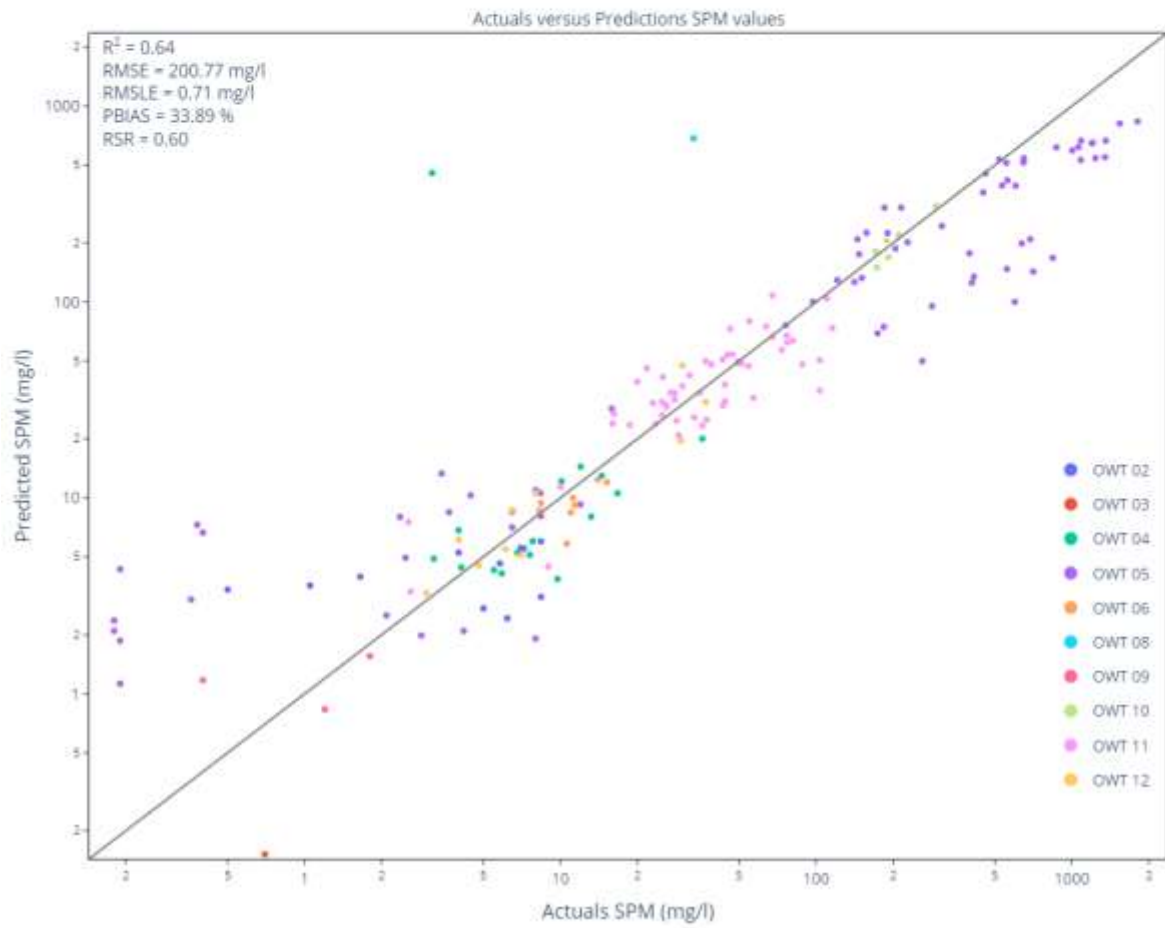
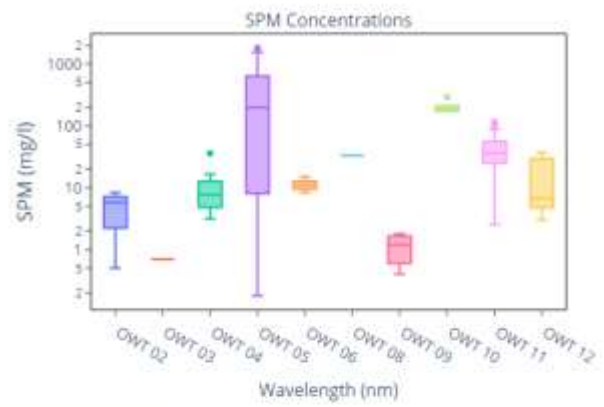
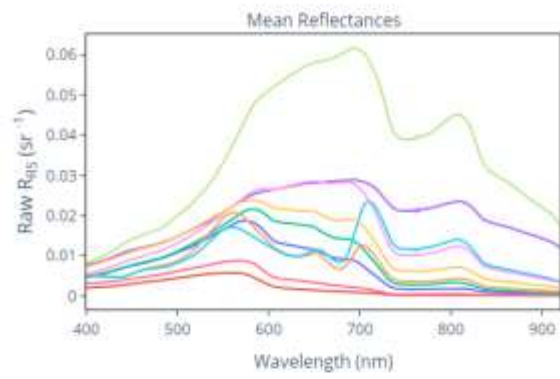


Figure II-12a presents the mean reflectance for each OWT, and

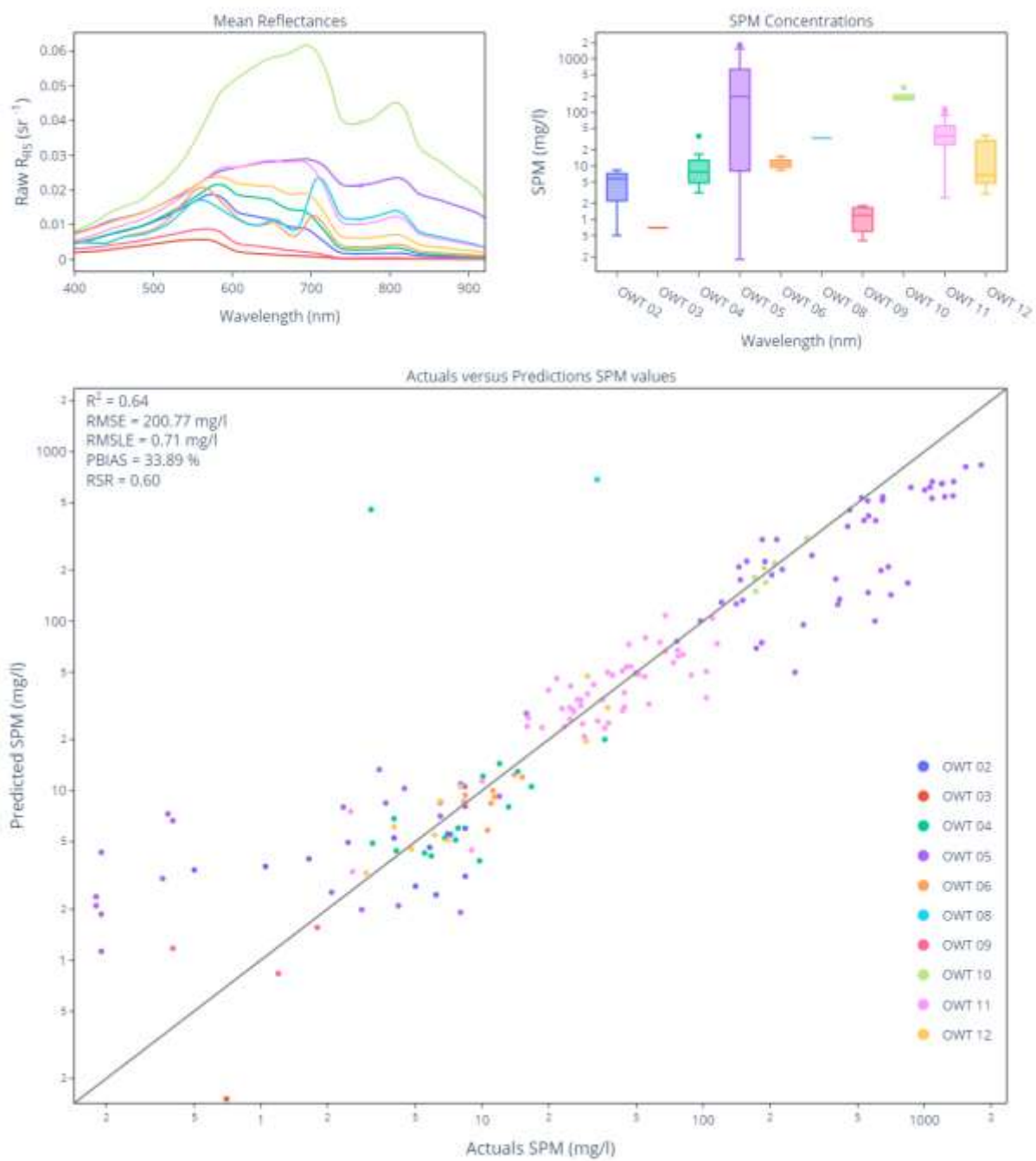


Figure II-12b the SPM concentration range corresponding to each OWT. Measured versus predicted SPM concentration (

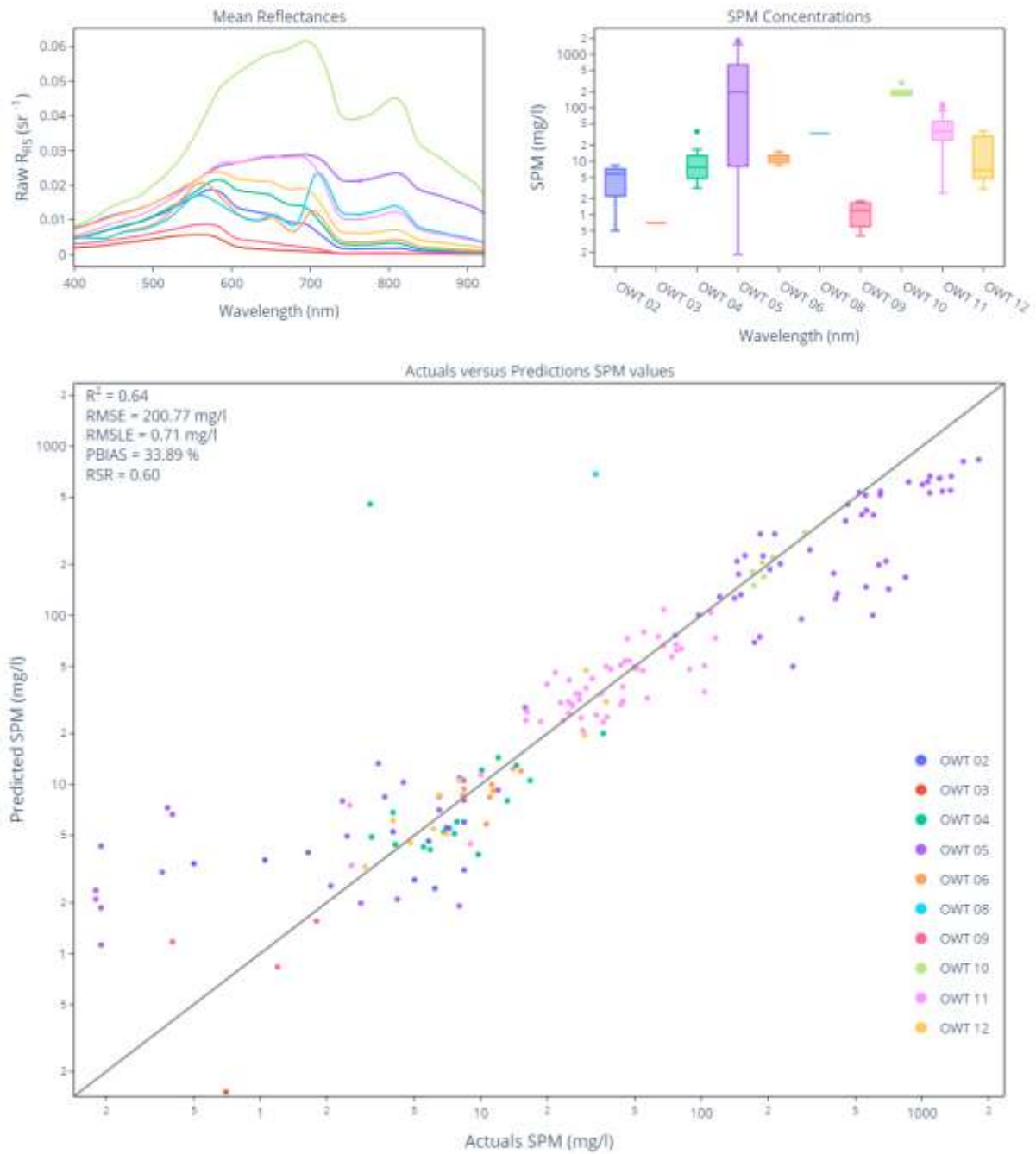


Figure II-12c) shows a higher dispersion across all range, when compared to the previous baselines. In particular, the retrieval based on Spyrakos' OWTs performed poorly for the largest SPM concentration, beyond 100 mg/l.

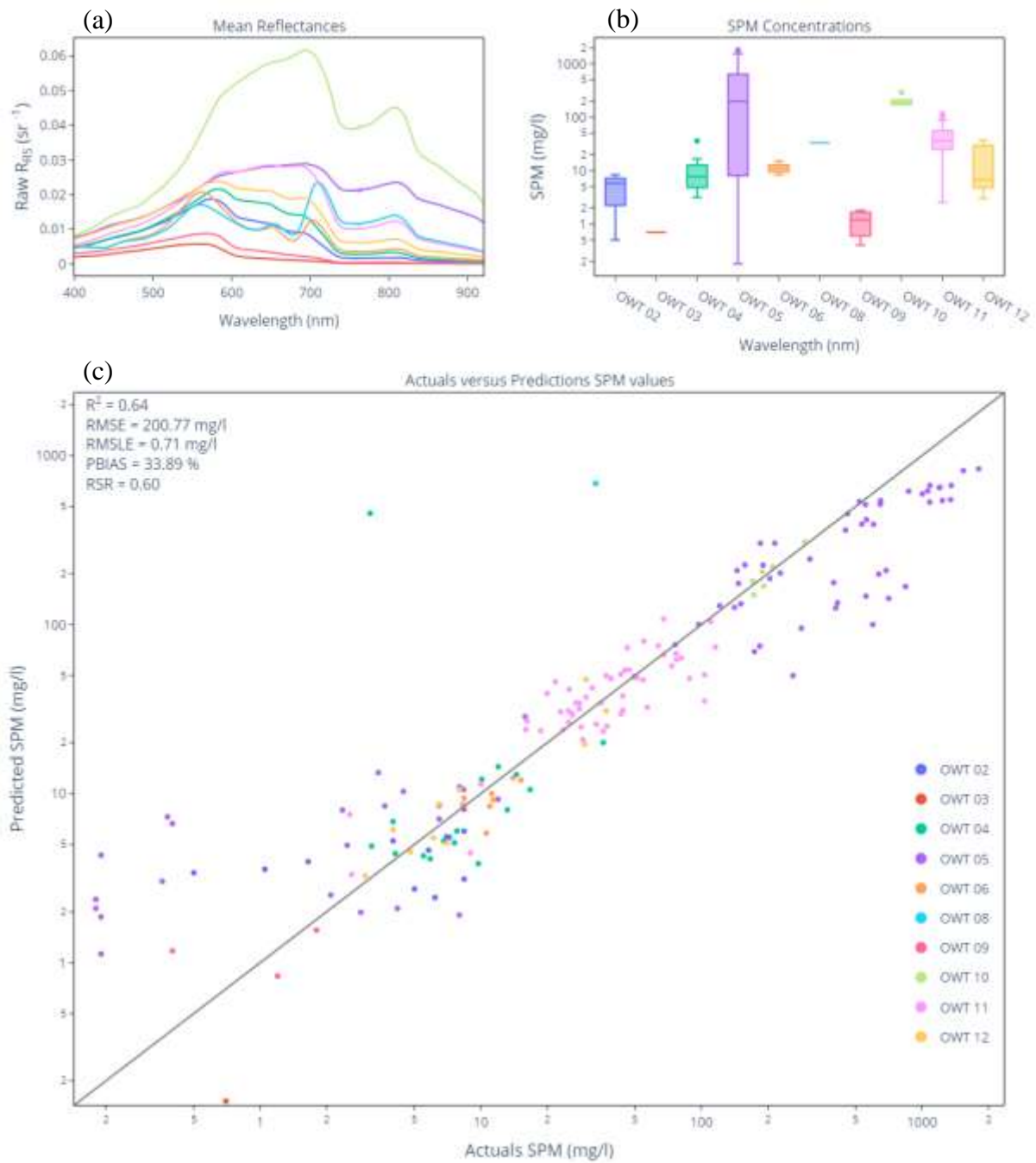


Figure II-12: Baseline retrieval model: Spyrakos' OWTs (a) Mean raw reflectance for each considered OWT; (b) SPM range for each OWT; (c) Actual vs prediction SPM concentrations in mg/l considering one algorithm for each OWT. The colors represent different classes.

Membership function assessment

One key aspect of hybrid algorithms is to properly select the model to be used for each pixel automatically. To achieve this, the membership function is responsible for matching a reflectance sample to the pre-defined spectral groups (clusters). Diverse factors impact the accuracy of the membership matching. One factor concerns the dimensions (or input features) used to create the clusters and to perform the assignment of new samples. For example, if the pre-defined optical classes were created using hyperspectral data, and the membership function uses multispectral data as input features (e.g., Sentinel-2 MSI bands), the membership matching may result in poor performances because some features detected using the high-resolution spectral data may not appear in lower resolution dataset.

Another factor concerns the coherence between the algorithm used for the clustering and the membership function characteristics. It has been proposed in the literature the use of Mahalanobis distance for spectral shape matching (Lubac and Loisel, 2007; Moore et al., 2014; Vantrepotte et al., 2012). As already noted, the advantage of Mahalanobis distance is that it considers scaling and correlation among the input features. However, if the rule for a point to belong to a cluster is the minimal distance to the cluster centroid, as it is the case for k-means and Fuzzy C-means clustering, the Euclidean distance should be preferred.

To test these hypotheses, several experiments have been conducted using the complete R_{RS} dataset. The first experiment considered a clustering using the K-means algorithm and the spectral dataset at “full” spectral resolution (i.e., 1nm steps) to assess how the assignment accuracy is affected by the spectral resolution of the sensor during the assignment step. For the assignment, the simple Euclidean distance considering an n -dimensional space has been adopted and the input features tested were 1) the same full hyperspectral resolution used for clustering; 2) a degraded

spectral resolution at 20nm steps; 3) Sentinel-2 spectral bands and 4) Sentinel-3 spectral bands. To assess the precision of the membership assessment, we used the accuracy score, which is defined as the number of correct assignments divided by the total of samples in the test (Eq. II-8). The results of this first test are presented in Figure II-13.

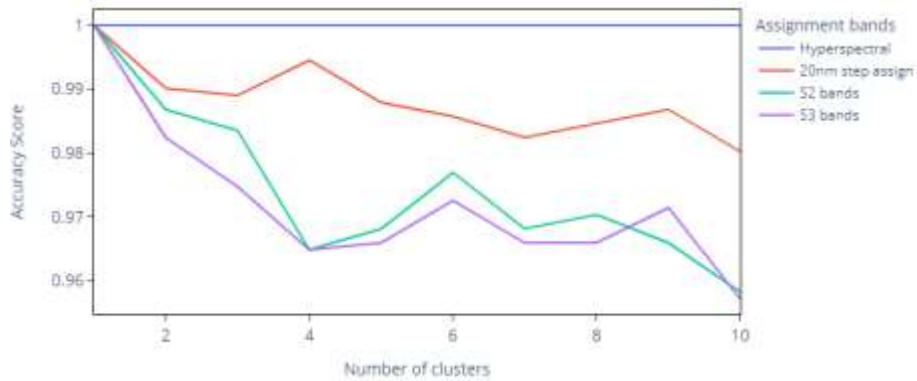


Figure II-13: Performance of the assignment accuracy (Euclidean distance) for different input features offering decreasing spectral resolutions. Cluster definition was realized using *k*-means and full resolution hyperspectral Rrs dataset.

It can be seen from Figure II-13 that the assignment accuracy slightly degrades as we coarsen the bands used for the assignment, and a perfect match (i.e., accuracy equals to 1) is obtained when the same bands are used for clustering and for the assignment. The assignment accuracy also decreases according to the number of clusters to identify, but this degradation is limited, representing less than 4% of loss of accuracy with a 10-cluster assumption.

As the final goal is to adapt the classification framework to Sentinel-2/MSI satellite images, input features used in the clustering step for the next experiments were constrained to the spectral bands listed in Table II-2. Setting the clustering input features to the Sentinel-2 MSI bands and clustering to K-means, Figure II-14 shows the impact of the distance function jointly with the spectral resolution during the assignment. Mahalanobis and Euclidean distance functions have been tested for membership assignment. We can see that Euclidean distance provides better assignment accuracy regardless of the bands used for clustering, with a perfect match (accuracy equals to 1)

when the assignment uses the same Sentinel-2 MSI bands as the clustering. This finding evidence that there must exist coherence among the spectral wavelengths used during the classification and the retrieval phases. Moreover, as already noted, there must also be a coherence between the distances considered in the clustering and assigning steps.

Figure II-15 presents an experiment similar to the previous one but employing the agglomerative method for clustering. It can be noted that, even when the clusters are not equidistant (clusters derived from the agglomerative method can have distinct sizes), the Euclidean distance surpasses the Mahalanobis distance in assignment accuracy for $K < 5$. For 5 clusters and more, the results from Euclidean and Mahalanobis mix up and are indistinguishable. Comparing results from Figure II-14 and Figure II-15, it is possible to note that regardless of the clustering algorithm used, the Euclidean distance is more stable in all scenarios, and the best accuracy is achieved when it is coupled with K-means clustering.

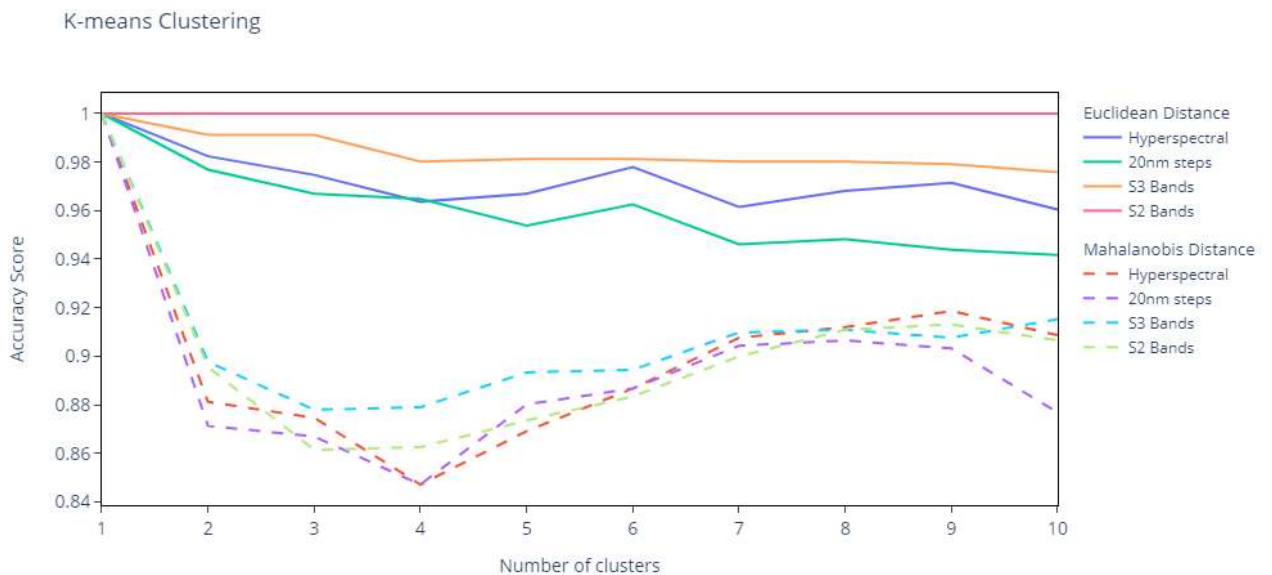


Figure II-14: Performance of the assignment accuracy as a function of distance metric used (Euclidean and Mahalanobis distances) for different input features offering decreasing spectral resolutions. Cluster definition was achieved using k-means applied to Sentinel-2/MSI spectral bands and for a varying number of clusters, from 1 to 10.

Agglomerative Clustering

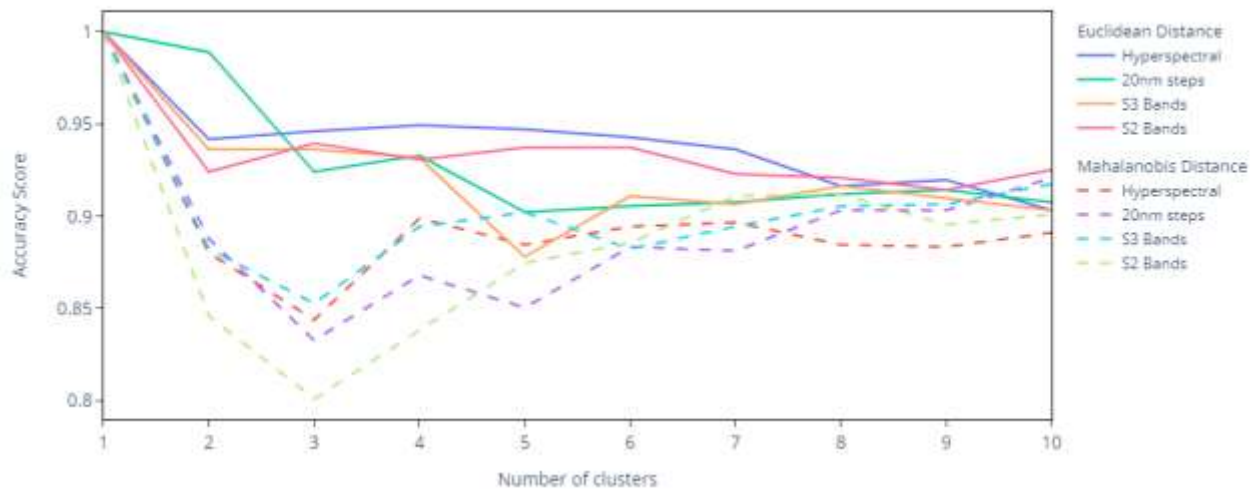


Figure II-15: Performance of the assignment accuracy as a function of distance metric used (Euclidean and Mahalanobis distances) for different input features offering decreasing spectral resolutions. Cluster definition was achieved using Agglomerative Clustering applied to Sentinel-2/MSI spectral bands and for a varying number of clusters, from 1 to 10.

Clustering experiments

We further analyzed how the OWTs definition driven by distinct clustering algorithms impacts the final SPM concentration inversion. Figure II-16 displays the retrieval error as a function of the number of clusters (OWTs) and for 3 different clustering methods. For this test, Sentinel-2 MSI spectral bands were used as input features for both clustering and membership assignment. Mahalanobis distance was used as it is the membership function primarily used in the literature (Moore et al., 2014; Vantrepotte et al., 2012).

Until $K=6$, no significant difference was noted among the algorithms in terms of SPM concentration retrieval performance. For 7 and 8 clusters, the algorithms K-means and FCM presented an increase in the retrieval error. This can be explained by the difference between the clustering methodology and the membership function, as discussed in the previous section. Considering K-means and FCM create clusters of equal sizes, the proper membership function would be the simple Euclidean distance. The agglomerative cluster seemed to adjust better to the

Mahalanobis distance and was used in the following experiments. It is worthwhile to note that from 4 clusters onwards, no improvement was detected in the retrieval error.

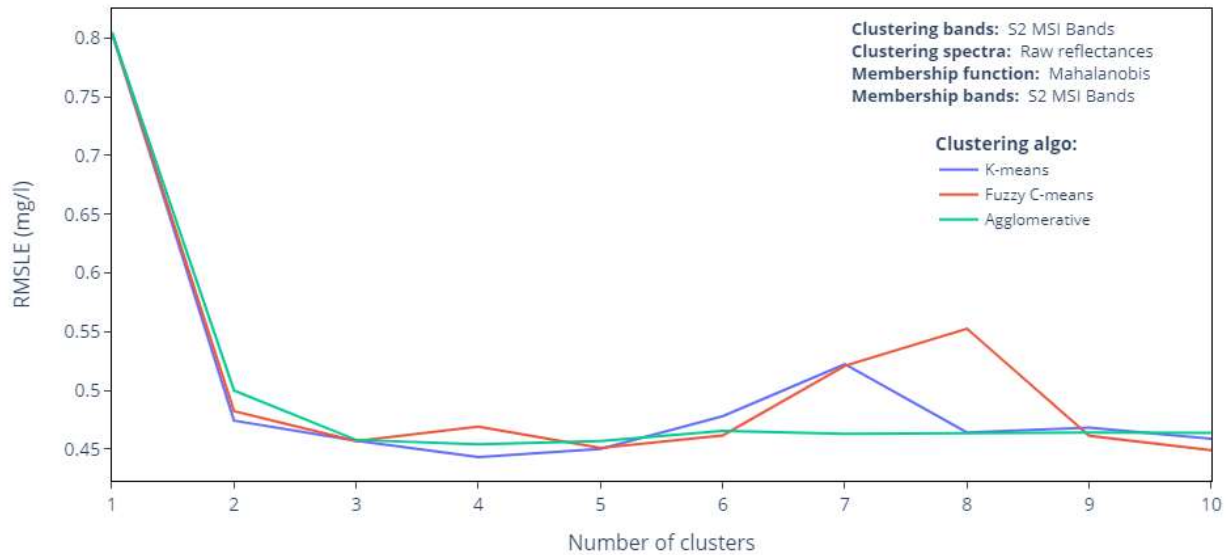


Figure II-16: Impact of the clustering step on SPM retrieval as a function of the number of OWTs and for three different algorithms: K-means, FCM and Agglomerative clustering. Retrieval performance is assessed using RMSLE on the R_{RS} validation dataset. The input features were R_{RS} sampled to Sentinel2 MSI spectral bands for both clustering and assignment (Mahalanobis distance).

Using normalized reflectance (NR_{RS}) as input features instead of the original R_{RS} data (Figure II-17), it took more OWTs to achieve the optimal performance and it did not provide any additional improvement in terms of overall SPM retrieval. It can be interpreted that for SPM retrieval it is more important the magnitude of the reflectances than their spectral shape.

An additional experiment compared the effect of different spectral sampling to be used as input for the clustering and the overall results (Figure II-18). The spectral curves used as input features for clustering were resampled to 10nm, 20nm and 30nm steps. Additionally, the Sentinel-2 and Sentinel-3 bands were also used as input features to the clustering algorithm.

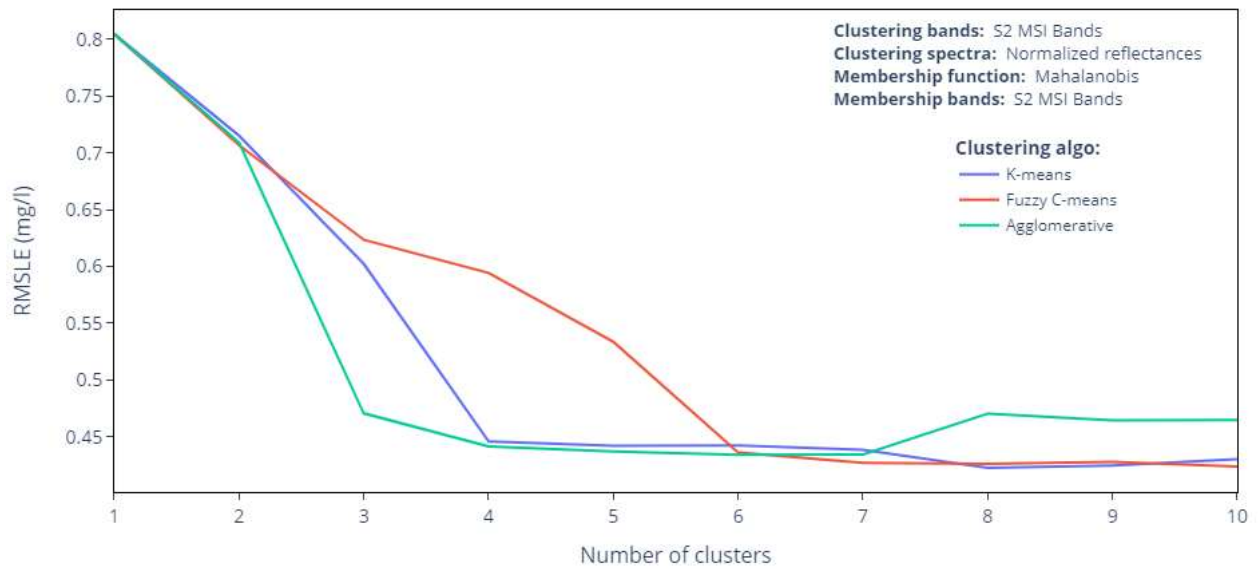


Figure II-17: Impact of the clustering step on SPM retrieval as a function of the number of OWTs and for three different algorithms: K-means, Fuzzy C-means and Agglomerative clustering. Retrieval performance is assessed using RMSLE on the R_{RS} validation dataset. The input features were normalized R_{RS} sampled to Sentinel2 MSI spectral bands for both clustering and assignment (Mahalanobis distance).

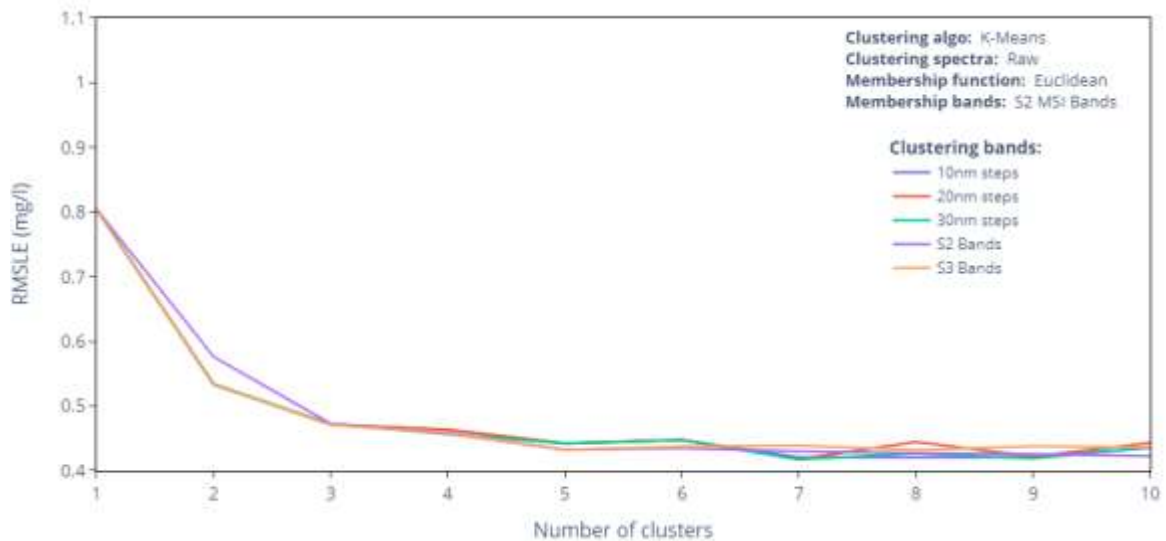


Figure II-18: Impact of the clustering step on SPM retrieval as a function of the number of OWTs and for R_{RS} input features sampled at decreasing spectral resolution: 10nm, 20nm, 30nm, Sentinel-2 and Sentinel-3 spectral bands. For this test, clustering used K-means algorithm while the assignment step was based on Euclidean distance and Sentinel-2 spectral resolution as input features.

Results showed that input features sampled at higher spectral rate during clustering did not increase the overall SPM retrieval performance when the subsequent membership function uses, as input features R_{RS} , spectral data with a coarser spectral resolution (i.e., Sentinel-2 MSI spectral bands).

Multiple Cluster model

Based on the results obtained in the previous section, the K-means algorithm and non-normalized R_{RS} were used to define the OWT classification delivering the most efficient SPM retrieval. The input features for clustering were set to the Sentinel-2 MSI bands to keep coherence between the clustering and assignment steps, and Euclidean distance was used for validation. Table II-6 presents the SPM retrieval performances (assessed on the validation dataset) using 1 to 10 OWTs.

Table II-6: SPM retrieval performance as a function of the number of OWTs. Accuracy was assessed on the validation dataset. OWT classification was realized using K-means algorithm and non-normalized R_{RS} , with spectral input features set to Sentinel-2 MSI spectral bands. Assignment used Euclidean distance and Sentinel-2 MSI spectral bands as input features.

	Number of OWTs									
Metric	1	2	3	4	5	6	7	8	9	10
R²	-24.35	0.82	0.89	0.93	0.94	0.94	0.95	0.95	0.95	0.95
RMSE	1688.55	143.47	110.52	86.08	82.10	80.30	73.58	74.91	73.77	77.12
RMSLE	0.80	0.58	0.47	0.46	0.43	0.43	0.43	0.43	0.43	0.42
RSR	5.02	0.43	0.33	0.26	0.24	0.24	0.22	0.22	0.22	0.23
PBIAS	159.36	14.75	10.20	0.93	-0.46	2.30	0.07	-0.24	0.32	-4.87

As expected, SPM retrieval accuracy appears to increase with the number of OWTs. However, as for previous examples (Figure II-16 to Figure II-18), the SPM retrieval accuracy does not improve much beyond a 4-OWT typology. In fact, defining more classes (i.e, 5 OWTs and beyond) does not imply systematic gain in the accuracy depending on the metric as for example for RMSLE, RSR or PBIAS. For example, SPM retrieval accuracy measured by RMSLE improved

from 0.46 mg/l to 0.43 when considering 4 and 9 OWTs. To quantify the significance in the improvement of the accuracy retrieval as a function of the number of input OWTs, we randomly repeated the experiment by re-shuffling the datasets and repeating the validation up to 10 times. Consequently, it was possible to assess a mean and standard deviation for each retrieval accuracy metric and OWTs configuration, allowing to use independent two-sample *t*-Student test for significance assessment. Student test was performed on pairwise adjacent OWTs configurations (i.e., 2-OWTs/3-OWTs, 3-OWTs/4-OWTs, and so on). The results showed that there was no significant difference (i.e., $p\text{-value} > 0.05$) between retrieval accuracy assessed with RMSLE when using 4 or more clusters. To confirm this finding, additional *t*-Student tests were performed to compare 4-OWTs typology to a larger number of OWTs, from 5 to 10. All the tests were performed by using the RMSLE metric and agreed that there was no significant accuracy improvement beyond the $K=4$ limit. There is probably a balance between complexity (number of OWTs and of retrieval models) and retrieval performance (low RMSLE). Accordingly, the final OWTs clustering that we propose is based on a 4-class OWT and Table II-7 presents the OWTs spectral characteristics (mean R_{rs} and the standard deviation) for the Sentinel-2 MSI spectral bands. The OWTs have been ordered ascending by median SPM values.

Table II-8 details the accuracy results and retrieval model for each of the four OWTs. The overall metrics for the validation dataset were $R^2 = 0.93$, $RMSE = 86.08$ mg/l, and $RMSLE = 0.46$ mg/l.

Table II-7: R_{RS} values, expressed as mean and standard deviation and for each Sentinel-2 MSI spectral bands, for the four OWTs that represents the optimal water type classification for inland water SPM concentration retrieval.

		Sentinel-2 Bands (wavelengths in nm)									
Cluster		443	490	560	665	705	740	783	842	865	940
0	Mean	0.0043	0.0058	0.0102	0.0077	0.0067	0.0024	0.0024	0.0015	0.0013	0.0008
	Std	0.0026	0.0036	0.0072	0.0058	0.0051	0.0020	0.0020	0.0014	0.0013	0.0009
1	Mean	0.0105	0.0146	0.0271	0.0323	0.0297	0.0140	0.0145	0.0097	0.0084	0.0043
	Std	0.0025	0.0032	0.0065	0.0064	0.0065	0.0056	0.0060	0.0050	0.0045	0.0028
2	Mean	0.0224	0.0287	0.0519	0.0722	0.0735	0.0530	0.0554	0.0433	0.0393	0.0183
	Std	0.0058	0.0068	0.0092	0.0103	0.0101	0.0091	0.0102	0.0107	0.0106	0.0052
3	Mean	0.0151	0.0198	0.0352	0.0459	0.0455	0.0317	0.0325	0.0258	0.0236	0.0130
	Std	0.0038	0.0045	0.0079	0.0088	0.0076	0.0057	0.0058	0.0068	0.0069	0.0058

Table II-8: Model type and accuracy for the 4-OWT classification retained as the most efficient spectral clustering for SPM retrieval. Models were calibrated using the training dataset and the accuracy metrics were calculated using the validation dataset.

Clusters	Model	Spectral Band	R²	RMSE (mg/l)	RMSLE (mg/l)	RSR	PBIAS (%)
0	Exponential	705	0.53	7.21	0.60	0.68	23.05
1	Power	b783/b705	0.82	119.73	0.34	0.42	18.55
2	Linear	865	0.97	23.56	0.10	0.16	-0.23
3	Nechad	b740/b705	0.92	121.81	0.19	0.27	-5.46
Overall metrics			0.93	86.08	0.46	0.26	0.93

Figure II-19a and Figure II-19b present, respectively, the mean R_{RS} for each OWT and the boxplot graph with the SPM variability within each of the 4 clusters. The median SPM concentrations for the clusters are: 6.8mg/l (cluster 0); 52.8mg/l (cluster 1); 188.7 (cluster 2); and 454.9 mg/l (cluster 3). In the measured versus predicted SPM concentration plot (Figure II-19c), it is possible to note that, although RMSLE was close to the one obtained from the hybrid baseline model, that R^2 value, RMSE and PBIAS were much improved, evidencing a better fit for the highest SPM concentrations.

Moreover, the R^2 value was close to the one obtained using one model for each basin/area (0.95) with the advantage of using just 4 models instead of 13. Additionally, it should be considered that the performance obtained from the multi-basin algorithm was tailored by the previous knowledge about the basin making it difficult to be used globally, while the 4 clusters algorithm considers automatically assignment, through the Euclidean membership function. Comparing these results with the second watershed-based model baseline (Table II-9), where we do not inform “a priori” the area being monitored, the 4 clusters algorithm matches R^2 and surpasses RMSE (86.08 vs 91.16 mg/l) and RMSLE (0.46 vs 0.59 mg/l) metrics.

Table II-9: Comparison of different classification schemas.

Classification Schema	R^2	RMSE (mg/l)	RMSLE (mg/l)	RSR	PBIAS (%)
Hybrid model (high/low)	0.87	97.62	0.44	0.36	7.50
Watershed-based (tailored by watershed)	0.95	74.36	0.46	0.22	0.22
Watershed-based (with automatic assignment)	0.93	91.16	0.59	0.27	-8.25
Spyrakos' OWT baseline	0.64	200.77	0.71	0.60	33.89
4 classes hybrid model	0.93	86.08	0.46	0.26	0.93

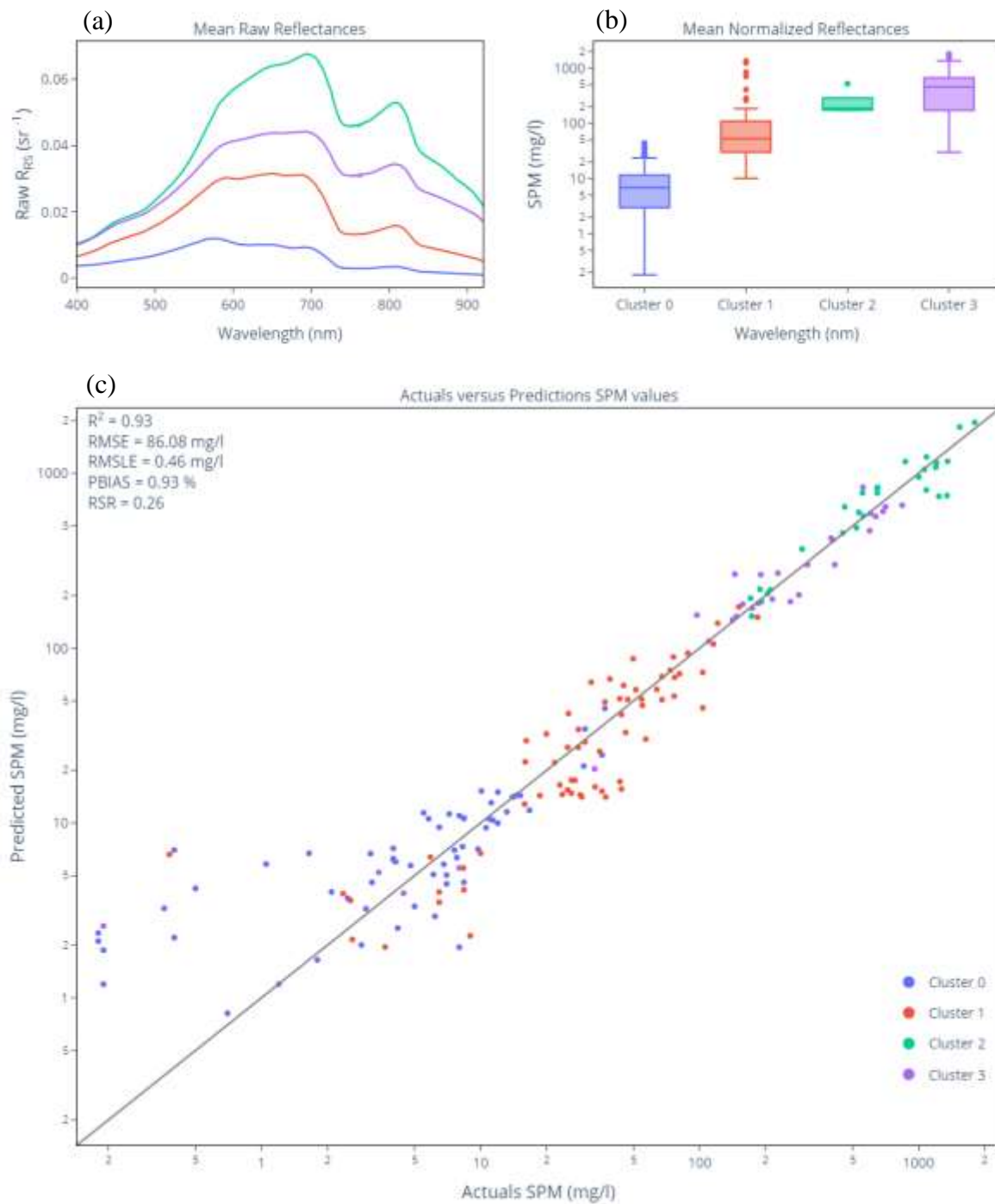


Figure II-19: optimal 4-OWT clustering for SPM concentration retrieval (a) Mean Rrs for each considered OWT; (b) SPM range for each OWT; (c) Actual vs prediction SPM concentrations in mg/l considering one algorithm for each OWT (Table II-9). The colors represent the four different classes.

1.5 Discussion

The analysis of a large dataset acquired over 13 catchments presenting contrasted conditions in the tropical region allowed to better understand the sensitivity of R_{RS} to SPM concentration over inland waters. Our results demonstrate that the use of a hybrid model based on 2 optical classes, as proposed by several authors (Han et al. 2016, Novoa et al. 2017), may not be enough accurate to robustly retrieve SPM load over a large range of concentration. However, our experiment demonstrates that the use of four optical classes allows to reach the best retrieval performances and that a larger number of optical classes does not results in significant accuracy improvement. This four-cluster classification matches the accuracy performance achieved using catchment-wise OWTs, but it does not require multiple models for each specific watershed to be studied.

Looking at the dependency of the results to the spectral resolution, our results show that there are no differences in the SPM retrieval performance if we increase the spectral resolution for clustering if this increase is not accompanied by the assignment and fitting steps. Most important, we observe the need for coherence between the OWT definition (clustering algo, and clustering bands) and the retrieval algorithm (membership function and sensor bands). Moreover, unlike photosynthetic pigments, such as Chl-*a*, that exhibit complex spectral absorption patterns, light absorption and scattering by suspended particles present smooth spectral behavior and does not present much variability as a function of wavelength, especially beyond the green range. This may explain why the spectral resolution may not play an important role for the definition of the best SPM retrieval method.

The most efficient optical water type typology matches a simple 4-class membership, which are mainly function of increasing SPM concentration except for the two last clusters (Cluster 2 and

Cluster 3, see Figure II-19b) that overlap in terms of mean SPM concentration. For the latter two, it appears that the difference may be more related to the suspended sediment type (expressed as median particle diameter -D50) than to its concentration as Cluster 2 samples match the finest suspended sediment median size, as shown in Figure II-20. The median D50 diameter for each cluster is: 16.75 μm (Cluster 0); 11.05 μm (Cluster 1); 5.43 μm (Cluster 2); and 7.90 μm (Cluster 3). This observation confirms that SPM retrieval is not fully dependent to the concentration range but also to the particle type or size distribution. Figure II-21 shows that the final OWTs covered most of the catchments and that each area usually may be represented by at least 2 OWTs. In fact, most individual catchments/region are represented by two different OWTs, which may explain why simple 2-class OWT hybrid models may reach fine accuracy over limited region but fail to represent all the water types in a much broader area.

Regarding the retrieval model type, most of the equations rely on power-like equations with more than half of them making use of spectral band ratio. It is worthwhile to note that most of the best retrieval models use NIR wavelengths especially in the 700-800 nm range. The use of this narrow spectrum range demonstrate that the new spectral bands brought by Sentinel-2/MSI sensor (e.g., bands 5 to 7) really represent an improvement towards better SPM mapping using resolution images in comparisons to more conventional platforms such as Landsat that offer only one spectral band in the red spectrum and one in the NIR spectrum beyond 800 nm.

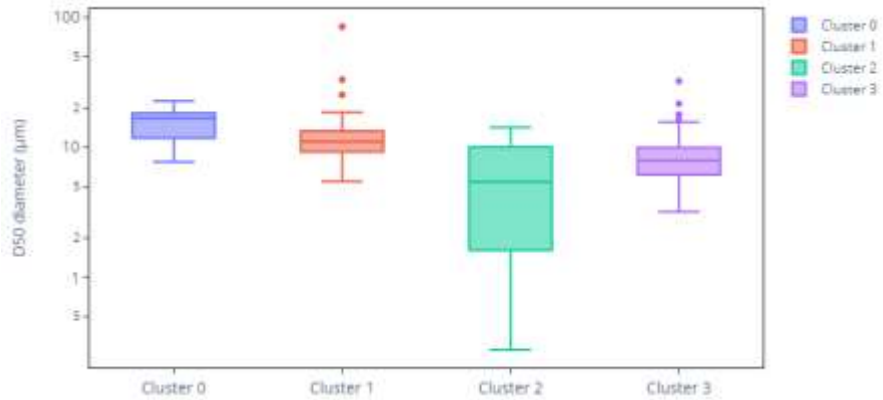


Figure II-20: Suspended sediment particle diameter (D50) distribution for the four optimal OWTs configuration, expressed in μm .

Analyzing the results obtained from the Spyrakos' OWTs baseline, which considers 8 different models, it is clear that those OWTs are not optimal for improving SPM retrieval. All analyzed metrics using Spyrakos' OWTs are worse than the simple hybrid (High/Low SPM concentration) OWT clustering. This poor performance can be credited to the normalized reflectances used to create the Spyrakos' OWTs. The normalized reflectances tend to privilege the spectral shape in detriment to its magnitudes, which is mostly useful to discriminate different water types. However, SPM concentration tends to affect the light scattering in the NIR range (especially 700nm to 800nm) regardless of the water type, and the magnitude of the reflectance, as well as the ratio between upper NIR bounds (800nm to 860nm) and lower NIR bounds (700nm to 740nm), seems to play a bigger role in SPM retrieval than the spectral shape. Therefore, reflectance magnitude should be preferred during the OWTs definitions.

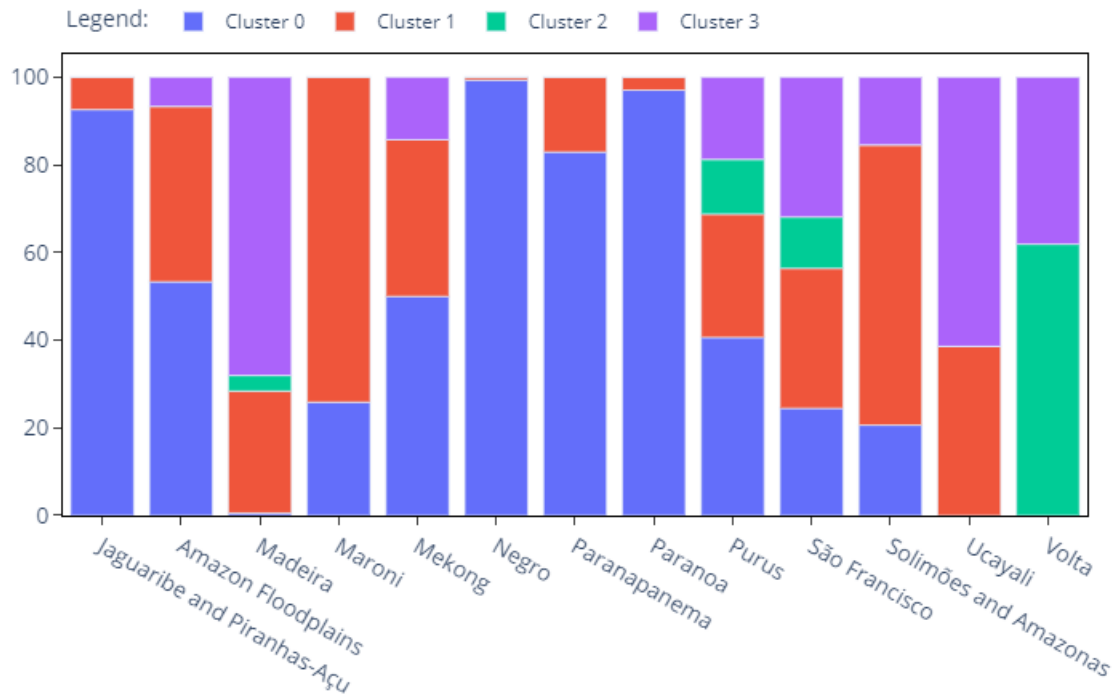


Figure II-21: Geographical distribution of the OWTs (“clusters”) accordingly to the 13 hydrological regions. OWT frequency was normalized to 100 % for each region.

1.6 Conclusion

Conventional monitoring of suspended sediment, as well as other water quality parameters, in inland waters do not offer adequate spatial and temporal resolutions due to the high costs involved in field surveys. The present study used a dataset of more than 1,000 hyperspectral measurements, coupled with corresponding field measurements, to propose a new framework to support the use of water color remote sensing images for SPM monitoring. The hyperspectral dataset has been collected in 13 different sub-watersheds in tropical regions, with SPM concentrations ranging from 0.1 to 1,800 mg/l.

The article analyzed the viability of using models tailored by water types to improve overall SPM retrieval accuracy. In this regard, three “baselines” OWTs have been defined: 1) a 2-class typology (low and high SPM concentration), as proposed in the literature (Han et al., 2016; Novoa

et al., 2017); 2) OWTs tailored to rivers or catchments; 3) a 13-class OWT for inland waters as previously proposed by Spyarakos et al. (2018).

To build up on these baselines, multiple combinations of clustering algorithms (K-means, Agglomerative, and Fuzzy C-Means) and input features (i.e., R_{RS} vs. NR_{RS}) were tested into an iterative process to achieve the best overall SPM retrieval accuracy. For each clustering test, unsupervised fitting of SPM retrieval models was achieved for each new OWT.

The results show that only 4 classes would be enough to achieve results that are similar to the one fitting one model for each specific river or catchment. The hybrid 2-class algorithm is a good alternative, but it fails to deliver good results, especially in the higher concentration points (above 1,000mg/l), probably due to a saturation effect on the NIR reflectance, as already discussed in the literature (Dogliotti et al., 2015; Nechad et al., 2010). Interestingly, the more detailed water types definition proposed by Spyarakos et al. (2018), with 13 distinct inland OWTs, delivered the poorest results. This weak performance is also seen in our clustering experiments, which showed a disadvantage of using normalized spectra for SPM retrieval.

To confirm the use of a 4-classes model in detriment to a higher number of OWTs, a comprehensive *t*-Students test has been performed in order to verify if there were significant error improvements by augmenting the number of clusters beyond this number.

This study also demonstrates that most models rely on NIR wavelengths, especially between 700nm and 800nm to deliver the best accuracies, regardless of the water type or median SPM concentration. Another important finding refers to the impact of the particle size of the sediment on the light scattering. According to the 4-classes proposed in the Multiple Cluster Model section, Cluster 2 appears with the highest reflectances, but it does not correspond to the higher

SPM concentrations, which appears on Cluster 3. However, Cluster 2 presents the smaller particles, as depicted in Figure II-20.

In summary, this study makes use of unsupervised Machine Learning methods, such as clustering algorithms, to maximize the results obtained from remote sensed data for inland water quality assessment. In the end, we present an operational 4 classes OWT framework designed specifically to improve SPM retrieval through Sentinel-2 MSI sensor. Although designed specifically for SPM concentration, a similar workflow could be used to improve the performance of other water quality parameters retrieval with distinct sensors. In particular, it is expected that the results of this work may be applied straightforwardly for turbidity monitoring which is an optical parameter mostly related to suspended sediment fluxes in inland waters and widely used in limnological studies for water quality trend assessment.

Acknowledgments

The field spectral measurements were acquired previously to this work in the frameworks of different research projects, MSc and PhDs works from 2008 to 2019. In particular, we thank the Brazilian National Water and Sanitation Agency (ANA), the French National Research Institute for Sustainable Development (IRD), the Brasilia University (UnB), the Federal University of Amazonas (UFAM), the Geological Survey of Brazil (CPRM) and the Geosciences Environment Toulouse (GET) laboratory for sharing the radiometric hyperspectral data and related water quality measurements.

Chapter III – Generalization and Analysis: A case study of Brazil’s 2021 drought

The main focus of this chapter is to demonstrate the feasibility of applying remote sensing techniques based on Sentinel-2/MSI high-resolution time series to assess water surface and some water quality parameters at a regional level and over an extended period.

As described in the Introduction, to the best of our knowledge, there are no operational tools available to complete these tasks in an automated way. When such a study is required, geospatial experts usually craft the processing and analysis steps manually, and the area of interest is generally cropped to very localized areas. Although it works for specific studies, this approach is inefficient for operational monitoring over large extents.

New challenges arise when scaling up the analysis to a regional level, from the initial processing that would require the acquisition of vast amounts of data and corresponding ancillary data to the final analysis that requires combining all the generated information into a meaningful result.

Supervised analysis through geospatial packages is impracticable due to the amount of data, and automation is strictly necessary. Furthermore, the methods developed to work into single scenes or at reservoir levels are insufficient for this new task, and new processing and analysis methodologies are required.

To operate with tens of thousands of inputs assets (i.e., reservoirs, river reaches) and derived subproducts, Spatio-Temporal Assets Catalogs (STAC) must be organized to keep track of the assets’ semantics and corresponding geospatial localization. Besides, Big Data concepts such as

cloud computing, lazy execution, and parallel computing must also be employed to process multi-dimensional cubes at this large scale.

In this context, the present chapter is divided into two parts. It starts with an introductory section to present in detail the methodologies developed in the context of the thesis to cope with these challenges and would not fit in the case study article due to length restrictions imposed by scientific journals. The methodology section covers the water detection step that goes from the satellital input data to the resulting inland water masks. The water surface and quality assessment section covers the stages that were used to aggregate, post-process, and subsequently analyze these data.

In the second part of this chapter, we propose an article draft that uses the presented methodologies to assess the impacts on the water surface and water quality caused by the drought that hit Brazil's southeast in 2021. The area of interest is within the Parana hydrographic region, which is very important for Brazil's economy, representing 40% of the country's Gross Domestic Product (GDP) and 45% of the country's electric power generation (ANA, 2018). An area composed of 36 Sentinel-2 tiles has been selected for the study. Each Sentinel-2 tile covers a square area of approximately 10,000 km² (100 x 100 km), totalizing 320,000 km², representing more than the half of French territory. The study considered four years of data, from 2018 to 2021, for which we had an average of 5-day coverage from both Sentinel-2 satellites. Before that period (2015-2017), Sentinel-2A delivered mostly 10-day interval images which was considered inconsistent for water body monitoring. Considering this time frame and the temporal resolution of the Sentinel-2 mission (one image every five days), over 12,000 images have been processed. At the end, the results for water surface and water quality have been combined, validated with field measurements, and interpreted accordingly.

1. Water Detection and Assessment Methodologies

1.1 Large scale water detection methodology

The **waterdetect** algorithm, when initially developed in 2019, has been validated over 15 Sentinel-2 high-resolution selected scenes over continental France (Cordeiro et al., 2021). As discussed in the article (Chapter 1), the challenge of classifying the pixels in a Sentinel-2 scene in a short processing time required certain considerations in terms of performance. The adopted solution included a random subsampling of the scene to cut down the number of pixels from more than 100 million (single Sentinel-2 scene) to tens of thousands and a subsequent generalization through a supervised machine learning classifier (Naïve Bayes). These solutions enabled fast processing of each image while keeping accuracy high, as demonstrated in the article.

However, the current case study presents a higher order of greatness regarding coverage area and time frame. Although the original algorithm has served as a starting point, it needed to be adjusted to face this more significant challenge towards scalability and robustness to provide a fully operational execution over tens of thousands of images. The following sections provide a high-level summary of the processes implemented on top of the original algorithm.

Multiprocessing

Due to the amount of data necessary for the study and to speed up the process, the **waterdetect** algorithm has been reimplemented to take advantage of the public geospatial clouds available, notably the Google Earth Engine (GEE) and the Microsoft Planetary Computer. Various processes were developed to work in parallel. To avoid storage costs involved in downloading all the 12,000 Sentinel-2 images necessary for the case study, which would take approximately 23Tb of disk space, just the necessary bands were downloaded from the Planetary Computer on-demand.

To save time, while the bands are being downloaded, a distinct thread is responsible for processing the clouds and cloud shadows on GEE, and just the final mask is downloaded locally, as explained in the Cloud and shadow masking section (Page 151). A third concurrent thread is also used, and it is responsible for calculating the probability of sunglint intensity from sun and satellite viewing angles with the information from the granule’s metadata, as described in the Glint Processor section (page 148).

For the water detection step, other improvements have been made to parallelize time-consuming tasks such as clustering and calculating the cluster metrics (i.e., Calinsk-Harabasz index) that are needed to find the best number of clusters (K). Instead of looping through each possible K value, ranging from 2 to 10, concurrent processes are submitted, improving the overall performance. A schematic diagram of the entire solution is presented in Figure III-1.

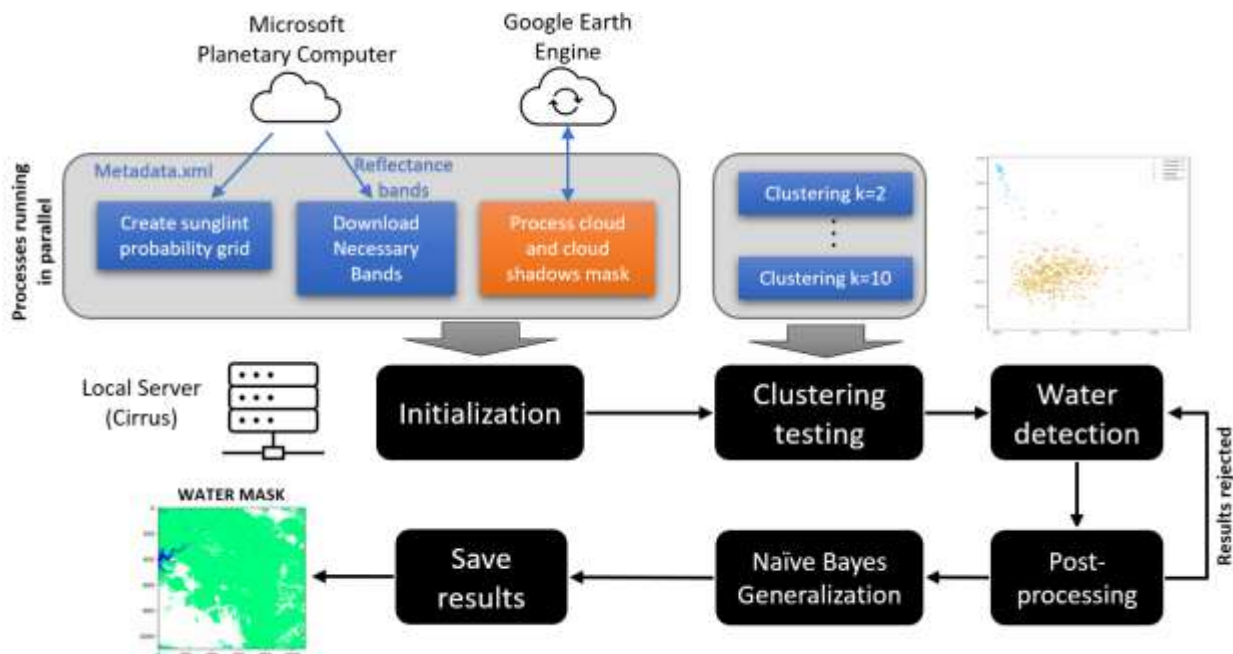


Figure III-1: Schematic diagram of the water detection process. Blue boxes represent different threads. Gray boxes group the threads that run in parallel. The orange box represents a process that runs entirely in the cloud.

In the end, the water mask is saved with one of the three possible values: 0 – not water; 1 – water; and 255 – no/masked data. Additionally, a simple report for each scene is also provided with quick views and some scatter plots for investigation purposes (Figure III-2).

The water detection step has been performed in a local server (Cirrus) and used specific Application Program Interfaces (APIs) to access the Microsoft Planetary Computer and the Google Earth Engine. With these improvements towards robustness and performance, taking advantage of

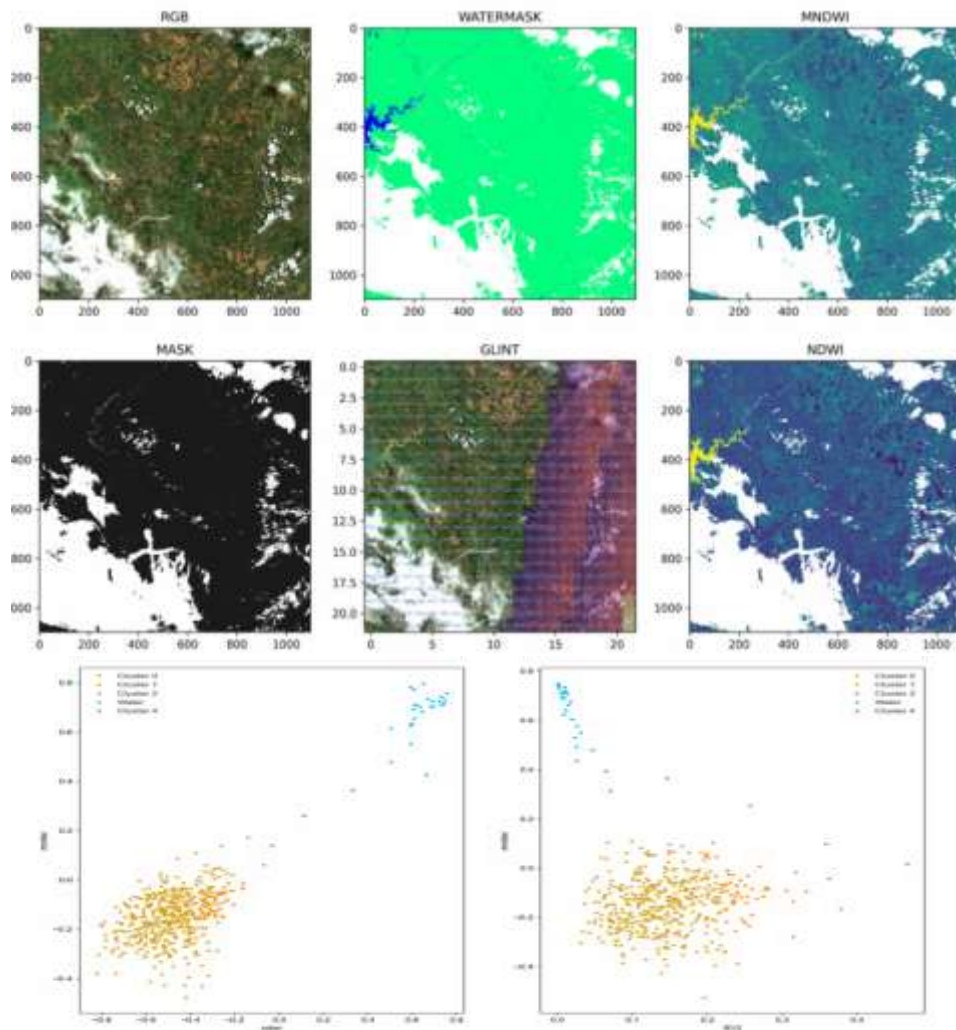


Figure III-2: Examples of the reports saved for each Sentinel-2 images during the water detection step. (a) Scene overviews with true color (RGB) image, final water mask, no/masked data, sunglint probability and the MNDWI and NDWI indices. (b) NDWI x MNDWI and MNDWI x B12 scatter plots with final clustering results.

these public cloud platforms, each Sentinel-2 scene is processed in approximately 1 minute. For the whole period and region (12,000 images), two weeks of processing were necessary.

Glint Processor

The optical remote sensing analysis depends on understanding the absorption and scattering of solar radiance processes on ground targets. As each material will absorb and reflect energy differently depending on the considered wavelength, the reflectance spectra are usually used to distinguish and measure the properties of the ground targets. However, many undesired effects can interfere with the observed reflectance when it is measured from space-borne sensors. One of these effects, called sunglint, is caused by specular (mirror-like) reflectance of the incident solar radiance, on water surfaces, in the direction of the sensor's field of view. The occurrence of the sunglint will depend on the zenithal (θ_s and θ_v for source and viewing zenithal angles, respectively) and azimuthal (Φ) angles (Figure III-3). In these cases, the sensor will not measure the characteristics of the observed target but rather a mixture with the spectra of the emitting source (sun).

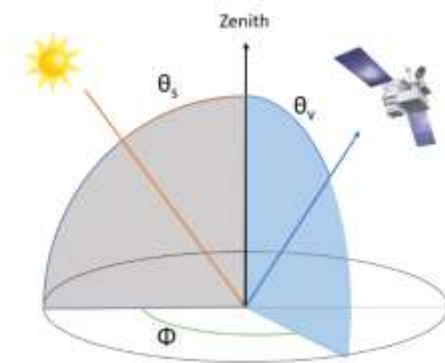


Figure III-3: Schematic representation of angles and geometry between light source, target and viewing sensor.

In water detection application, this sunglint will raise the reflectance values across all wavelengths, specially NIR and SWIR regions, which are usually used to identify water because

of the high absorption on these bands. This would cause failure for the algorithm to detect water surfaces in these highly glinted waters (Figure III-4).

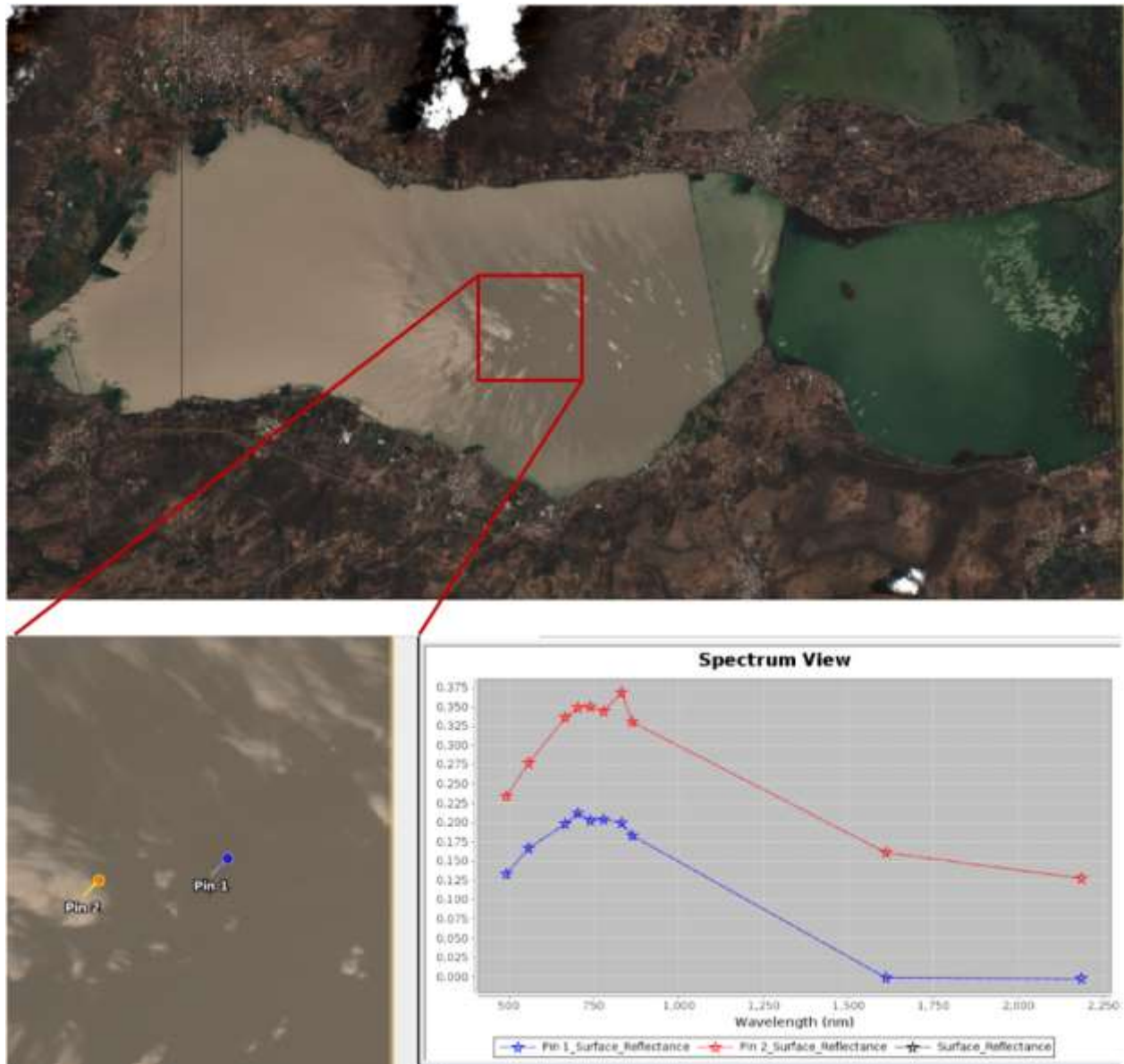


Figure III-4: Effect of the sunglint in a Sentinel-2 image on the surface of the Lake of Cuitzeo, Mexico.

Although there exist de-glinting techniques for Sentinel 2 in the literature (Harmel et al., 2018) that have been used in the case study's water quality assessment, the process is costly and requires prior knowledge of the water pixels in the scene. One alternative solution to minimize this problem is to use sun and viewing angles to identify sub-regions in the scene which are prone to

have sunglint and adjust **waterdetect** parameters (notably the NIR and SWIR thresholds) accordingly. The methodology selected to estimate the probability of sunglint is the one proposed by Giglio et al. (2003) for the Moderate Resolution Imaging Spectroradiometer – MODIS sensor, and that has been implemented for several works dealing with water quality retrieval (Espinoza Villar et al., 2013; Martinez et al., 2009; Santos et al., 2018):

$$\cos(\theta_g) = \cos(\theta_v) \cos(\theta_s) - \sin(\theta_v) \sin(\theta_s) \cos(\phi), \quad \text{Eq. III-1}$$

where θ_v is the zenithal viewing (sensor) angle, θ_s is the zenithal source (solar) angle and ϕ is the azimuthal angle between the source and sensor. The angle θ_g is the sunglint angle, and the lower its value, the higher probability of sunglint in the sensed image.

All the required inputs necessary for calculation are available in the Sentinel-2 granule's metadata file in grids of 5km x 5km (23 x 23 grid). This way, a glint angle map can be produced for the whole scene, as represented in Figure III-5. The glint angle is then used to adjust the thresholds in each grid cell to fine-tune the water detection process. The threshold value has been adapted from the previous works dealing with water quality retrieval cited above.

The sunglint processing has been added to the **waterdetect** package and it can be turned on/off through the software's configuration file.

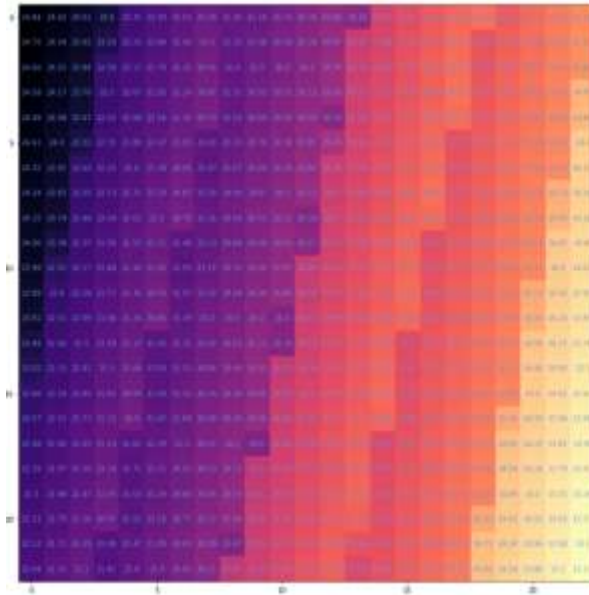


Figure III-5: Representation of the glint angles in the 5km grid cells. The lower the sunglint angle (yellow color), the higher probability of sunglint in the region. Angle values higher than 25° are considered exempt from sunglint (dark colors)

Cloud and shadow masking

As mentioned in Chapter I, the **waterdetect** package was originally designed to work with surface reflectance images atmospherically corrected by the MAJA processor.

As MAJA was originally developed to be a multi-temporal cloud detector (Hagolle et al., 2010), its performance concerning cloud screening surpasses most other algorithms available for Sentinel-2 imagery (Baetens et al., 2019).

However, MAJA-corrected products are unavailable globally, and our area of interest was outside MAJA coverage. Therefore, the solution was to use Sen2Cor corrected images, which are available globally in the Microsoft Planetary Computer. However, the cloud masks provided by the Sen2Cor processor were not reliable enough for the automation and could deteriorate the performance of the **waterdetect** algorithm. To overcome this limitation, an adaptor for the s2cloudless cloud masks has been developed. The s2cloudless is a pixel-based cloud detector that

uses a Random Forest algorithm trained with MAJA cloud masks as a reference on 14,140 Sentinel-2 scenes covering 77 different countries from Europe, Asia, Africa, North and South Americas, and Oceania. This machine learning approach is able to deliver state-of-the-art results surpassing the traditional processors such as MAJA, FMASK, and Sen2Cor (Zupanc, 2020).

Instead of running the trained classifier on the selected scenes, we opted to get the processed results directly from the Google Earth Engine platform, which is available in the form of a Cloud Probability mask, with values ranging from 0 (low cloud probability) to 255 (high cloud probability). In addition, a cloud threshold has been defined as 60 and pixels above this value were marked as clouds.

One problem concerning clouds shadows detection remains, as the S2cloudless is limited to identifying just the clouds. This is an essential step as shadows are usually misinterpreted as water in many algorithms (due to their low albedo), and that could also deteriorate the performance of the **waterdetect**. To overcome this limitation, a potential cloud shadow area is projected on the ground in the opposite direction given by the solar azimuthal angle (\emptyset) for the scene. This information is available in the image's metadata. As the cloud altitude is unknown, the solar zenithal angle is discarded, and an arbitrary distance of 1 km is used to project the potential shadow area in the ground. Finally, to derive the final cloud shadow layer, dark pixels on the near-infrared (NIR) band, herein defined as $NIR < 0.15$, that lie within the potential cloud shadow area are marked as actual shadows (Figure III-6). This process represents a simplified version of the cloud shadow object matching implemented in FMASK (Qiu et al., 2019; Zhu and Woodcock, 2012), and it is entirely performed within the Google Earth Engine platform before downloading the masks, following the code implementation is proposed by Braaten (Google Developers, n.d.).

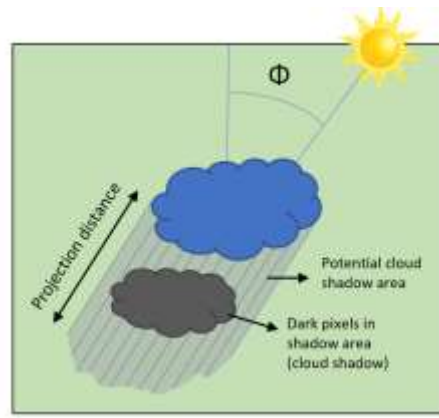


Figure III-6: Representation of the cloud shadow matching algorithm used to identify shadows on the ground.

The cloud and shadow masks are then downloaded from the GEE platform through another package, **GEES2Downloader** (<https://github.com/cordmaur/GEES2Downloader>), developed within the scope of this thesis. To download images using HTTP protocol and overcome size limitations imposed by the GEE platform, the **GEES2Downloader** subdivides the original asset into smaller tiles, download them in parallel and then recreates the original asset in the client computer.

Average Sampling and Post-Processing

In terms of robustness, one last improvement in the water detection process was implemented to avoid instabilities during the clustering process and water identification. Such instabilities may occur when the scene has little or no water or when there are mainly water pixels. As the clustering is done with a random subsample of the image, the water pixel percentage can affect the algorithm's performance (Cordeiro et al., 2021). Although not usual for an entire Sentinel-2 scene, this can occur, for example, when clouds dominate the image, and there is no water to be classified or when the cloud mask fails, and cloud pixels remain in the clustering process. One possibility to smooth out the results regarding water surface extent is to define a multitemporal window and average the results, as proposed by Peña-Luque et al. (2021). Instead, another solution has been implemented to deal with such scenarios. The water cluster parameters

(i.e., mean MNDWI and mean SWIR values) are analyzed during a post-processing phase. The cluster identified as the water cluster can be rejected if these parameters' values do not match an average water spectral signature, herein considered as mean MNDWI value above 0.2 and mean SWIR below 0.2. If this happens, a new sampling is made, ensuring that a minimum percentage of water (defined as $MNDWI > 0.4$) is present in the subset. This process is repeated for minimum percentages ranging from 20% to 45% in 5% steps and the scene is discarded if no solution passes the post-processing parameters.

1.2 Water surface and quality Assessment Methodology

One key difference between the present study to other surface water assessments at a regional level is that instead of using low spatial resolution to cover a larger area (e.g., Khandelwal et al., 2017) we are producing the water masks in full spatial resolution (10m) in order to maintain the ability to detect reservoirs as small as 0.5 ha that **waterdetect** can map with fine accuracy (Cordeiro et al., 2021). Keeping this resolution permits the assessment of water surface and water quality of thousands of reservoirs stratified by size ranges. The methodologies used to combine all these output data to deliver meaningful results are presented in this section.

Data Processing

Considering the number of files to be processed and intermediate sub-products, a Spatio Temporal Assets Catalog (STAC) has been created to organize all the assets. The advantage of STAC catalogs is that each asset is referenced with its own metadata. The metadata can describe the semantics of each asset stored in the catalog. Additionally, geospatial attributes such as scene geometry, footprint, coordinate reference system, and other properties are saved into XML files for quick search. This way, it is possible to find assets and combine them into the same projection effortlessly for further operations.

Once the catalog was prepared, the monthly water extent maps were created for each one of the 36 tiles and 48 months, ranging from 2018 to 2021. A total of 1,728 monthly maps were created by combining all images of the same month into a data cube, then applying the median operator Figure III-7a. No data values due to clouds or masked pixels have been skipped during the median operation. The use of time windows for water detection has been the object of other studies. Peña-Luque et al. (2021) analyzed time windows ranging from 10 to 20 days on Sentinel-2 and Sentinel-1 data and found significant improvements in large-scale water maps and reservoir area estimation.

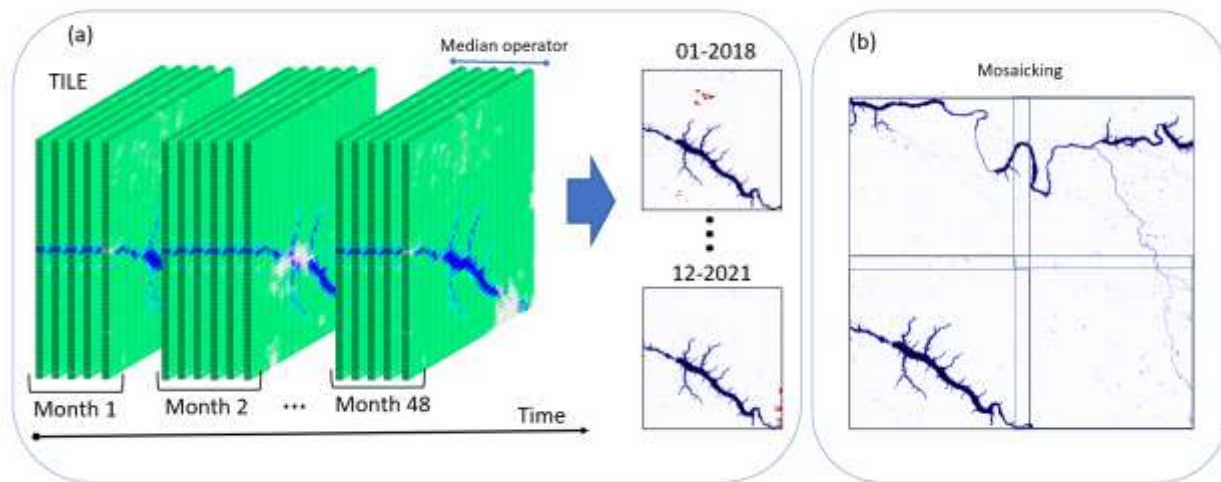


Figure III-7: (a) Schematic view of the monthly water masks process creation for each tile. (b) Schematic view of the monthly masks combination into a single mosaic.

These maps were then combined into monthly mosaics, covering the whole area of interest. The overlapping regions were treated with the same median operation to avoid discontinuity along the borders Figure III-7b.

Even with one image being sensed every five days by the Sentinel-2 constellation, approximately, the monthly mosaics can still present many no-data pixels, due to the high persistence of clouds in the region, blocking the satellites' view. This happens mainly in the wet season, from October to March, when rains are abundant in the region. Figure III-8 presents the

permanently cloud covered pixels for January 2018 that represented the worst scenario during the whole time series with about 6% of the scene presenting no data pixels.

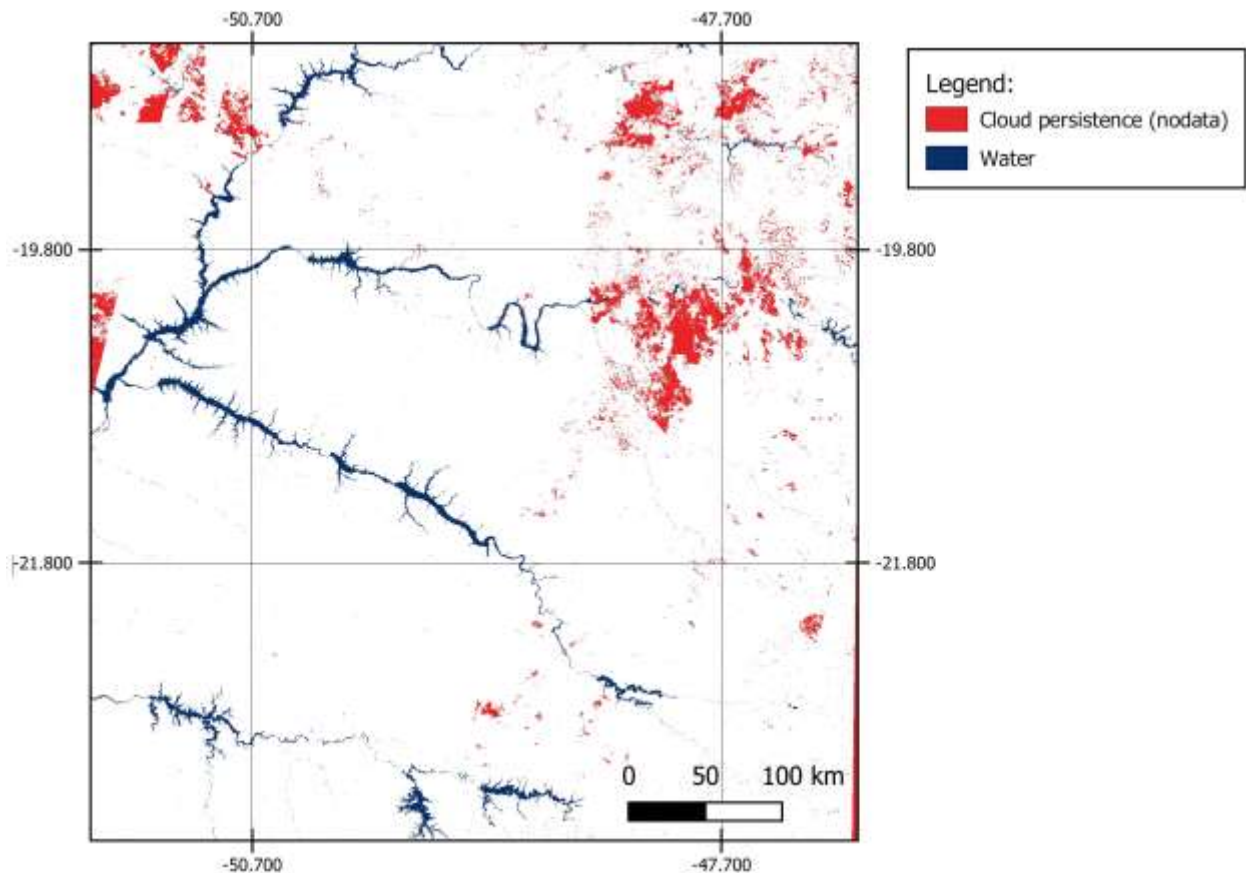


Figure III-8: Water surface extent mosaic for January 2018. In red, the pixels that remained covered by clouds or without valid data during the whole month.

Data Interpolation

Although not covering a huge percentage of the scene, these no-data areas can create discontinuity in the time series and impact the overall water surface being measured. Figure III-9 shows the total water surface extent for the area of interest in the blue line and the percentage of no data pixels in the red bars.

It can be noticed that the months with higher cloud persistence (higher amount of no data pixels) present discontinuities. This discontinuity was especially significant in January 2018 and December 2018, when cloud persistence reached over 5%.

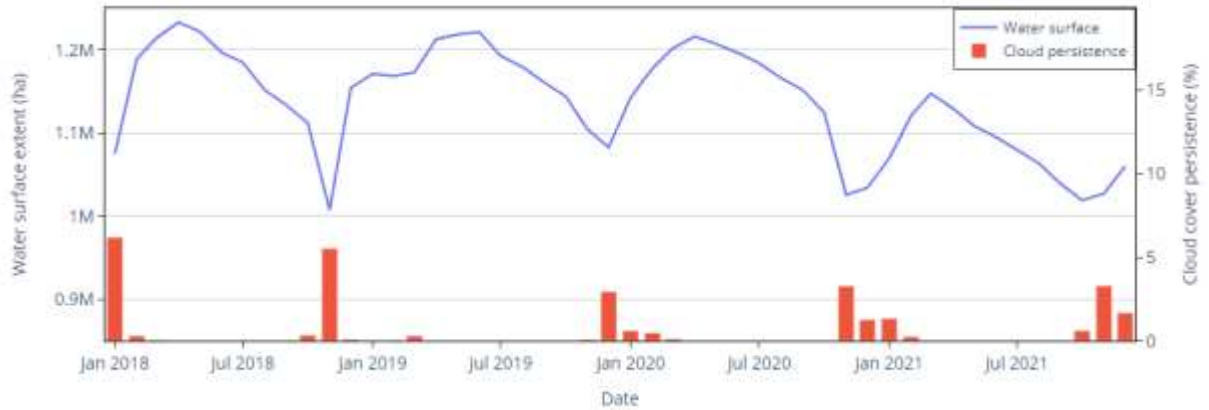


Figure III-3: Monthly water surface estimation in the area of interest and the percentage of invalid pixels due to cloud persistence.

Two additional steps were taken to deal with these remaining gaps in the times series. First, the Boolean values (0 – not water, 1 – water) have been replaced by real numbers through image downscaling. Each group of 10x10 pixels (100m x 100m), corresponding to 1ha, has been merged through the average operator and skipping no-data values. This first step result is a lower image resolution (100m) with values ranging from 0.0 to 1.0, representing the water cover percentage in the pixel area (1ha). This resolution remains consistent with our objective to analyze water body dynamics for an area of about 1ha. In this specific case, the pixel values can be directly interpreted as water surface in hectares. This processing has been done for each monthly mosaic. However, the impact of missing pixels because of cloud coverage can still affect the surface area calculation.

To overcome the problem of the missing pixels, the second step consisted of interpolating gaps linearly through time. For that, a data cube consisting of the 48 mosaics has been stacked and a linear interpolation has been performed along the time axis (i.e., for each coarse pixel), as shown in Figure III-10. As the first (January 2018) and last (December 2021) months of the series cannot be filled through the interpolation due to a lack of data outside its bounds, an additional step has been taken to fill these gaps. The first month has been filled with a backward-fill method, which

propagates the values backwards in time to fill existing gaps. In this case, remaining gaps occurring in January were filled with values present in February. Conversely, the last month has been filled with a forward-fill method that operates the opposite direction. Then, the total water extent area was obtained by adding up the values for each pixel in each considered month. It is worthwhile to note that both spatial resolution degradation and time mosaicking can be adjusted depending on the area of interest of the seasonality of the hydrological processes.

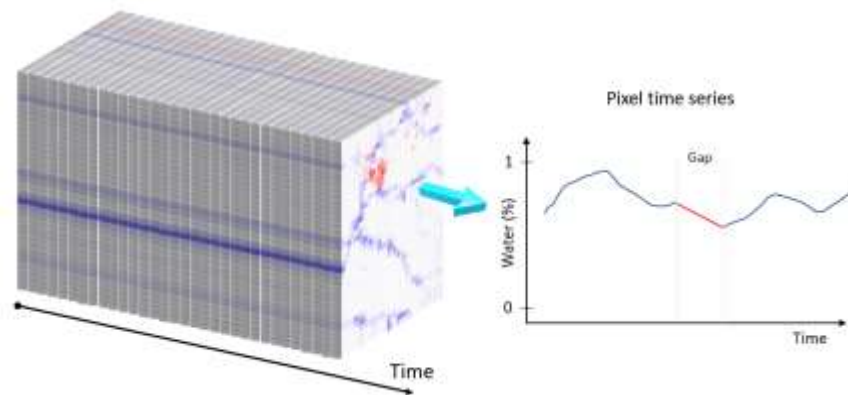


Figure III-10: Schematic representation of the pixel level interpolation along the time axis.

Anomaly Maps

To produce spatial water surface anomaly maps, the interpolated cube described in Data Interpolation section has been used as a starting point. As already explained in the previous section, this data cube has a downscaled resolution (10 pixels by 10 pixels), where each pixel has a real value representing the water surface in hectares. This original cube has been coarsened one more time to create regions with 200 by 200 pixels, using the average method, so the new coarse pixel value would represent the water percentage over a 40,000ha area.

Considering that the water surface is unevenly distributed spatially, we have coarse pixels in areas with large water surfaces and other ones with almost complete absence of water coverage.

In this scenario, the anomalies cannot be presented in absolute values. Otherwise, they would highlight the pixels that touch or contain large reservoirs and ignore the pixels with smaller water bodies within. Instead, they will be represented as the z -score metric, that is the number of standard deviations away from the mean. The z -score for each coarse pixel can be computed through the following equations:

$$z_i = \frac{x_i - \bar{x}}{\sigma}, \quad \text{Eq. III-2}$$

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}, \quad \text{Eq. III-3}$$

where σ is the standard deviation for each coarse pixel, x_i is the value each coarse pixel assumes through time, N is the number of periods considered (48 months), \bar{x} is the mean value and z_i is the z -score for element i .

The mean and the standard deviation of the coarsened cube were calculated through the time dimension and the results are presented in Figure III-11.

Anomaly maps were created for each month, according to Eq. III-2, with the z -score representing the amount of discrepancy from the mean value. The results of these anomalies for each 40,000ha coarse pixel were then layered over the hydrography to create the final map representations.

Figure III-12 presents the hydrographic network for the whole area of interest in the left panel and the anomaly map produced for November 2021, in the right pane. This month represented the peak of the crisis period in the basin.

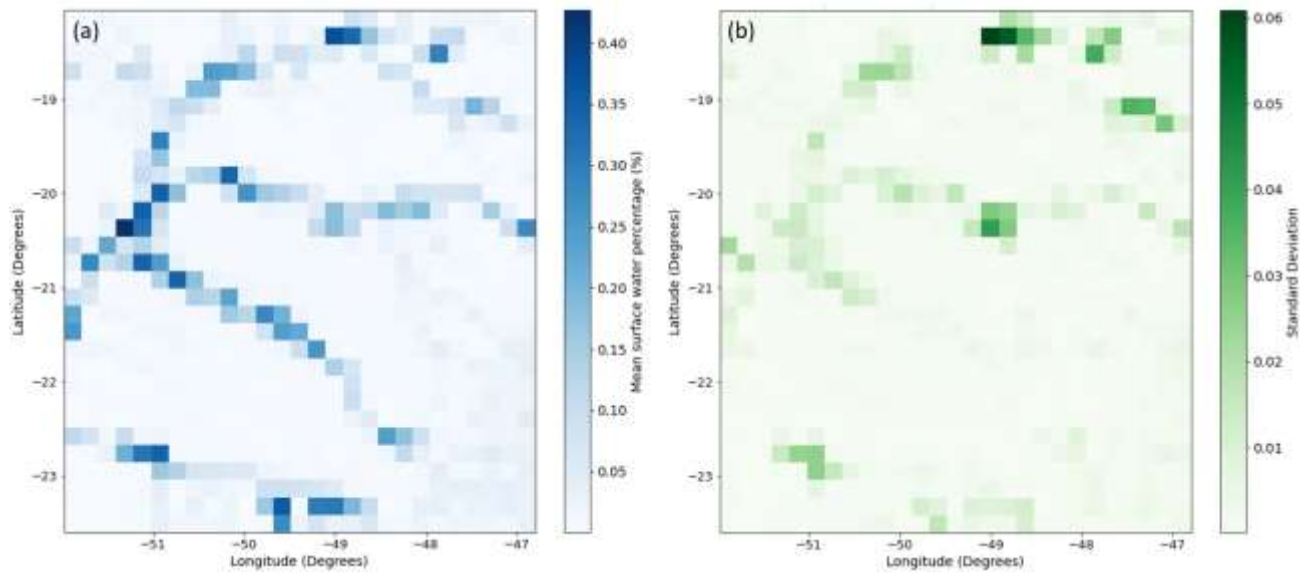


Figure III-11: (a) Mean value for each coarse pixel, calculated through time dimension, representing percentage of surface water over 40,000ha area; (b) standard deviation for each coarse pixel.

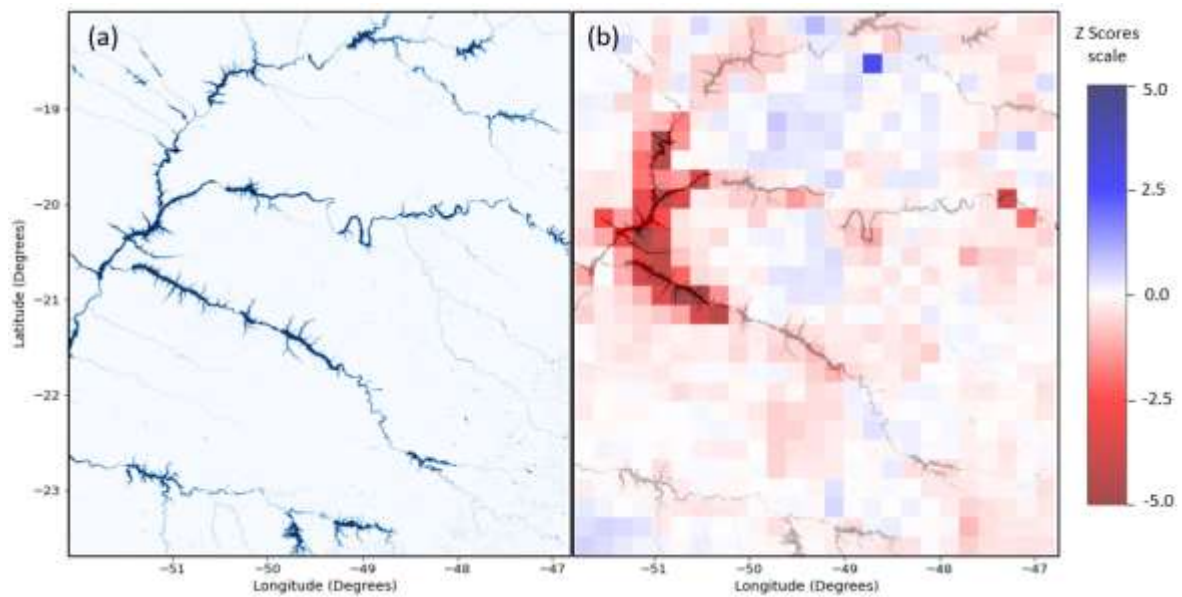


Figure III-4: (a) In the left panel, the hydrographic network of the area of interest. (b) In the right panel, the anomaly map produced for November 2021.

Stratified Analysis

One advantage of the current study over other studies in regional level is the possibility to assess what happens in smaller reservoirs that are not monitored by the water authorities. To accomplish this, it is fundamental to perform an analysis stratified by reservoirs' size. To sort the reservoirs by size, the official *shapefile* containing all mapped water bodies from Brazil was downloaded from the Sistema Nacional de Informações de Recursos Hídricos – SNIRH which is a comprehensive database of water resources provided by the National Water and Sanitation Agency (ANA, 2019a).

First, a spatial clipping was performed to the region of interest, and a total of 26,955 water bodies were identified. A 100m buffer was created around each polygon to consider any possible flood that could spread outside the official polygon. To retrieve the area of each reservoir from the raster satellite imagery, the number of water pixels within each reservoir's polygon was retrieved. This processing has been applied to each reservoir, individually, within the 48 months period. An example is presented on Figure III-13.



Figure III-5: Example of the stratified analysis for an unnamed water body. (a) Shape file with its area and coordinates; (b) high resolution imagery; and (c) overlaid water mask for June 2020.

As a result, it was possible to create a data table to keep track of each reservoir across the entire time frame. Then, the reservoirs were grouped in 3 categories, according to their surface areas: Large – area greater than 10,000ha; Medium – area between 100ha and 10,000 ha; and Small – area between 1ha and 100ha. The corresponding number of reservoirs within each group and their areas are represented in Figure III-14.

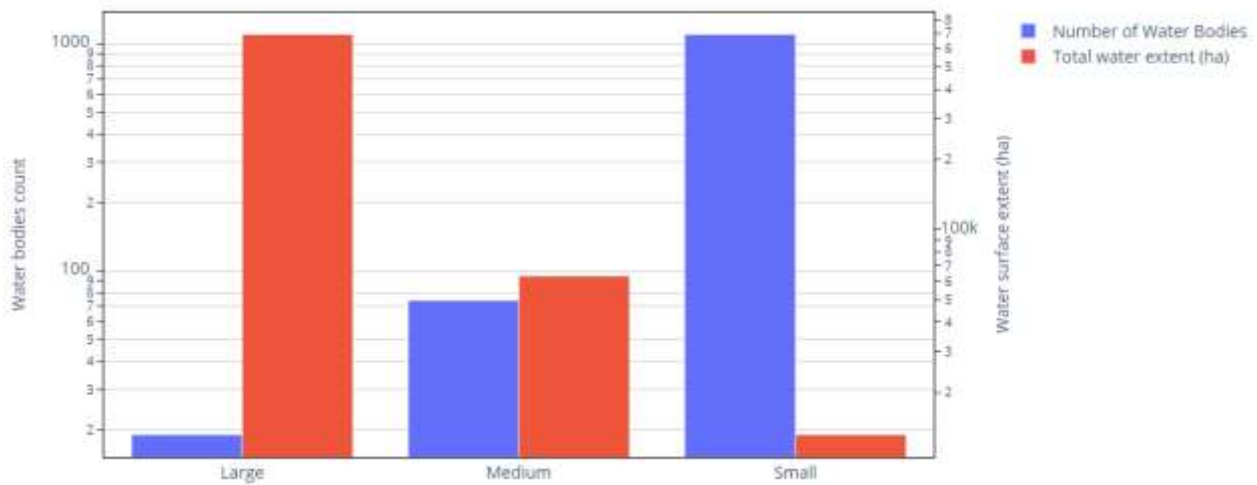


Figure III-14: Reservoirs' statistics (quantity and total area) grouped by category for the area of interest

2. Article Draft - Assessing the Impacts on Water Surface and Quality during Drought Events Using High Resolution Optical Satellite Imagery: Parana River 2021 drought (Brazil).

2.1 Abstract

The United Nations have elected water security as one of the Sustainable Development Goals for the Agenda 2030, and continuous monitoring is essential for efficient water resources management. In this context, Remote Sensing (RS) for Earth observations can be essential in complementing traditional hydrometric data, especially for water quality parameters, but most studies focus on localized scales and lack regional coverage. This study presents the application of several RS techniques toward inland water monitoring in terms of quantity and quality on a regional scale and proposes novelty in terms of data visualization. The significant drought that hit Brazil's southeast has been selected as a case study. A total of 12,000 scenes from the Sentinel-2 MSI sensor with 10m of spatial resolution have been used, covering an area of 320,000km² and 4 years of data (from 2018 to 2022) in the Parana Hydrographic Region. In the quantity analysis, the monthly water surface for the whole region has been estimated through the Waterdetect package. RS estimations were validated with field data from 16 reservoirs representing more than 7,400 km² of water surface area with an R² score of 0.96. Besides the overall area surface, a spatial analysis evidencing the water surface anomalies has been performed as well as a stratified assessment to investigate the impacts of the drought on reservoirs of different sizes. In the water quality domain, turbidity and Chl-a parameters were retrieved for the Paranapanema sub-basin and displayed as monthly Ridgeline plots. The results show a significant augmentation of eutrophication levels during the dry periods, especially during the drought when Chl-a concentration reached its peak of 16.27 mg/m³. Turbidity was substantially higher in wet periods evidencing different mechanisms

involved and a strong correlation with the rains. Moreover, the stratified analysis identified that the more extensive reservoirs have a higher seasonal variation in the water quality parameters. The results evidence an excellent potential for RS in large-scale water monitoring for both water quantity and quality, and for possible identification of trends and anomalies.

2.2 Introduction

Water availability, sustainable management of water resources, and sanitation for all, have been elected as targets by the United Nations under the 2030 Agenda for Sustainable Development Goals (UNDP, 2022b). In terms of availability, in 2020, over 2 billion people were living in areas subject to water stress, and this situation will worsen with climate changes, population growth, and land-use changes (UNDP, 2020).

In Brazil, from 2017 to 2020, drought events affected almost 90 million people, most located in the northeast region. Although the country has the world's largest amount of fresh water, almost 80% of the water discharge is in the remote Amazon River basin, away from the important economic centers. Moreover, droughts have become more frequent and severe; during the last years, they have shifted toward the southeast region. Increasing water use for agriculture and industrial activities also pressures water availability in the region (ANA, 2021a).

In 2021, the Parana Hydrographic Region faced a severe drought with a recurrence time estimated at more than 100 years (ANA, 2021a). The region is vital for the country's economics and electric power generation, representing 40% of the country's Gross Domestic Product – GDP and 45% of the total active storage of the national electric grid, called Interconnected National System (SIN) (ANA, 2018). Due to low water storage volumes, the government imposed water-use restrictions.

One aspect to consider is that most of the world's water-related issues are not driven by a lack of water but result from poor and inefficient water resources governance (UNESCO-iWSSM, 2019). Therefore, robust water monitoring and accounting are key steps toward the sustainable management of water resources and a better understanding of the surface water dynamics (UNESCO-WWAP, 2021).

In this context, since the beginning of the drought event, the Brazilian National Water and Sanitation Agency (ANA) began monitoring the daily levels of 14 large reservoirs in the region, during selected among those that contribute to the SIN. Besides, this hydrographic region has the highest stream gage density of stations operated by ANA, with 329 points. Even so, this amount is far below the estimated number of reservoirs within the region, which according to the official water bodies vectorized dataset (ANA, 2019b), has more than 20,000 water bodies with a surface area above 1ha.

To fill this gap, remote sensing (RS) for earth observations can play an important role in complementing national hydrometric data and supporting countries to monitor changes in their surface waters. In this regard, the Sentinel-2 satellite, part of the European Copernicus initiative, stands out, offering global coverage imagery free of charge. Moreover, these data have a high spatial resolution (10m to 20m) and revisit frequency of a few days, and it highly improved systematic earth observations.

To monitor inland surface water, using supervised models is preferable for accuracy. However, considering their dependency on training data, they may not be the best option at a regional scale and long time frames due to the presence of different water types, atmospheric conditions, and a diverse land coverage (Bangira et al., 2019). Based on the Landsat program, the Global Surface Water (GSW) database provides a comprehensive knowledge base about global

water extent coverage and its long-term changes over the past decades on a monthly basis (Pekel et al., 2016). However, it is not suitable for a per-scene operational analysis, and the spatial resolution of 30m can limit the detection of smaller reservoirs.

Peña-Luque et al. (2021) tested both optical and radar approaches to evaluate water surface detection at regional and reservoirs level and concluded that optical was more accurate with a lower recall error in smaller water bodies. Among the various optical algorithms tested, the better performance was achieved with agglomerative clustering issued from Waterdetect package (Cordeiro et al., 2021). It has outperformed other commonly used methods for automatic water mapping, such as Otsu Masking and Canny Edge filter, especially in smaller water bodies. Additionally, it has been shown that a windowed approach, with time windows ranging from 5 to 30 days, could improve water detection efficiency using optical images due to the persistent presence of clouds.

Similarly, a round robin intercomparison of water detection algorithms has been recently conducted in the context of the World Water project, sponsored by the European Space Agency (ESA). Several peer-reviewed algorithms using optical or Synthetic Aperture Radar (SAR) or even a combination of both were tested in 5 distinct sites. The sites covered various ecologic and climatic regions and included challenging features for water detection, such as clouds, canopy shading, and fire scars, among others. Despite the combined models (optical + SAR) having achieved the highest accuracies, this study requires optical images without the presence of clouds to infer water quality parameters. Among the optical algorithms, the Waterdetect outperformed its peers, with the advantage of avoiding using ancillary data.

In brief, the Waterdetect algorithm is an unsupervised automatic algorithm based on multidimensional agglomerative clustering. It achieves high-performance and low memory

consumption for high-resolution images by selecting a random sample of the image's pixels and then applying a Naïve Bayes classifier responsible for generalizing the results to recreate the whole scene.

In this context, this article is divided into two main axes to better understand water extent and water quality dynamics during the severe drought event that hit the southeast of Brazil at a regional level and by reservoir sizes. The first axis focuses on mapping the water surface occurrences for each month from 2018 to 2021 and evaluating the drought impact by stratifying water bodies according to their surface area. The Brazilian water body database is used as a baseline.

In the second axis, a water quality assessment considering two parameters, turbidity and Chlorophyll-a (Chl-*a*), is performed in the Paranapanema basin, a sub-basin of the region of interest with national importance. The rationale for this axis is to evaluate how the hydrological cycle and, more specifically, the 2021s drought, contribute to the variation in these water quality parameters and study the impact on smaller reservoirs that are not covered by conventional field monitoring.

2.3 Materials and Methods

Water Surface Study Area (axis 1)

The Parana is the third biggest Brazilian Hydrographic Region. It has the most significant industrial park in Brazil and settles more than 32% of the Brazilian population. With a predominantly tropical climate and a mean annual temperature of 22°C the region accounts for 30% of national consumptive water uses. However, it stores less than 7% of Brazil's water availability, evidencing potential pressures on water resources (SRHU-MMA, 2006).

The region has been subject to two significant droughts in the early 2000s and in 2014. The 2000s/2001s event was responsible for Brazil's most significant energy crisis, leading to energy-

rationing impositions due to limited interconnection among national powerplants. The 2014s event hit Brazil's largest metropolis, São Paulo, compromising its water supplies (Melo et al., 2016). More recently, in 2021, the same region faced a new drought event leading the National Water and Sanitation Agency (ANA) to raise an alert and declare water scarcity in the region on June 1st, 2021. Additionally, it established a technical follow-up group to continuously monitor and advise decision-makers about the situation in the basin (ANA, 2021b).

The water authority in Brazil has selected 14 main reservoirs within the Parana hydrographic region to be monitored continuously as they represent approximately 45% of hydropower production provided by the SIN. However, due to size restrictions, the region of interest has been defined to encompass 13 of these reservoirs, and it covers an area of approximately 320,000km² (Figure III-15).

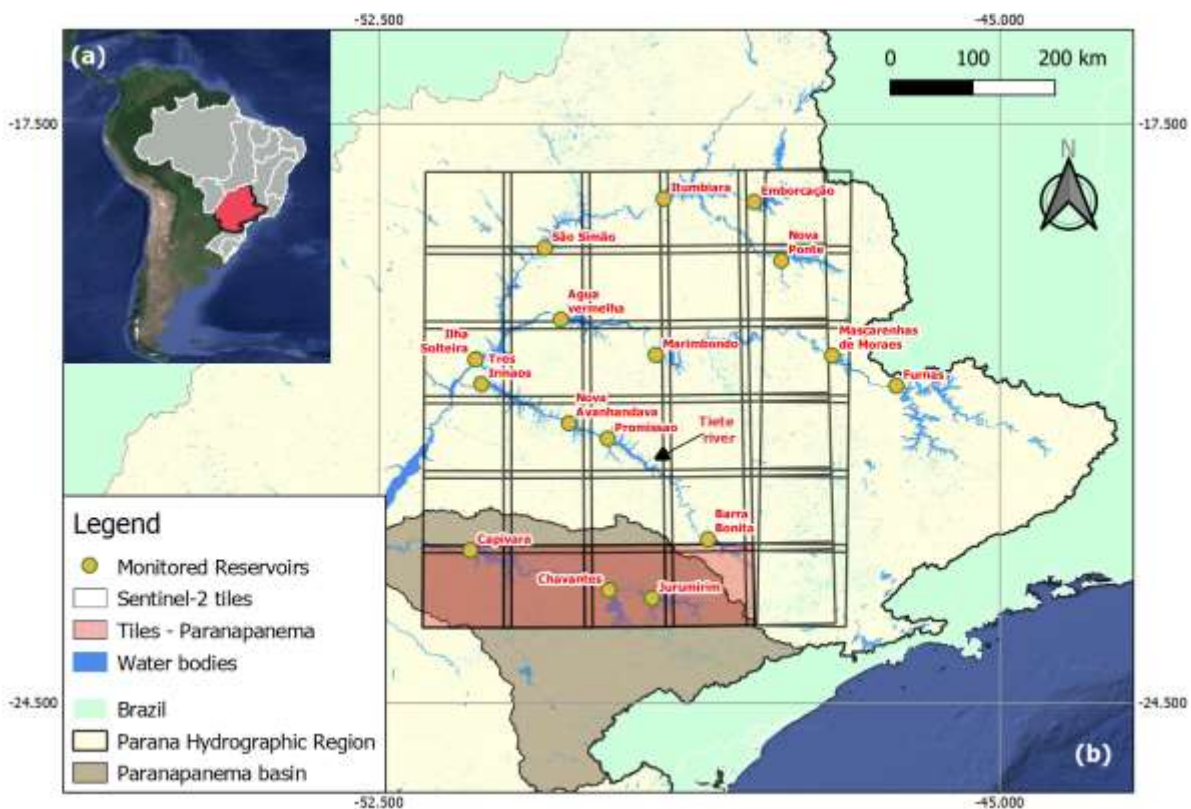


Figure III-15: (a) Parana Hydrographic Region in the national context; (b) Regions' water bodies and reservoirs being monitored by ANA. The black grid represents the region of interest selected for the current study and the Sentinel-2 tiles considered.

Water quality study area (axis 2)

For the water quality assessment, we focused on the Paranapanema River basin that covers approximately 100,000 km² with the river's main stream extending over 929 km. It flows to the Paraná River, with a mean discharge of 1500 m³/s (Rocha and de Araújo, 2011). The Paranapanema River has a total of eight hydropower dams, from which Capivara, Jurumirim and, Chavantes are the biggest ones considering its reservoirs' areas.

The area of study selected for the water quality assessment includes 4 Sentinel-2 tiles, which are highlighted in red in Figure III-15. Although it does not strictly match the Paranapanema's boundaries, 40% of the basin is considered, as well as the two hydroelectric dams of Jurumirim and Chavantes comprised in the ANA alert report. A schematic flowchart of the study, main steps, datasets, and methods used is presented in Figure III-16.

Datasets

The Sentinel-2 Multispectral Imagery satellite (Sentinel-2) was launched in June 2015 by the European Space Agency (ESA) as part of Europe's Copernicus program to provide frequent global coverage of Earth's land surface. With two satellites in orbit (Sentinel-2A and Sentinel 2B), it offers a combined revisit time of 5 days near the Equator line. Each Sentinel-2 tile covers an area of 100x100km, offering a medium spatial resolution (10m in visible and 20m in shortwave infrared – SWIR bands) with broad wavelength coverage from 443nm to 2190nm. As the objective of the current study is to assess the impact of drought in waterbodies stratified by size, this spatial resolution permits better identification of smaller scale targets when compared to other medium resolution platforms such as the Landsat program that has a pixel of 30m (Lima et al., 2019).

For the water surface extent variation analysis, a total of 36 Sentinel 2 tiles have been selected to cover all the regions of interest, as presented in Figure III-15. In addition, level 2A images atmospherically corrected by Sen2Cor (Mueller-Wilm et al., 2019) processor were used instead of MAJA (Hagolle et al., 2010), as proposed initially by Cordeiro et al. (2021) due to

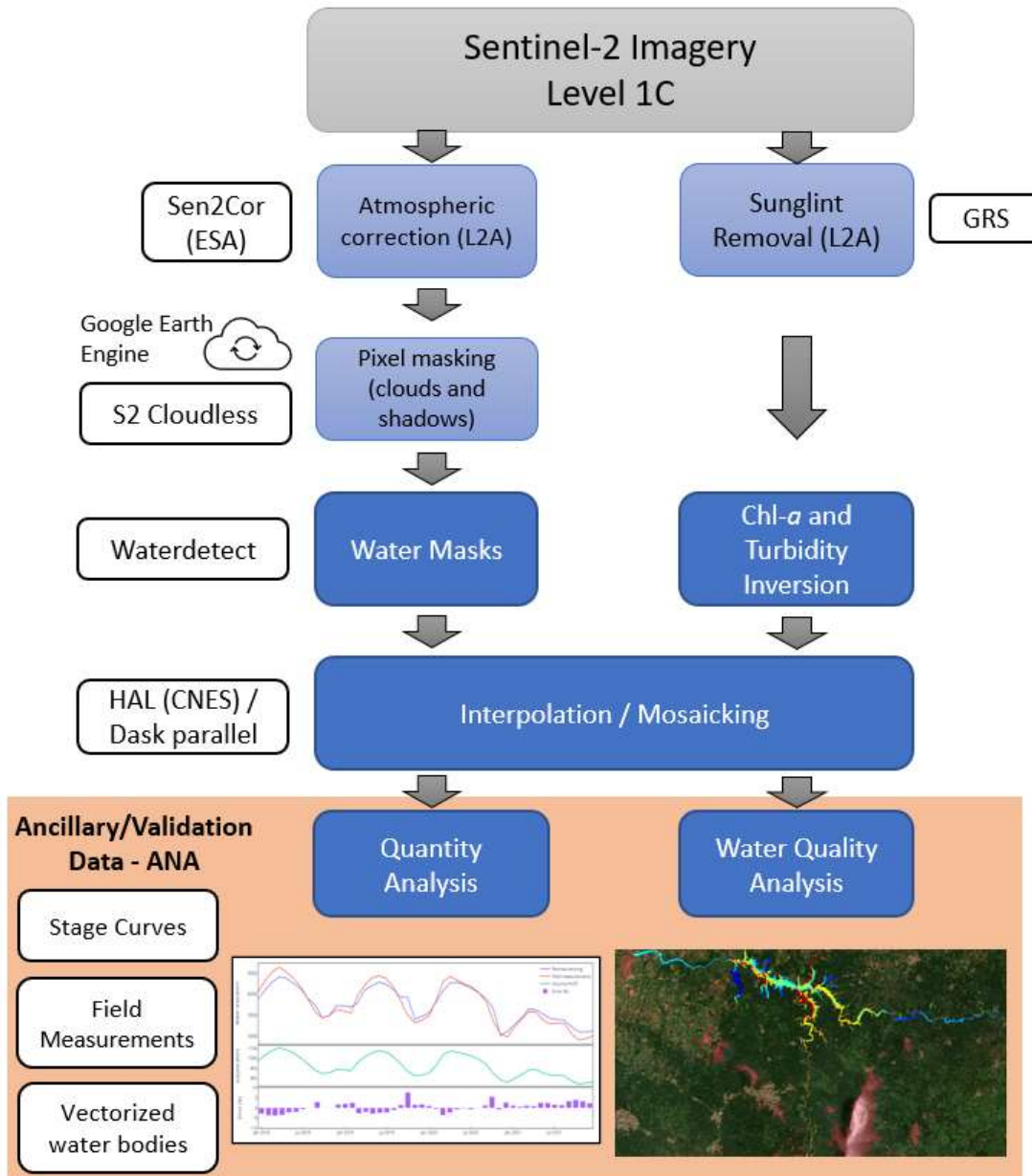


Figure III-6: Schematic flowchart of the main steps, data and methods used in the current study.

coverage availability in the selected region. A total of 12,000 scenes were processed acquired from January 2018 to December 2021. Instead of downloading the full dataset, which would take more than 25Tb of disk storage, an adaptor has been developed for Waterdetect to retrieve the single bands (Level 2A) on-demand from the Microsoft Planetary Computer cloud.

For the water quality dynamic analysis, we focused on 4 Sentinel-2 tiles covering the Paranapanema basin for which we had Level-1C images processed by the Glint Removal for Sentinel (GRS) algorithm. GRS allows removing sunglint reflection at the water surface and provides atmospheric correction and has shown to provide very robust reflectance assessment over inland and optically complex waters (Harmel et al., 2018).

It has been shown by Baetens et al. (2019) that MAJA outperforms both FMASK and specially SEN2COR in terms of cloud masks accuracy. However, the region of interest is a tropical climate zone with high cloud cover persistence and MAJA cloud masks are not available. Therefore, an alternative for cloud masking was obtained from the S2Cloudless algorithm, developed by Sinergise (Zupanc, 2020). S2Cloudless is a pixel-wise supervised Random Forest algorithm that has been trained globally with MAJA cloud masks to serve as a reference. The S2Cloudless is available as a cloud probability layer, and the dataset has been downloaded from the Google Earth Engine (GEE) Platform. The processing to derive cloud shadows is explained in the Water Detection Method section.

For the water surface analysis, besides the satellite images, the official water body vectorial dataset, provided by ANA, has also been used as a reference (<https://metadados.snirh.gov.br/>). The dataset has a total of 240,899 natural and artificial water bodies with surface areas ranging from 0.1 ha to 21,000 ha (SPR/ANA, 2020), of which 62,566 are located in the Parana hydrographic region.

To estimate rainfall in the regions, we've used data from the MERGE product from the INPE (Rozante et al., 2010). This product covers the South American continent, and is derived from the global Integrated Multi-Satellite Retrievals for GPM (IMERGE) from the National Space Agency – NASA with the addition of field data from 3,000 rain gauges in the region for model calibration (Huffman et al., 2019). The MERGE product has 10km of spatial resolution and daily data from the 1st of June 2000 to the present, the same period used to compute average rainfall.

Field Data

Additional ground data from the 18 biggest reservoirs that belong to SIN and have a surface area bigger than 10,000ha have been used to validate the water extents obtained from the satellite images. These reservoirs also include the ones monitored by ANA, as detailed in Table III-1. Since the alert has been raised, 14 of these reservoirs are subject to daily reports issued by ANA and are identified in the Table III-1 through the Alert Report column.

Historical daily stage data has been downloaded from the Reservoir's Monitoring System (SAR), available online at: <https://www.ana.gov.br/sar/>. As the historical data provides only the stage and volume percentage, the stage-area-volume curves for these reservoirs were also downloaded from the Sistema from the National Information System for Water Resources (SNIRH), available at: www.snirh.gov.br. The stage-area-volume curves are derived from bathymetric surveys and updated periodically by each reservoir manager in compliance with the technical requirements imposed by the Brazilian Electricity Regulatory Agency (ANEEL) and ANA (ANEEL and ANA, 2010).

For the turbidity validation, field data from 3 stations (Jurumirim, Taciba and Ourinhos), operated by the Companhia Estadual do Estado de São Paulo (CETESB), were downloaded from <https://sistemainfoaguas.cetesb.sp.gov.br/>.

Table III-1: Reservoirs with area greater than 10,000ha, considered for the data validation. Reservoirs without stage curve were excluded from the analysis.

Reservoir	Alert Report	Stage Curve	Reference Area (km²)
Agua Vermelha	Yes	Yes	593.53
Barra Bonita	Yes	Yes	271.77
Capivara	No	Yes	563.61
Emborcação	Yes	Yes	403.87
Ilha Solteira	Yes	Yes	1172.90
Itumbiara	Yes	Yes	714.30
Jurumirim	Yes	Yes	437.13
Porto Colombia	No	Yes	129.04
Três Irmãos	Yes	No	653.78
Promissão	Yes	Yes	537.82
Nova Ponte	Yes	Yes	405.25
Volta Grande	No	Yes	187.00
São Simão	Yes	Yes	614.43
Marimondo	Yes	Yes	365.41
Chavantes	Yes	Yes	363.16
Jupia	Yes	No	296.16
Mascarenhas	Yes	Yes	235.65
Nova Avanhandava	No	Yes	194.62

As Taciba and Ourinhos stations are located in the river stream, they are less prone to high chl-*a* concentrations. To add relevant Chl-*a* measurements, additional data from the stations Canoas I and Canoas II (station's ids 64345078 and 64345075 respectively), located inside reservoirs, were downloaded from the HIDROWEB/SNIRH website (<https://www.snirh.gov.br/hidroweb/>). The stations used in water quality validation and their corresponding geographic coordinates are listed in Table III-2.

Table III-2: Stations used in water quality parameters validation.

Station Name	Source	Turbidity	Chl-a	Latitude	Longitude
Jurumirim	CETESB	Yes	Yes	-23.2608	-49.0011
Taciba	CETESB	Yes	No	-22.6598	-51.3798
Ourinhos	CETESB	Yes	No	-22.9965	-49.9069
Canoas I – 64345078	ANA	No	Yes	-22.9122	-50.4181
Canoas II – 64345075	ANA	No	Yes	-22.9378	-50.2492

Water Detection Method

The process to generate the water masks for each Sentinel-2 scene was developed using the Waterdetect package as the main algorithm. The main workflow has been redesigned due to the amount of data to be processed. It employs multiple threads to handle concurrent tasks such as sunglint detection, band downloading, cloud masking, and cloud shadows processing (Figure III-17).

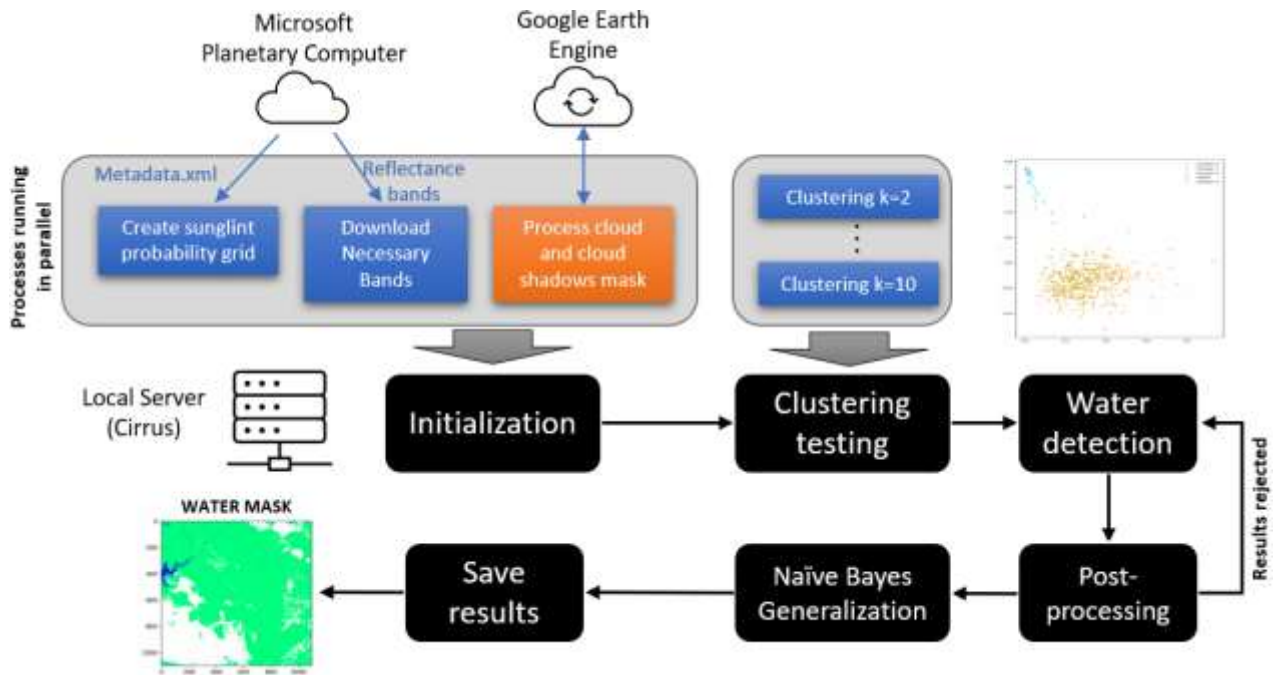


Figure III-7: Schematic diagram of the water detection process. Blue boxes represent different threads. Gray boxes group the threads that run in parallel. The orange box represents a process that runs entirely in the cloud.

The Microsoft Planetary Computer platform was used to retrieve just the necessary bands on-demand, avoiding the necessity to download complete images beforehand. The Google Earth Engine was used to process the cloud masks and project their shadows on the ground according to the sun angle in the scene, similarly to the object matching proposed in the FMask processor (Qiu et al., 2019).

Additionally, a post-processing phase has been added to the original **waterdetect** workflow to check if the water cluster output is within predefined thresholds to avoid wrong detections. If the post-processing rejects the solution, new sampling is performed, but instead of a completely random sampling, the algorithm forces a minimum amount of water pixels (defined as MNDWI > 0.4) to be included in the subset.

Surface Area Estimation

Water masks for the 36 tiles were created from January 2018 to December 2021 to estimate the surface area in the region of interest. First, these masks were combined in 1728 monthly (48 months x 36 tiles) masks, using the median operator as the combining logic. Then, each month was combined into a single mosaic, and its final resolution was downscaled to 100m of resolution, so each pixel covered an area of 1 ha. Finally, the downscaling has been performed using the average operator, so each new downscaled pixel represents the area percentage relative to 1 ha.

To fill up no data gaps that occurs mainly in the wet season months, due to high cloud persistence, the 48 monthly mosaics were stacked into a data cube. Then, a linear interpolation has been performed along the time axis to fill the remaining gaps. This whole process is presented in Figure III-18.

To provide the surface area estimation stratified by reservoir size, a subsequent processing has been performed with the aid of the shapefile containing all mapped water bodies from Brazil,

downloaded from the SNIRH. A 100m buffer was applied to each reservoir, and then the number of the water pixels inside each polygon were summed up for considered months, with each pixel representing an area of 0.01ha (10 x 10m).

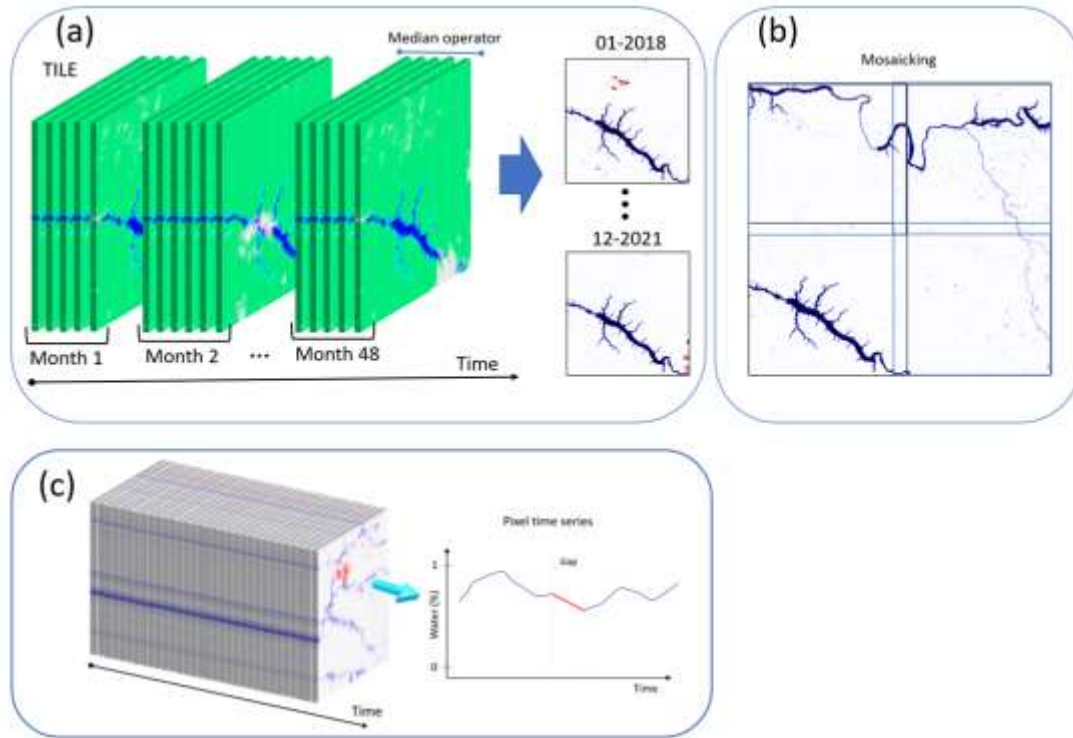


Figure III-8: Schematic view of the surface water estimation process steps. (a) Monthly water masks combination; (b) monthly masks combination into mosaics; (c) time interpolation.

Additionally, to understand the spatial dynamics of the drought period, anomaly maps were created in regions of 2,000 by 2,000 pixels (40,000 ha) that were overlaid on the hydrography for visualization purposes. The anomaly was calculated by estimating the number of deviations (z-scores) away from the mean value in each region, through to the equations:

$$z_i = \frac{x_i - \bar{x}}{\sigma}, \quad \text{Eq. III-4}$$

$$\sigma = \sqrt{\frac{\sum(x_i - \bar{x})^2}{N}}, \quad \text{Eq. III-5}$$

where σ is the standard deviation for each coarse pixel, x_i is the value each coarse pixel assumes through time, N is the number of periods considered (48 months), \bar{x} is the mean value and z_i is the z -score for element i .

Water Quality Parameters Inversion

The water quality analysis included two parameters: turbidity (T) and Chl- a . For turbidity, the single-band switching algorithm proposed by Dogliotti et al. (2015) was implemented. The algorithm makes use of the red band (645 nm) for low turbidity, defined as $R_{RS}(665) < 0.05$, which corresponds to $T \sim 15$ FNU and the NIR band (842 nm) for higher values. The single-band equation is defined as follows (Eq. III-6):

$$T_\lambda = \frac{A^\lambda R_{RS}(\lambda)}{\left(1 - \frac{R_{RS}(\lambda)}{C^\lambda}\right)}, \quad \text{Eq. III-6}$$

where A^λ and C^λ are wavelength-dependent coefficients, with values 228.1 (A^{red}) and 0.1641 (C^{red}) for the red band and values 3078.9 (A^{NIR}) and 0.2112 (C^{NIR}) for the NIR band.

A transition zone between the algorithms is proposed for $R_{RS}(665)$ ranging from 0.05 to 0.07. In this zone, the weight of the NIR algorithm (w) changes linearly from 0 at $R_{RS}(665) = 0.05$ to 1 at $R_{RS}(665) = 0.07$, and the blending is done according to

$$T = (1 - w)T_{red} + wT_{NIR} \quad \text{Eq. III-7}$$

For the Chl- a parameter, several retrieval algorithms selected by Neil et al. (2019) have been tested and validated using field measurements. In the end, the model that proved to be the

most adherent with the region of study and Chl-*a* values range of the present study is a simple ratio between the red-edge (704 nm) and the red (665 nm) bands. This model (Eq. III-8) has been originally proposed by Gitelson and Kondratiev.

$$Chla = a \left(\frac{R_{RS}(704)}{R_{RS}(665)} \right) + b, \quad \text{Eq. III-8}$$

where $a = 61.324$ and $b = -37.94$ are determined empirically.

Metrics

The metrics used for evaluating the proposed methodologies are the standard Coefficient of Determination (R^2), relative error, Root Mean Squared Error (RMSE) and the Percentage Bias (PBIAS) that measures the average tendency of the simulated data to be larger or smaller than their observed counterparts:

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(y_i - \bar{y})^2}, \quad \text{Eq. III-9}$$

$$\text{relative error} = \frac{\hat{y}_i - y_i}{y_i} \quad \text{Eq. III-10}$$

$$RMSE = \sqrt{\sum \frac{(\hat{y}_i - y_i)^2}{n}}, \quad \text{Eq. III-11}$$

$$PBIAS = 100 * \left(\frac{\sum(\hat{y}_i - y_i)}{\sum y_i} \right), \quad \text{Eq. III-12}$$

where y_i is the true observed measurement, \hat{y}_i is the value estimated by the model, \bar{y} is the mean of the observed measurements, and n is the sample size.

2.4 Results

Water Surface Analysis (axis 1)

The proposed methodology has been applied to assess the 2021 Brazilian drought by analyzing the water surface variability as a function of time for a large number of water bodies, most of them not being monitored by local authorities. Initially, data validation was performed to compare the results obtained from remote sensing with the areas benefiting from field monitoring, which correspond to the largest reservoirs (Table III-1). Then, three analyses were proposed. First, an overall analysis, considering all the water bodies within the area of interest to assess how the basin behaved during the stress period. Then, a second analysis in which the water bodies were grouped by area was performed to assess how the drought has impacted reservoirs of different sizes. In the end, anomaly maps were produced to assess how the drought was distributed spatially in the region.

Validation of remote sensing derived water surface

The proposed methodology for surface water extraction has been validated with field measurements provided by ANA from 16 reservoirs (Table III-1). Surface areas sum up to 8,144 km² and represent approximately 70% of the estimated water area for the entire region (11,888 km²). Unfortunately, two reservoirs don't have stage curves (Três Irmãos and Jupia) and were excluded from the validation.

The field areas for each reservoir were obtained from the historical stage data and the stage curves. The monthly average was computed to convert from daily data. The results of the field-measured areas and comparison with remote sensed obtained values are shown in

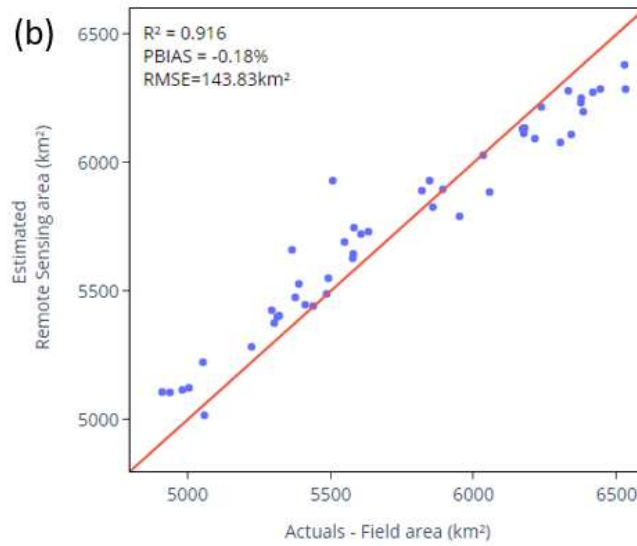
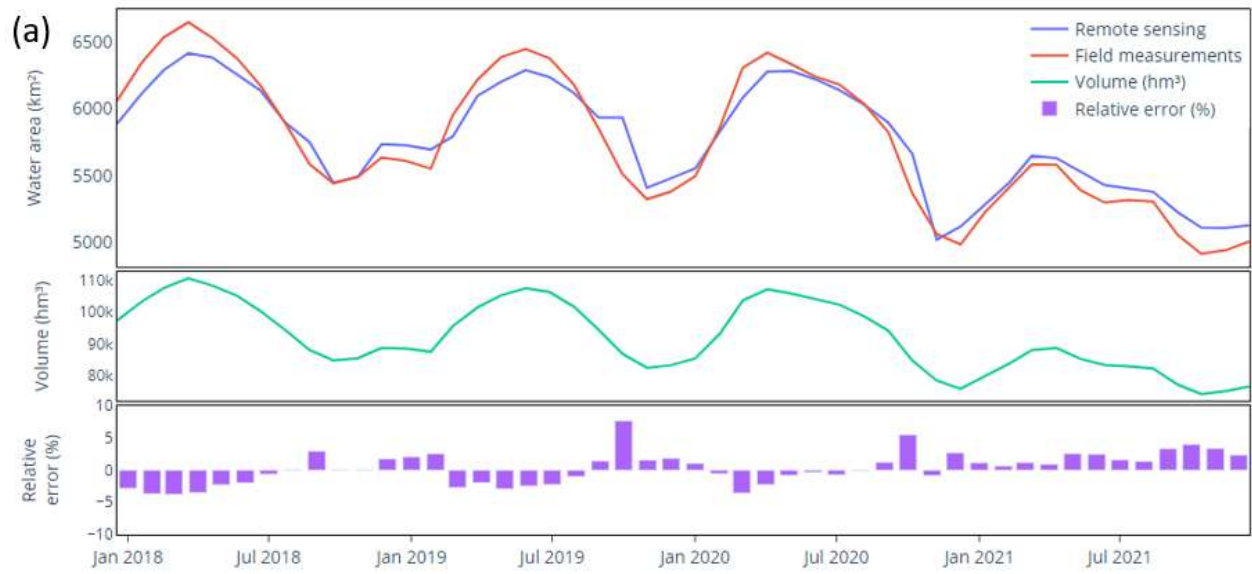


Figure III-19a.

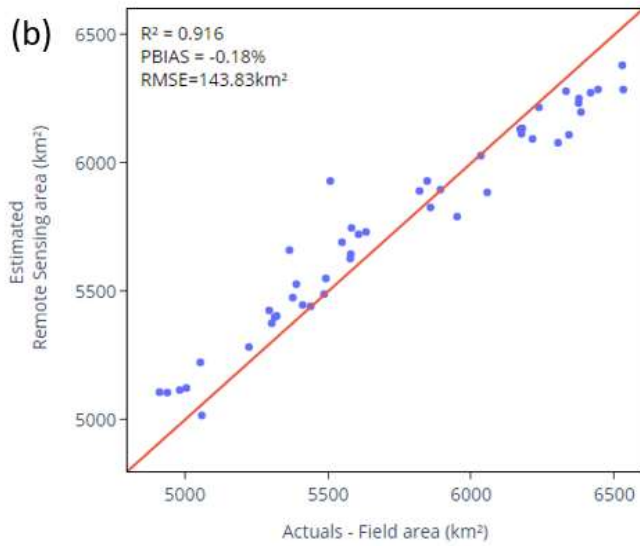
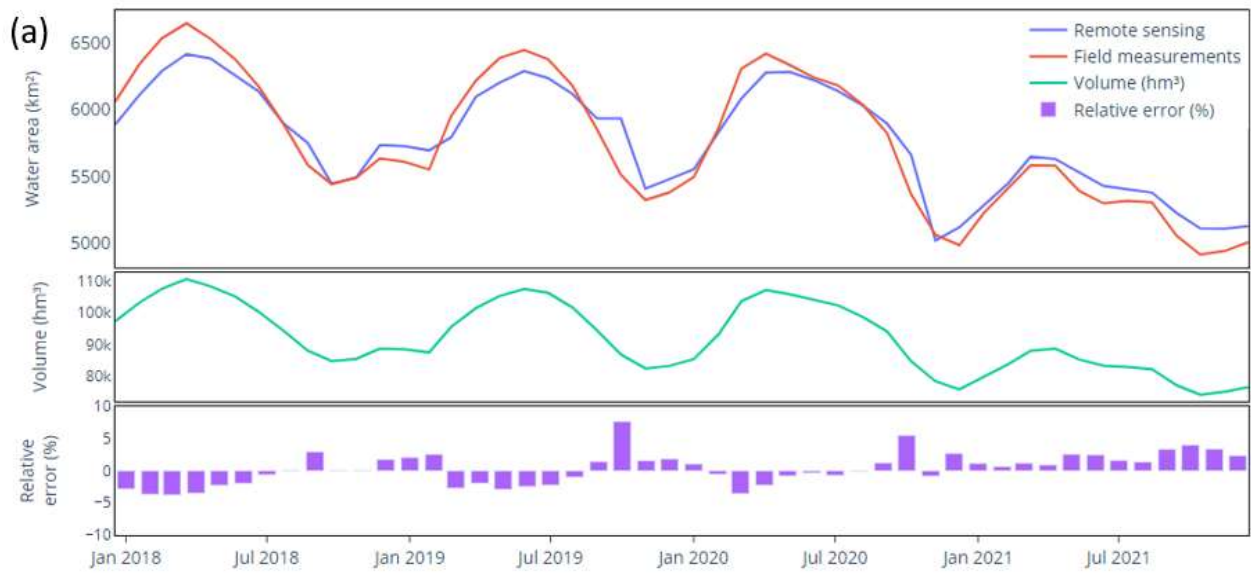


Figure III-9: (a) Monthly water surface area time-series obtained from field measurements (red line) and from the satellite imagery (blue line) for the 16 selected reservoirs, total volume as measured in the field (green line) and errors between field and satellite measurements (purple bars); (b) Correlation between actual field measured area and estimated remote sensing area.

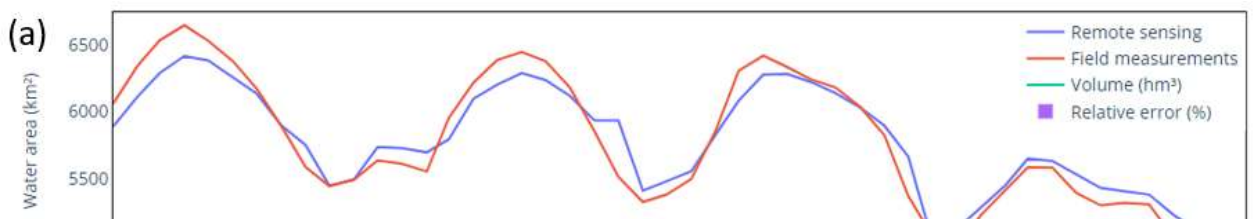


Figure III-19b presents the correlation between actual field area versus predicted remote sensing area. The graph shows an excellent overall adherence with a coefficient of determination $R^2 = 0.92$, and a maximum relative error of 7.6% occurred in October 2019 (

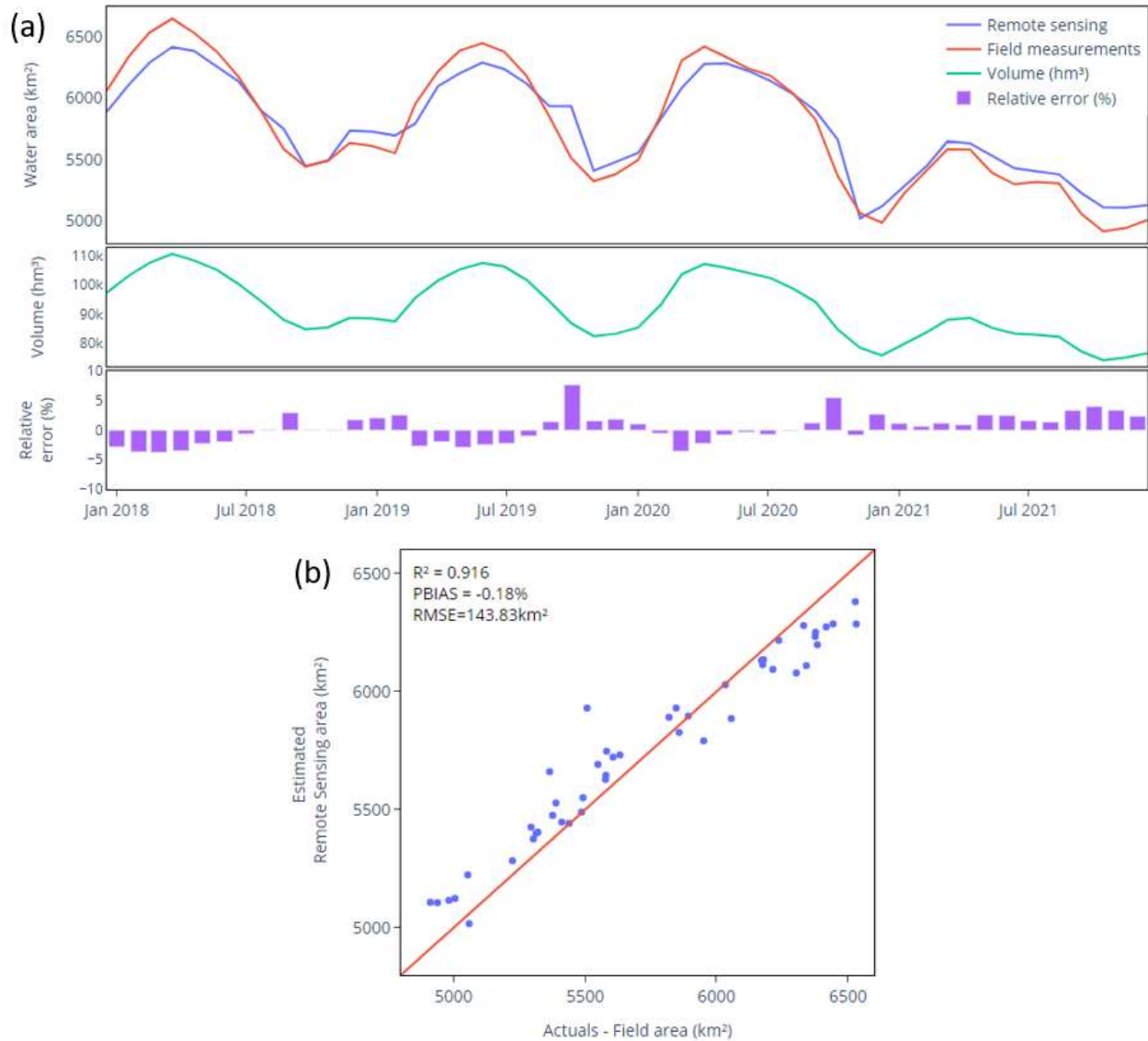


Figure III-19a, error bars). To check for possible bias, PBIAS was evaluated and resulted in -0.18%, indicating an almost unbiased result. However, a visual inspection of

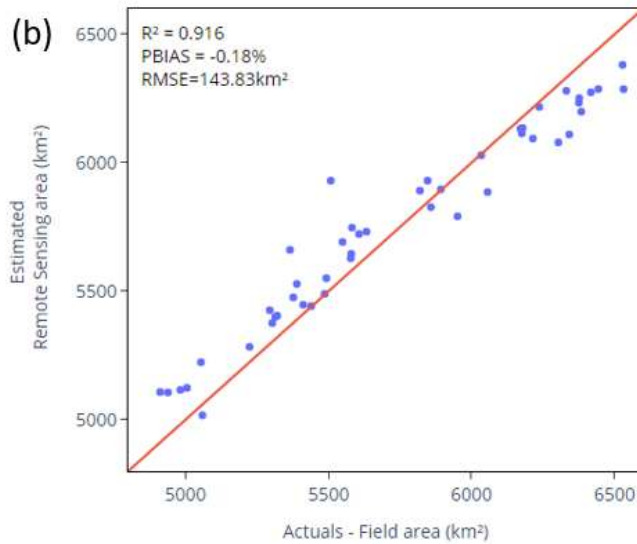
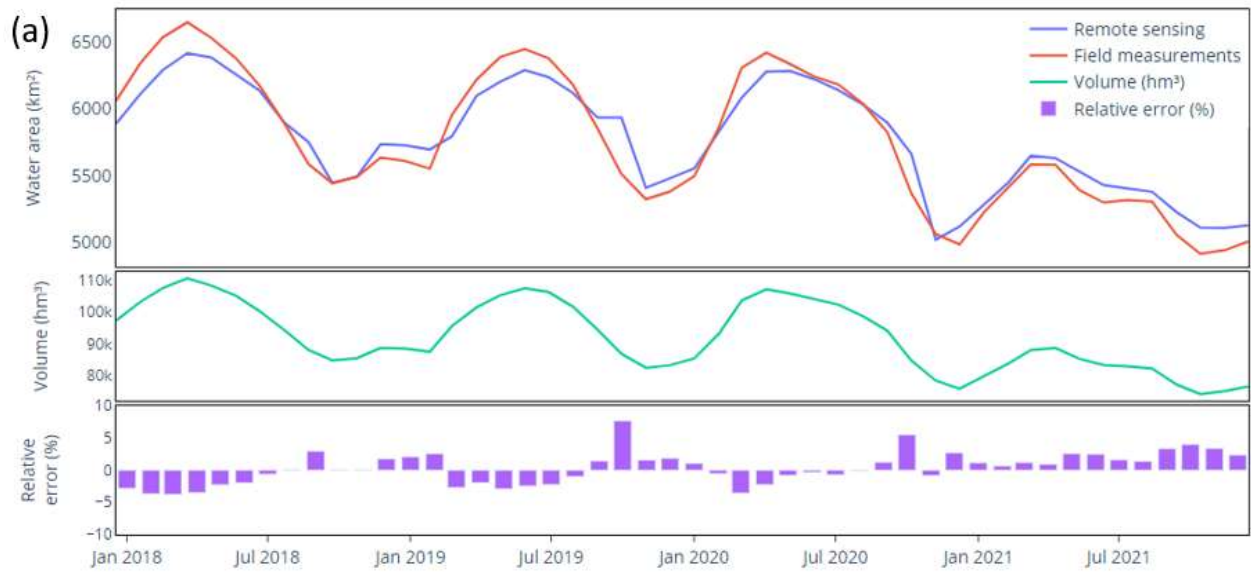


Figure III-19 shows that most of the underestimation occurs in the wet season. This generalized underestimation of the area extent identified for high filling periods might be explained by the dense vegetation often covering the rarely flooded shores and upstream parts of the reservoirs, as noted by Peña-Luque et al. (2021). Other minor differences may be explained by the amount of no data in the considered month, mainly due to clouds or other uncertainties in the water detection process due to confusing pixels such as shadows, dark vegetation, and others (Cordeiro et al., 2021).

Overall water extent

The overall analysis assessed how the total water extent decreased in the 2021s drought and compared it to the previous years. All water pixels in the region of interest, in full 10m resolution, have been included, independent of whether they represent rivers, reservoirs, or smaller ponds. The time series is presented in Figure III-20. The red dot indicates the date when ANA declared the critical drought situation in the basin, on June 1, 2021. It can be noted that the high filling rates usually occur between April and June, corresponding to the end of the wet season, which goes from September to May. The lower accumulation for 2020 occurred in November and represented approximately 6,3% less surface area than the lower accumulation months of the two preceding years (2018 and 2019), which also happened to be in November.

Moreover, the maximum surface area that typically surpasses 12,000km² was only 11,478 km² (7% lower) on Mars 2021 due to the scarcity of rains in the region during the 2020/21 rainy season. According to a Climate Situation Report produced by the National Institute for Spatial Research (INPE) issued on May 27, 2021, the accumulated monthly rains were below average from February 2021 to April 2021, and the forecast pointed to rains lower than the mean up to September 2021. At the end of 2021, the surface reached its minimum area of 10,200km² in October, a value slightly lower than the minimum that occurred the previous year (10,350km²).

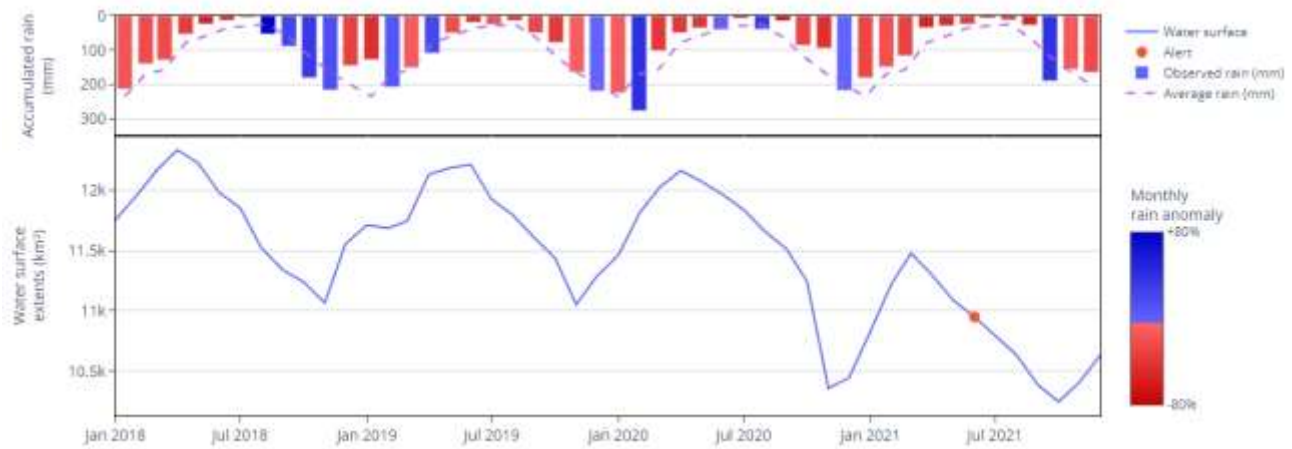
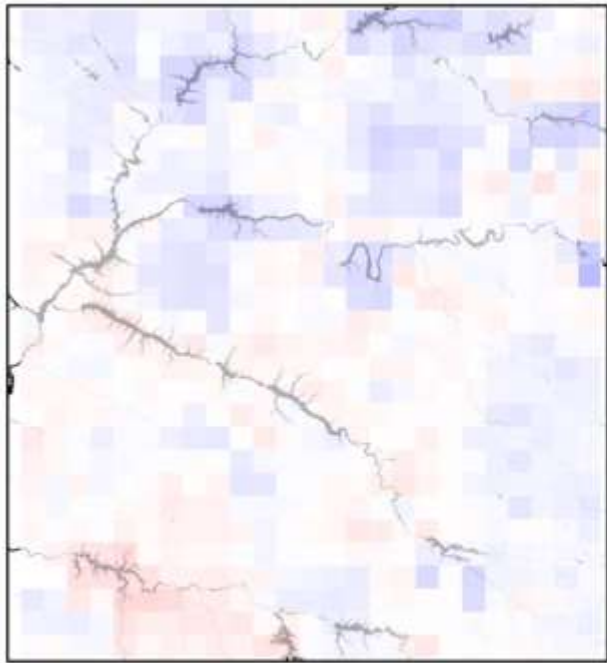


Figure III-20: Overall water surface extents (lower panel) in the study area. Red dot indicates when ANA raised the drought alert in the basin. In the upper panel, red bars indicate rainfall below average and blue bars indicate rainfall above average, according to MERGE product. The dashed line represents the average rainfall computed between 2000-2019.

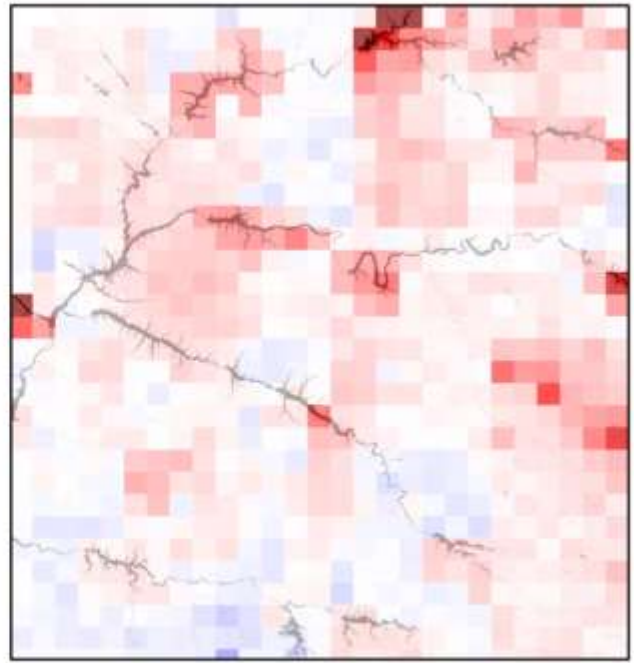
Assessment of the spatial pattern of the drought

The spatial assessment has been conducted to analyze how the basin has been affected spatially by the 2021's drought. The years 2018 and 2019 were considered as references for the computation of average area and standard deviation in order to calculate the z-scores for 2020 and 2021. Figure III-21 shows the distribution of the water surface anomalies, expressed in terms of z-score statistics (Eq. III-2) for the whole region of interest. The water surface anomalies for the period of maximum water storage (April 2020) (Figure III-21a), corresponding to the final stage of the rainy season, showed that the northern portion of the region of interest has a water surface above average, as expected for the period. However, the southern part is slightly below average, notably the Capivara reservoir and the Tietê river (see Figure III-15).

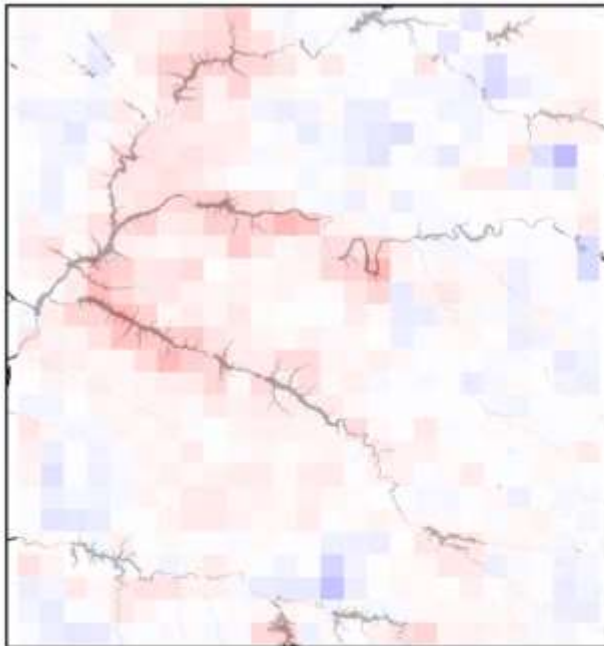
Comparing the dry period of November 2020 to the water storage capacity peak in April 2020 (Figure III-21b), we can note a substantial drop in surface area in all locations but the southwest. The most significant drops occurred in the northern areas, especially in the reservoirs Itumbiara (highest area reduction), São Simão, Água Vermelha, and Marimbondo (Figure III-15).



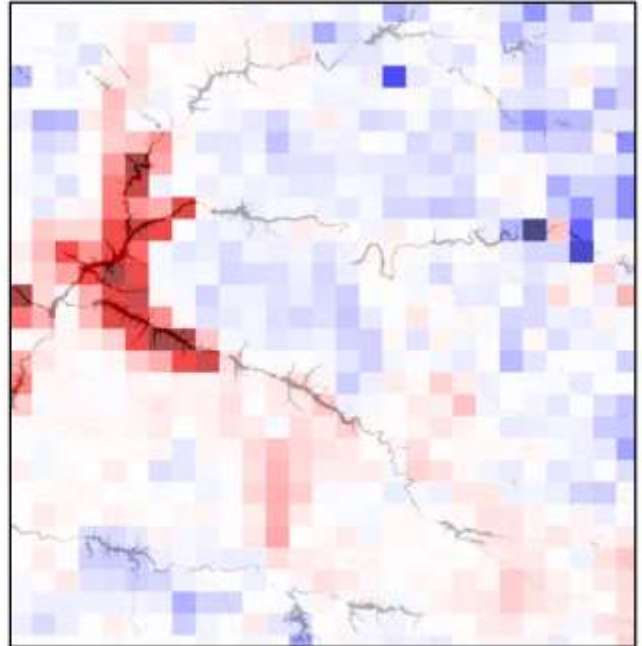
(a) April's 2020 anomaly to the mean



(b) November 2020 vs April 2020



(c) April's 2021 anomaly to the mean



(d) November 2021 vs November 2020

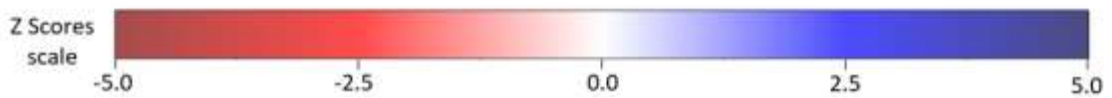


Figure III-10: Water surface anomaly maps, represented in percentage of change. (a) April's 2020 anomaly compared to the mean; (b) April's 2021 anomaly compared to the mean of 2018 and 2019; (c) November 2020 compared to April 2020; (d) November 2021 compared to November 2020.

Moving forward the analysis to April 2021, which marked the beginning of the dry season and the period when the reservoirs were expected to be fully recovered, it can be seen from Figure III-21c that none of the biggest reservoirs in the region are above average. The situation worsened until November 2021, when the overall water storage reached its minimum for the whole period since 2018. Figure III-21d compares the drought of 2021 to the drought of the previous year (i.e., 2020). Most reservoirs remained stable, with percentage variations smaller than 2%. However, a critical depletion can be noted in the reservoirs of Ilha Solteira, located in the Parana river, and Três Irmãos, located in the Tietê river.

Impact of the drought as a function of water body size

In the stratified analysis, the water bodies were categorized into three groups according to their surface areas: **Large** – area greater than 10,000ha; **Medium** – the area between 100ha and 10,000 ha; and **Small** – area between 1ha and 100ha. As the order of greatness of the water surface area is different for the three categories, the results were plotted in a graph with the anomaly measured in percentage referenced to the maximum water extent in the period.

Figure III-22 shows there exists similar seasonal trend between small and large water bodies, with both groups presenting the same water depletion pattern, especially during the 2021s drought, but with the smaller water body class showing greater losses (-30% in October relatively to the maximum extent over the 2018/2021 period) than the larger body classes (-25% in November

2021). The small water bodies present rapid variations, especially during the dry period while the medium size water bodies exhibit almost no variations.



Figure III-12: Water surface variation, measured in percentage, in comparison to the maximum extents for the water bodies grouped by size.

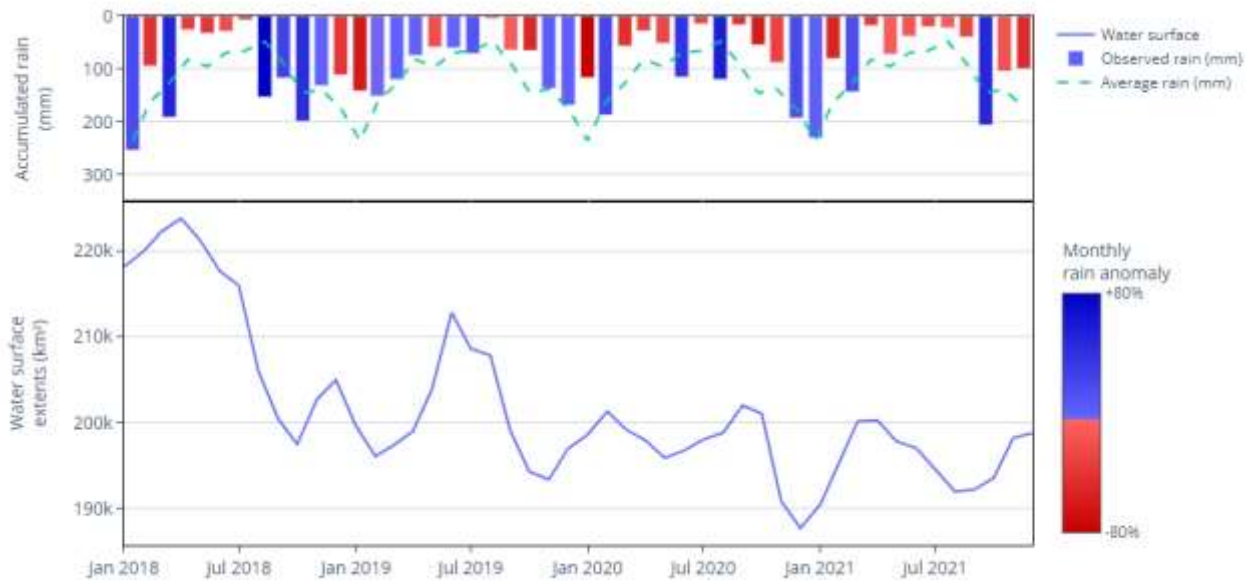


Figure III-11: Overall water surface extents (lower panel) in the Paranapanema basin. In the upper panel, red bars indicate rainfall below average and blue bars indicate rainfall above average, according to MERGE product. The dashed line represents the average rainfall computed between 2000-2019.

Analysis of the Water Quality trend during the drought period (axis 2)

Parapanema Water Surface Extents

To better understand the behavior of water quality parameters in the Parapanema basin, it is essential to note that its surface extents have a unique behavior, differing from the overall behavior of the region of interest, as depicted in Figure III-23. It can be seen from the graph that the basin presents a descending trend from January 2018 onwards, with the lower level occurring in November 2020, before the 2021 drought event. This behavior suggests a more complex pattern involving interannual variability in the local water resource availability and management.

Water Quality Data Validation

Turbidity retrievals have been validated with field measurements from Jurumirim, Taciba, and Ourinhos stations, with values up to 100 FNU. It can be noted from the graphs (Figure III-24) a positive BIAS with the algorithm delivering values slightly higher than those measured, especially in Jurumirim and Taciba stations, during 2016.

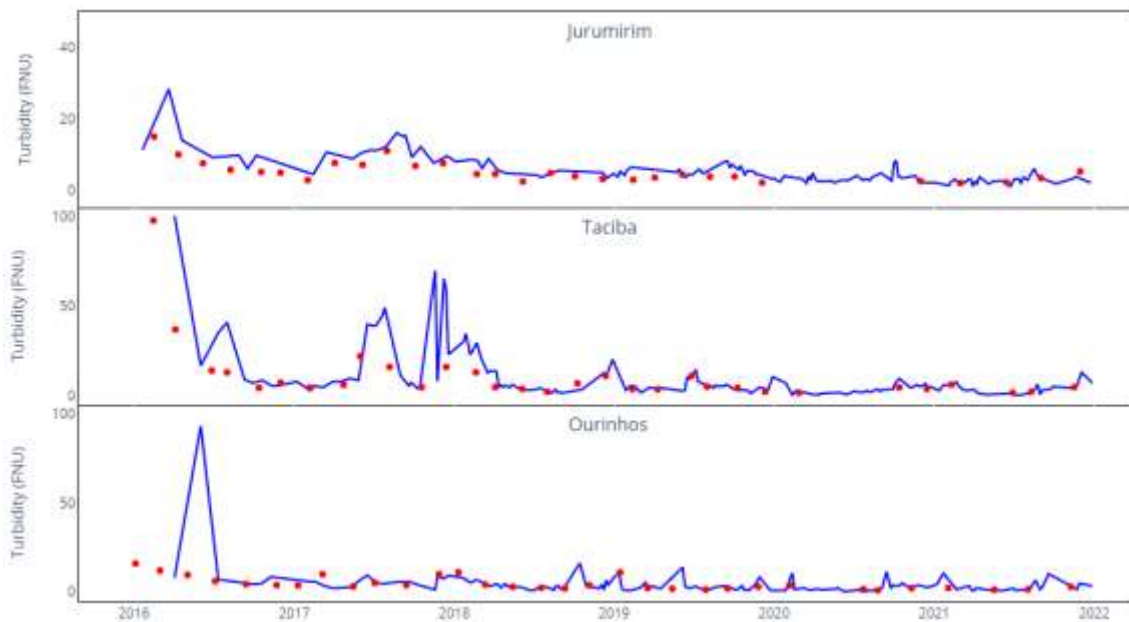


Figure III-13: Comparison between remote sensing derived turbidity (blue lines) and field measurements (red dots) for Jurumirim, Taciba and Ourinhos stations in the Parapanema basin.

Some spikes in all three satellite retrieved series can be seen, but it is difficult to confirm or discard them due to the absence of field measurements in matching periods. The actual versus predicted scatter plot has been prepared by matching field measurements with the closest available image and limiting the maximum date difference to 10 days. The graph (Figure III-25a) shows a good correlation, with RMSE = 10.6 FNU and a RMSLE of 0.46 FNU.

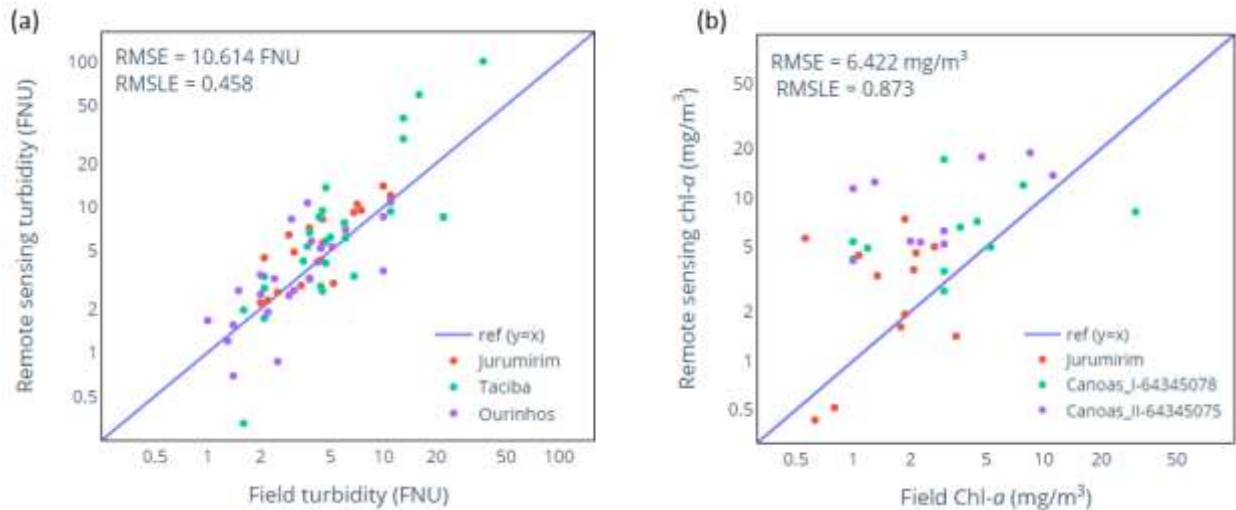


Figure III-14: (a) Comparison between field turbidity (x axis) and retrieved remote sensing turbidity (y axis), for Jurumirim, Taciba and Ourinhos stations; (a) Comparison between field Chl-a (x axis) and retrieved remote sensing Chl-a (y axis), for Jurumirim, Canoas I and Canoas II stations. Mashup performed considering a maximum data lag of 10 days.

A similar validation has been performed for Chl-a but replacing Taciba and Ourinhos river stations with Canoas I and Canoas II (ANA's station ids: 64345078 and 64345075 respectively) reservoirs stations, as presented in Figure III-26. Although these stations are located inside medium-sized reservoirs, field measurements are mainly in the low range, with a maximum value of 30.4 mg/m³ in Canoas I, which occurred on March 20.

The actual versus predicted scatter plot for Chl-*a* (Figure III-25b) shows a more significant dispersion when compared to turbidity with a RMSE of 6.4 mg/m³ and a RMSLE of 0.873m³.

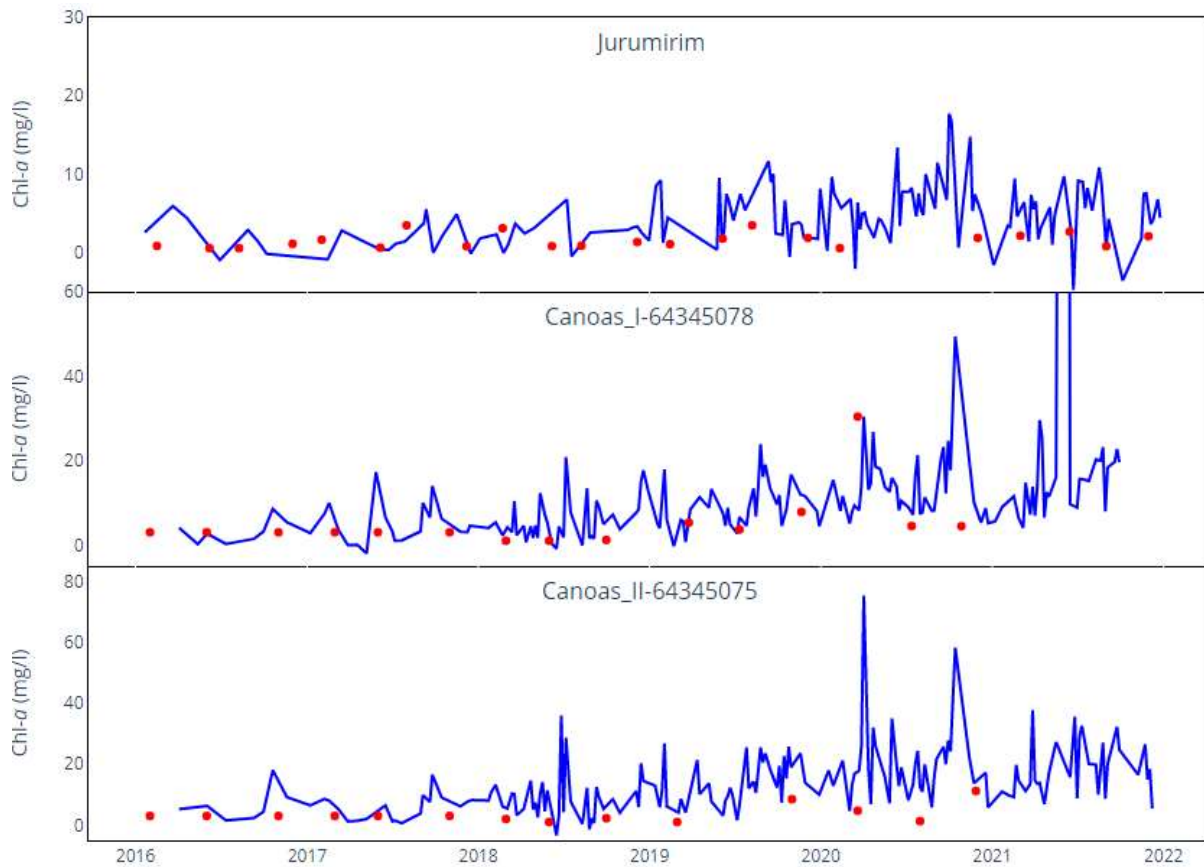


Figure III-15: Comparison between remote sensing derived Chl-*a* (blue lines) and field measurements (red dots) for Jurumirim, Canoas I and Canoas II stations in the Paranapanema basin

Water Quality Maps

To analyze the effects of the 2021s drought on the water quality and understand its spatial dynamics within the basin, monthly Chl-*a* and turbidity maps have been elaborated at 20-meter resolution. These maps represent the median value obtained from all valid pixels in the specified month, discarding invalid pixels such as clouds, shadows, negative reflectance values, and other anomalies. Considering April 2018 has the higher surface area and November 2020 the lowest, these two dates are plotted as examples in Figure III-27 and Figure III-28 for Chl-*a* and turbidity, respectively.

In Figure III-27, we can note an overall higher level of eutrophication in the November 2020 image on all water bodies within the region of interest. Higher water residence time (WRT) can lead to higher nutrients accumulation and stratification, resulting in excess phytoplankton growth (Olsson et al., 2022). Therefore, that observation matches the expected behavior, as November 2020 is the end of the dry season, representing the lowest water storage and inflow in the basin, which diminishes WRT.

On the other hand, turbidity maps (Figure III-28) doesn't show the same behavior, suggesting that turbidity and Chl-*a* are driven by different factors.

Water Quality Histograms

To go beyond a simple visualization of the water quality parameters for individual dates, we produced monthly Chl-*a* and turbidity histograms elaborated using the pixel values within all the water bodies with surface size greater than 1 ha identified in the Paranapanema region of interest. A total of 48 monthly Chl-*a* and turbidity histograms have been combined into Ridges' graphs for easy visualization. Due to the range of Chl-*a* values (from 0 up to 1,000 mg/m³), the histograms' bins were defined using logarithmic scale, from 10⁰ to 10³ in 0.1 steps in the exponent.

Figure III-29 presents the Chl-*a* histograms. The colors and annotated values represent the median Chl-*a* value for each month, ranging from 4.18 mg/m³ in Mars 2019 to 17.63 mg/m³ in September 2021. There is an overall uptrend in eutrophication levels through these years, resulting in higher Chl-*a* values, calculated as the mean from the monthly medians, from 7.04 mg/m³ (2018) to 12.81 mg/m³ (2021), which represents an increase of 82%.

Chl-a concentrations maps

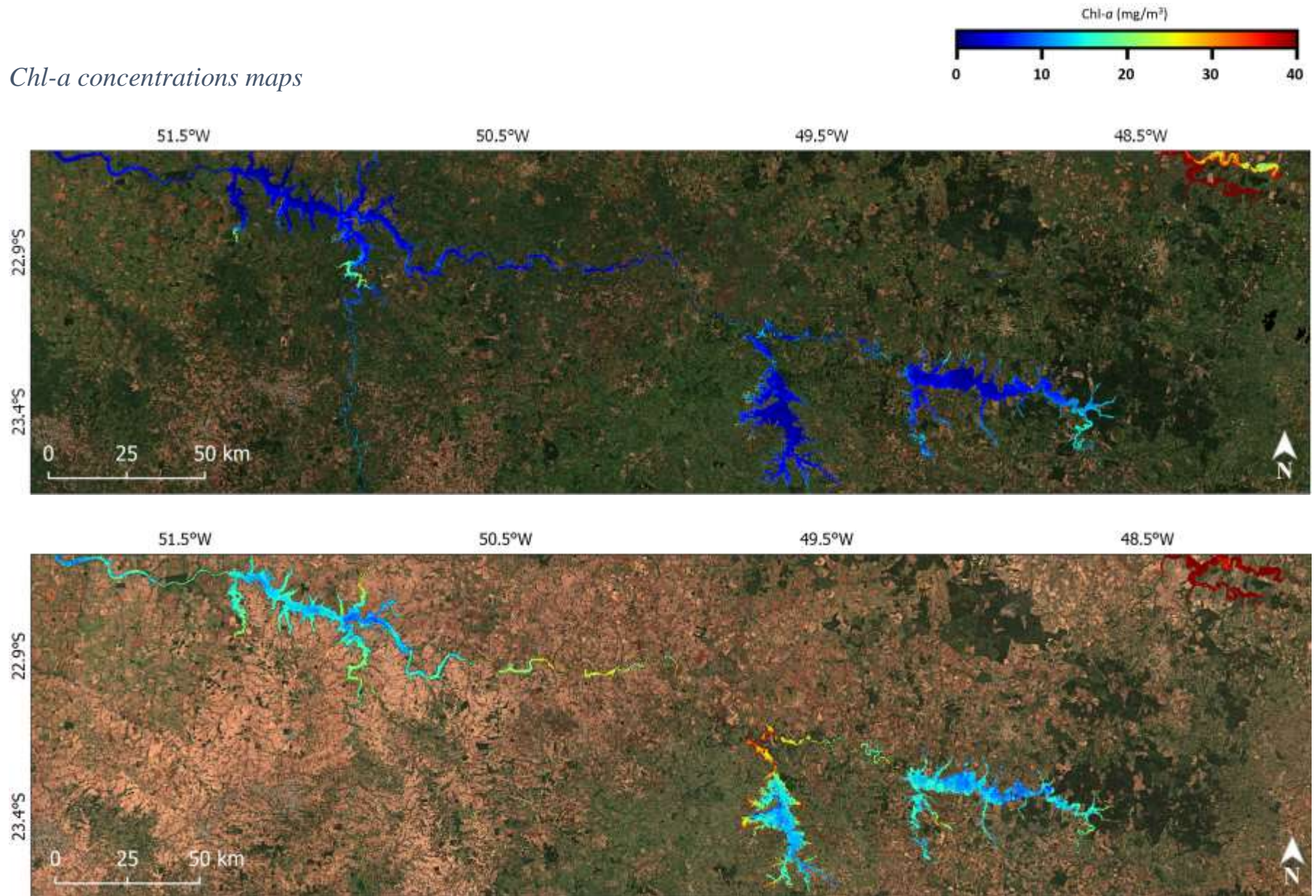


Figure III-16: Chl-a concentration maps for the Paranapanema region. (a) monthly composite for 04/2018; and (b) monthly composite for 11/2020

Turbidity concentrations maps

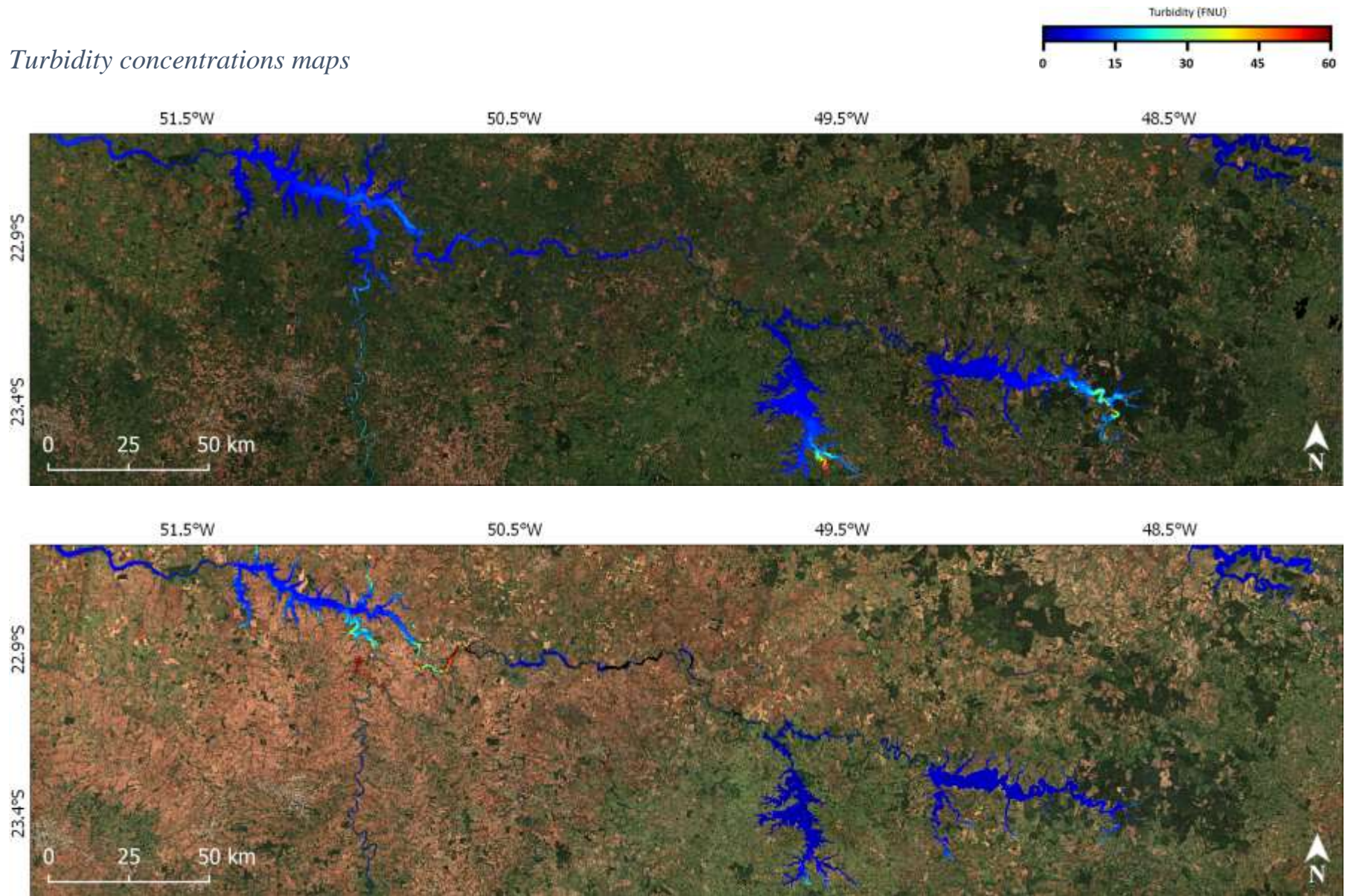


Figure III-17: Turbidity concentration maps for the Paranapanema region. (a) monthly composite for 04/2018; and (b) monthly composite for 11/2020

When analyzing each year separately, eutrophication levels remained stable throughout 2018 and did not present a visible pattern. However, in 2019, a slight increase in Chl-*a* is noticeable in the dry season, from June to November. In 2020, when the basin reached the lowest accumulation levels, the mean Chl-*a* increased, especially in September and October, achieving the highest median Chl-*a* in October (16.73 mg/m³). The last quarter of 2020 is also marked by a significant drop in water surface (from 202,000 km² in September 2020 to 187,000km² in December 2020), as seen in **Error! Reference source not found.** In 2021, Chl-*a* levels remained higher than the mean during the whole year and eutrophication reached its peaks from June to October with median Chl-*a* consistently above 15 mg/m³ and a maximum value of 17.63 mg/m³ in September 2021. These values indicate that most of the water surface would be rated as mesotrophic or eutrophic according to the classification schema proposed by Carlson and Simpson (1996).

Additionally, the histograms become bimodal with second peak of occurrences around 70 – 100 mg/m³ in several months (e.g., June, July, August, and September 2018; September 2019; among others).

Turbidity histograms are presented in Figure III-30. Similar to Chl-*a*, colors and annotated values represent the median turbidity retrieved in FNU for each month, ranging from 2.18 FNU, which occurred in May 2021, to 13.92 FNU In January 2018. Contrary to the Chl-*a* behavior, turbidity presented an overall downtrend in the period. Yearly mean turbidity monotonically decreased from 7.34 FNU in 2018 to 3.42 FNU in 2021. The behavior within each year is also opposite to the one seen in Chl-*a*, with higher turbidity on rainy months (October to Mars) and lower turbidity in the dry season (April to September). Higher turbidity values could be observed in January, and February 2018 (13.92 and 12.27 FNU, respectively), and the lowest values occurred in May, June, and July 2021 (2.18, 2.20, and 2.19 FNU, respectively).

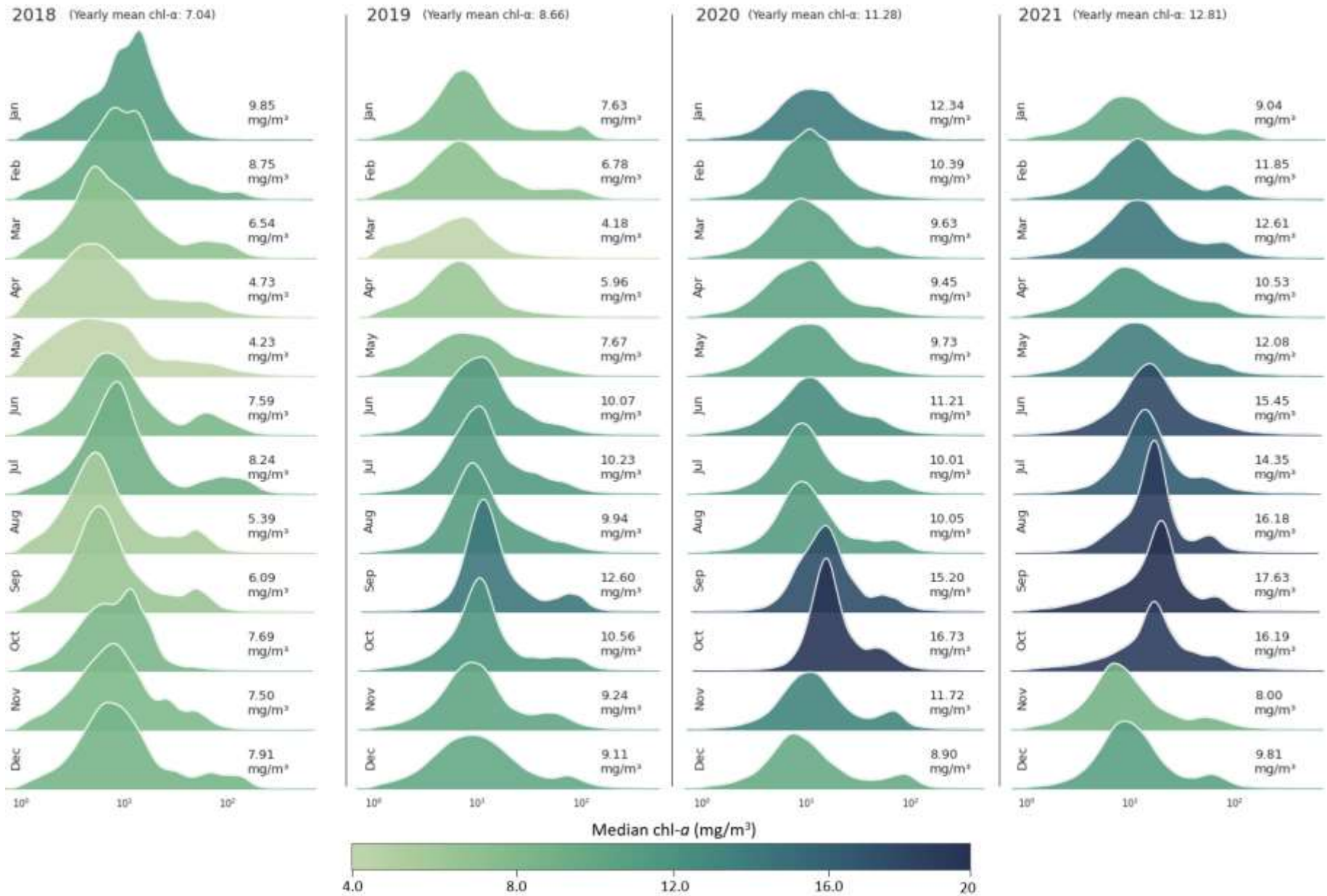


Figure III-18: Chl-a histograms for the Paranapanema region of interest. Annotated values and color scale indicate the median Chl-a value in the month.

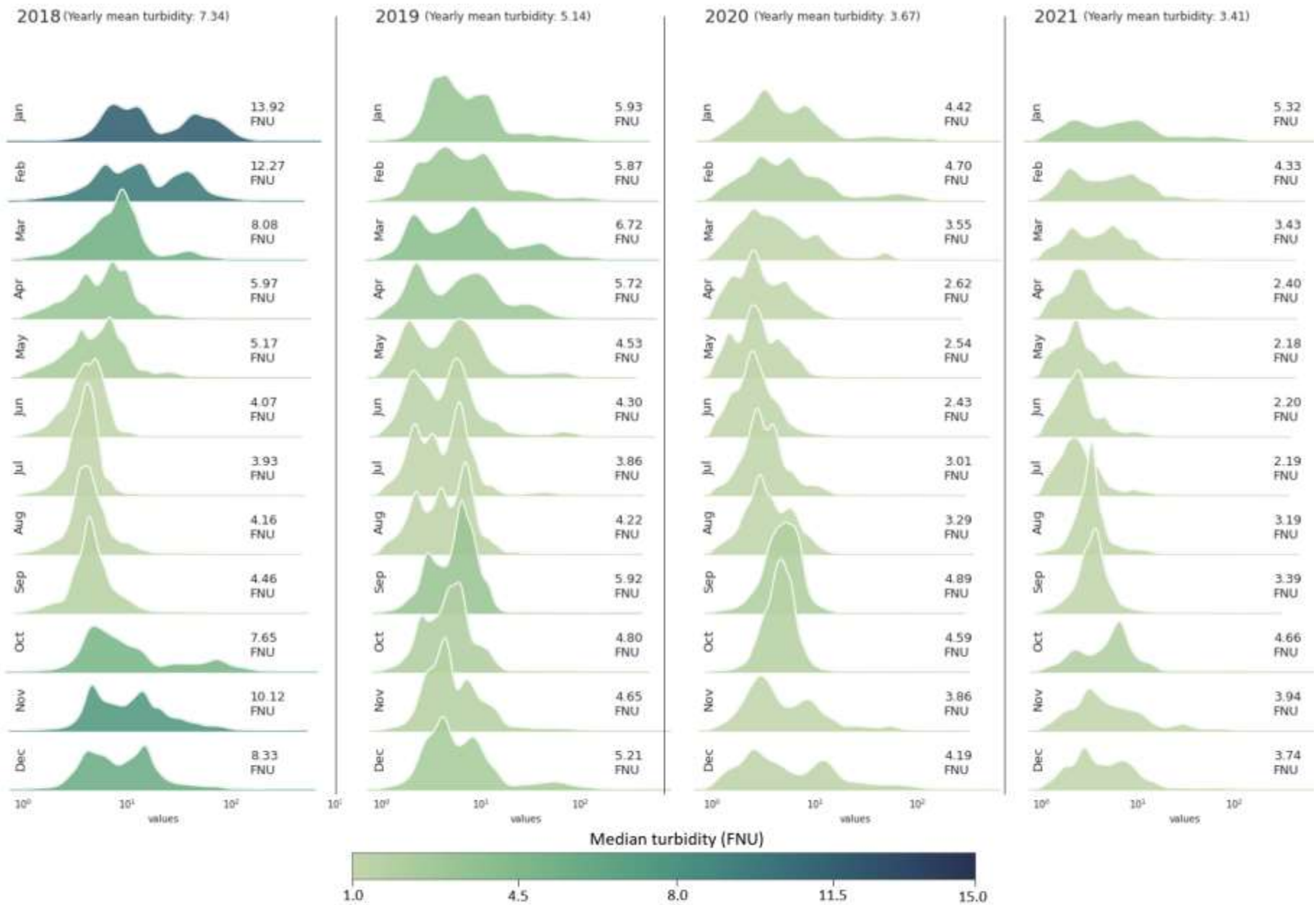


Figure III-19: Turbidity histograms for the Paranapanema region of interest. Annotated values and color scale indicate the median turbidity (FNU) value in the month.

However, it can be noted that the turbidity histograms show multimodal patterns depending on the period of the hydrological cycle, probably revealing different dynamic as a function of the water body use and size class

2.5 Discussion

Water surface variation during the 2021 drought

In this study, different assessments have been performed to better understand the effects of the 2021's drought and its impact on the water surface availability in the Parana hydrographic region through high-resolution satellite imagery. Results from the validation performed with field measurements from 16 of the biggest reservoirs in the region, with an area greater than 10,000ha, showed a very good agreement between field measurements and remote sensing derived data (R^2 of 0.91). It is well-known that area and volume do not follow a linear relationship and will depend on the bathymetry of each reservoir. However, this linear relationship can be established depending on the operational range. These finds reveal a great potential of the presented methodology for assessing water availability on a regional scale during hydrological droughts.

The overall surface assessment (Figure III-20), which also included the smaller reservoirs, showed a behavior close to the one observed for the biggest reservoirs. This result is mostly driven by the fact that the biggest reservoirs represent almost 70% of the water surface in the study region. However, an important finding is that despite the 2021 event being noticed as a major drought, the overall water area and water storage remained similar to those of the preceding year (2020). We inferred that 2020's dry period represented a significant drop in terms of water surface compared to the previous years of 2018 and 2019, and that the region did not recover average water storage levels in 2021 due to lower rainfall rates.

In addition to the high drop in the water areas that occurred in November 2020, the rainfall graph (Figure III-20, top panel), obtained from the MERGE product, shows that from September 2020 to September 2021, all the months but one (December 2020) had observed rainfall rates lower than the historical mean. Together with the low volume of the monitored reservoirs in May 2021, these observations led the Agency to raise the alert on June 1. If this rain anomaly had been timely predicted, the authorities could have raised the alert in advance, instead of waiting to the end of the wet season, to better mitigate the impacts. From this date, several rules have been imposed concerning water withdrawals and hydropower generation in the basin and the water extents dropped continuously to reach a new minimum in November 2021. Even with these adverse events, the overall situation of the basin in November 2021 remained comparable to the areas from November 2020, which can highlight the efficacy of the efforts envisaged by the authorities to control water consumption in the basin.

The stratified analysis (Figure III-22) was intended to assess how the drought impacted water bodies of different sizes. Results are noisier for small-size water bodies. This can be due to detection issues related to sensor limitations in this size range. Although the water detection methodology employed in this study can detect water bodies as small as 1ha with an overall kappa score superior to 0.8 (Cordeiro et al., 2021), the spatial resolution may struggle to capture the minor area variations of such small bodies.

Additionally, these results reveal that smaller water bodies lost slightly more area than the biggest ones, -31% versus -25% respectively, compared to the highest accumulation in the period. That's probably due to a higher percentage of natural water bodies and a lack of proper management in these smaller reservoirs. This is also evidence that good water use management can mitigate drought impacts.

Conversely, medium water bodies remained stable throughout the analysis, showing good resilience. To understand this unexpected behavior, it is essential to investigate its members. The medium-size group has 60 water bodies with a total area of 750 km². The 14 biggest water bodies in this group account for 80% of the total water surface and, among these, 13 are run-of-river (without regularization) hydroelectric dams (in descending order of water surface: Ibitinga, Taquaruçu, Cachoeira Dourada, Salto, Estreito, Bariri, Salto do Rio Verdinho, Tibaji, Canoas I, Ituparanga, Coqueiros, Canoas II and Piraju) and just one has regulated storage reservoir (Caçu). In run-of-river dams the discharge upstream and downstream of the dam are the same, by definition, so the reservoir is not designed to store water and its level remains constant.

The last analysis was intended to assess the spatial distribution of the drought impacts in the region. The results (Figure III-21) reveal that the drought impacted sub-regions differently. In the direct comparison of November 2021 to November 2020 (Figure III-21d), it is clear that the effects were mostly localized in the middle west, around the reservoirs of Ilha Solteira (Parana river) and Três Irmãos (Tietê river). Moreover, water surfaces along the Tietê river and between Tietê and Paranapanema rivers were also below the previous year. All other regions presented an increase in the water surfaces. That was the case with the reservoirs Agua Vermelha and Marimbondo in the Grande river, Itumbiara and Emborcação in the Paranaíba river and Nova Ponte in the Araguari river. Considering all these reservoirs drain to the Ilha Solteira reservoir, this spatial behavior may be due to operational rules enforced by the regulatory authorities.

Water Quality retrieval accuracy

The actual versus predicted graph for turbidity retrievals (Figure III-25) presented a good correlation with RMSE = 10.6 FTU. However, in the time series (Figure III-24), it is possible to note some outliers in the RS time series without a corresponding elevation on the field

measurements. These spikes can be verified in Jurumirim (Mars 21, 2016), Taciba (July 22 and November 19, 2017), and Ourinhos (June 2, 2016). The steepest spike occurred in Ourinhos, with a peak of 93 FNU, while field measurements remained around 10 FNU.

A visual inspection of the satellite images was done to verify the spike in Ourinhos. The satellite time series is already cleaned for invalid pixels and cloud occurrences, as explained in the 2.3- Materials and Methods section. Considering the high occurrence of clouds in this period, the valid images are only available on the following days: April 3, June 2, and July 17, and RGB image crops of 100 by 100 pixels have been selected (Figure III-31). The top images correspond to RGB crops from L2A images, and the bottom images from GRS corrected images. Red dots indicate the exact retrieval position. From the image crops, it seems that a punctual event occurred on June 2. This occurrence could have elevated the turbidity to levels far above the surrounding dates, corroborating with the satellite-retrieved time series.

Chl-*a* validation is more challenging due to the absence of relevant concentrations (most field measurements fall between 0 and 10 mg/m³) and to the noisier signal produced. This led to a higher RMSLE error (0.873 mg/m³) when compared to turbidity RMSLE (0.458 FNU). Several reasons may have contributed to this worse performance: difficulty to retrieve low level of Chl-*a* concentration over inland waters with varying level of inorganic turbidity that may dominate the overall water optical properties, adjacency effects and atmospheric interference.

A visual inspection like the one performed for turbidity was done on the October 14, 2020, spike on the Canoas I series (Figure III-32). The crop from October 14 shows a greenish pattern equivalent to an algae bloom. This pattern is not present on nearby dates. Additionally, we can note a strong sunglint in the image from November 8, but it has been almost entirely suppressed on the GRS-corrected image and did not affect the Chl-*a* retrieval.

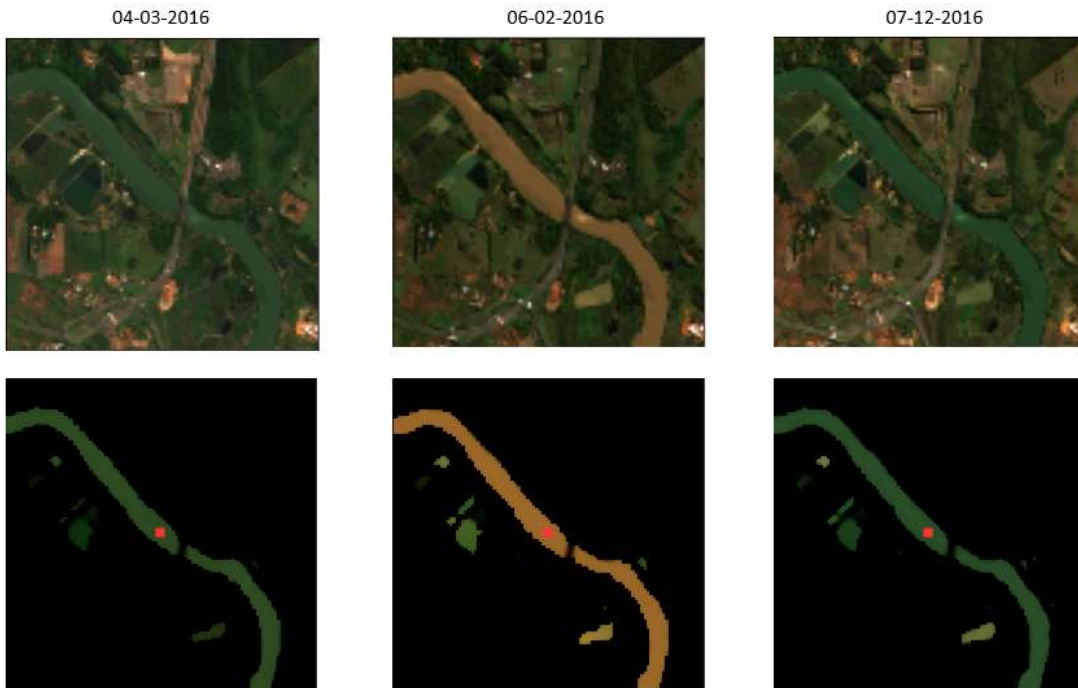


Figure III-20: Crops from the Ourinhos station, on dates surrounding the turbidity peak value occurred on June 2, 2016. Top images represent L2A RGB crops and bottom images GRS corrected water bodies. Red dot marks the retrieval position.

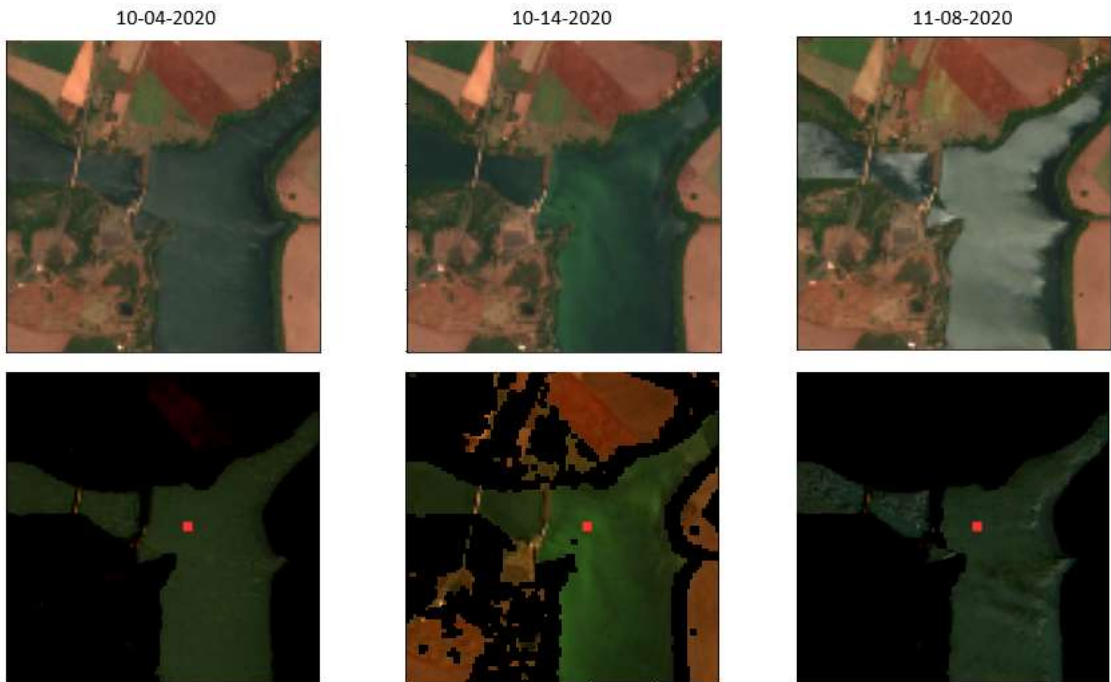


Figure III-21: Crops from the Canoas I station, on dates surrounding the Chl-a peak value occurred on October 14, 2020. Top images represent L2A RGB crops and bottom images GRS corrected water bodies. Red dot marks the retrieval position.

Chl-a Analysis

The histograms presented in section **Error! Reference source not found. Error! Reference source not found.** show an inverse relationship between Chl-*a* levels and water surface and, thus, water storage. Considering all water bodies in the area of interest, there is an overall water quality decay during dry months (from June to October) and drought events. In contrast, eutrophication is lower throughout the wet months (from November to April). This phenomenon remains true if we observe it from a yearly scale. In Figure III-23, it is possible to note that the overall area decreased from 218,000 km² in January 2018 to less than 198,000 km² in December 2021 (-9%). In the same period, there was a monotonic increase in the yearly mean Chl-*a* from 7.04 mg/m³ to 12.81 mg/m³.

Interestingly, the quarterly field measurements (Figure III-26) do not give us the whole picture due to the difficulty to capture slight variation in eutrophication patterns in space and time using pointwise estimates on a very limited number of water bodies. It is possible to note from Figure III-27 that the biggest reservoirs may present distinct trophic states simultaneously, as a function of the location within the water body.

Table III-3: Reservoirs' size classes considered for the stratified analysis.

Size Class	Minimum size (ha)	Maximum size (ha)
Very Small	2	5
Small	5	10
Medium	10	100
Large	100	-

To understand the impact on water quality caused by the drought on different size groups, we also performed a water quality analysis stratified by water body surface. A total of 4 size classes (Table III-3) have been defined to assess the impact on reservoirs of diverse sizes.

To highlight a significant Chl-*a* difference, the histograms for May 2018 (4.23 mg/m³) and September 2021 (17,63 mg/m³) are plotted for comparison (Figure III-33a, c). For visualization purposes, the bars have been replaced by the Kernel Density Estimation density function. Vertical lines represent the mode for each size class for visual reference. In May 2018, the median Chl-*a* for larger water bodies was 4.0 mg/m³, indicating mostly oligotrophic and mesotrophic waters in this size range in this month, according to the trophic class divisions proposed by Carlson and Simpson (1996). In contrast, smaller water bodies presented similar eutrophic states with median Chl-*a* values between 25.4 and 26,1 mg/m³. Moving forward to the end of the 2021s dry season, when the area achieved the worst month in terms of eutrophication (September), the large group presented a significant rise in the median Chl-*a* (17.4 mg/m³). This change represents an increment of 513% compared to May 2018 and this value corresponds to a eutrophic state. Very small, small, and medium water bodies presented a negligible variation with median values of 24.52, 24.75 and 22.92 mg/m³, in September 2021, respectively.

These results unveil that, in general, larger water bodies are prone to a greater impact in terms of water quality compared to smaller water bodies during the dry period, and these two groups have distinct dynamics for the region under study. Larger water bodies eutrophication is inversely correlated to general inflow in the basin, with higher values during the dry seasons. To visualize this difference, Figure III-34 presents bimonthly crops of the 2019s dry season, when the Chl-*a* raised from 4.18 mg/m³ on Mars to 12.6 mg/m³ in September. Small and very small water bodies are presented in Figure III-34a (center coordinates: 49.206W and 23.514S), with surfaces varying from 1 to 6 ha. Figure III-34b shows a medium-size reservoir called Couro de Boi, with 70ha (center coordinates: 51.164W and 23.132S). The third row (Figure III-34c) is the hydroelectric dam Piraju, with 1400ha (center coordinates: 49.334W and 23.173S).

Concerning the bimodal histogram that occurs in September 2019 and other months with high eutrophication, a careful analysis of the maps (Figure III-27) gives us a clue that calls for further investigation. The hypereutrophic water body that appears in the upper right section of the map is the Barra Bonita reservoir. This specific reservoir does not belong to the Paranapanema

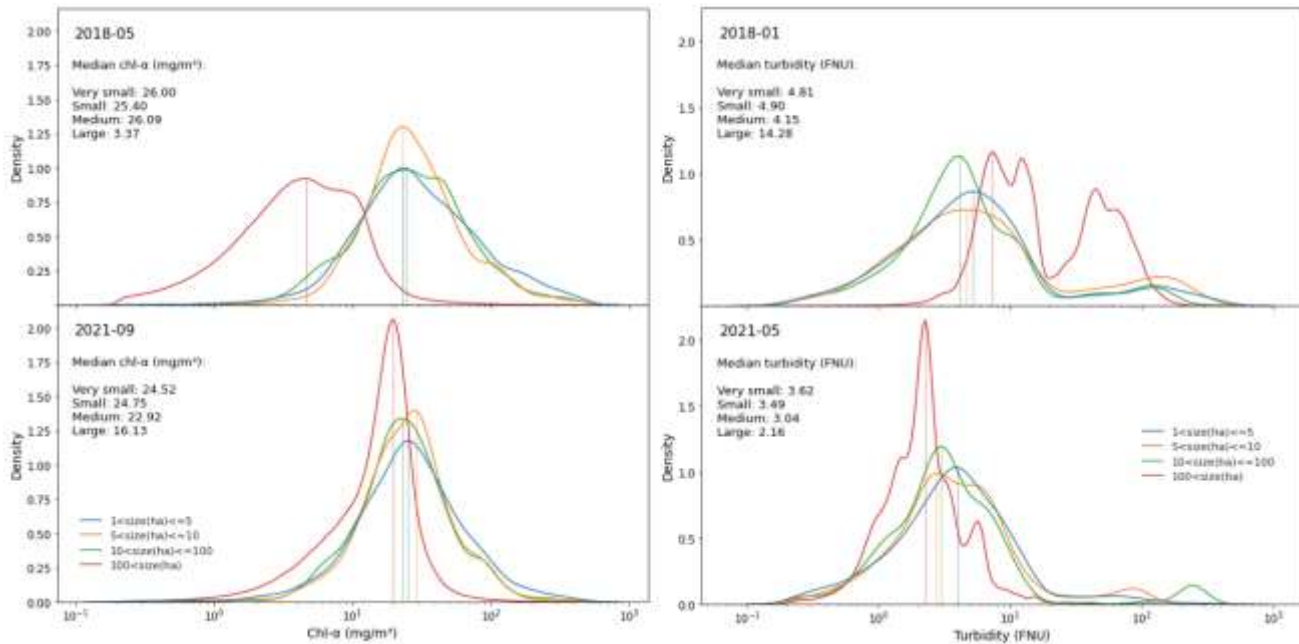


Figure III-22: (a, c) Chl-a density histograms and (b, c) turbidity density histograms for the Paranapanema basin.

basin. Instead, it is located in the Tietê basin, that is known for being a very polluted river in southeast of Brazil. To test this hypothesis, the monthly histograms have been re-generated discarding the effects of Barra Bonita reservoir and the results are seen on Figure III-35. It is possible to note that the peak around $100\text{mg}/\text{m}^3$ disappears as expected.

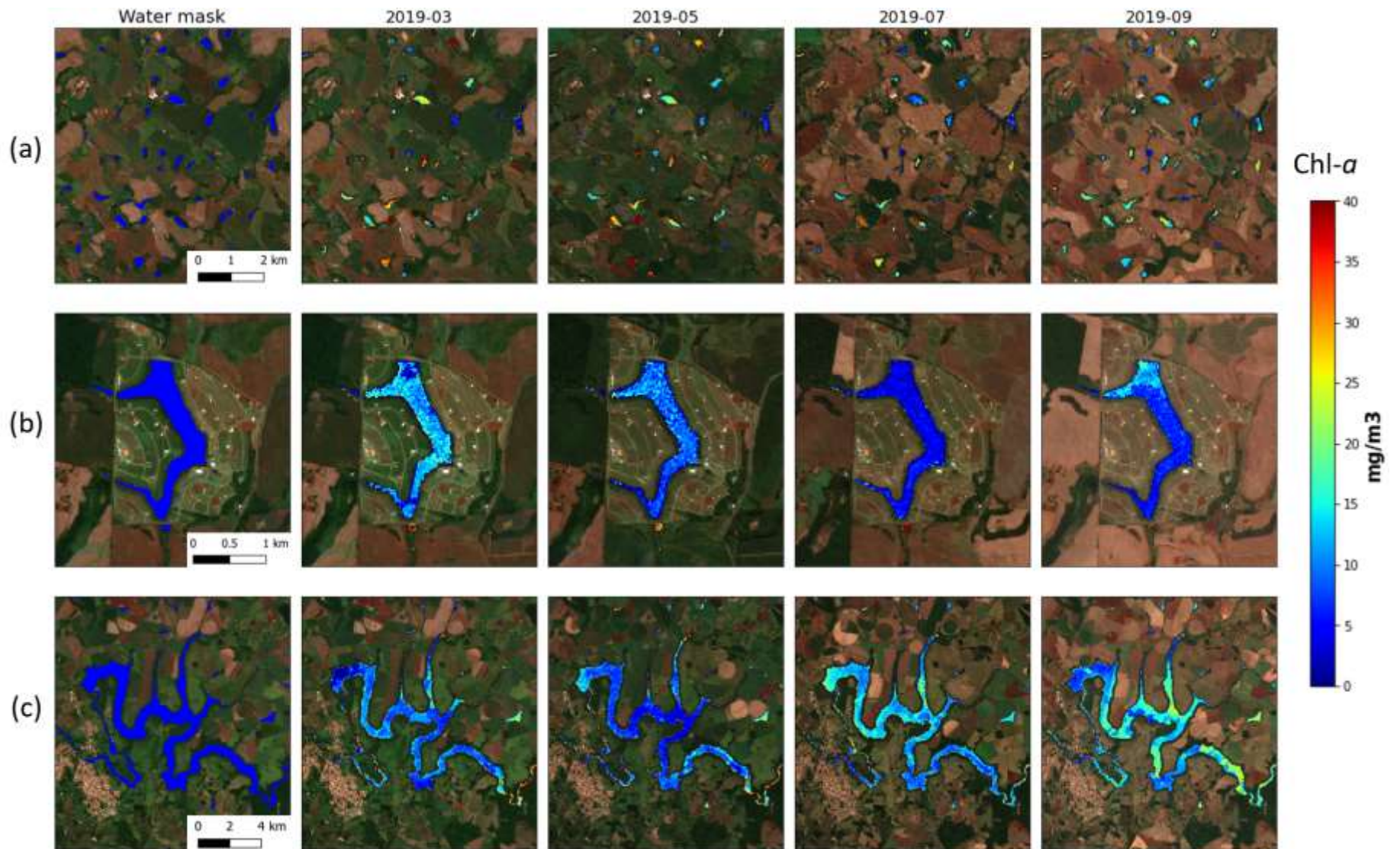


Figure III-23: Evolution in Chl-a during 2019's dry season for the following size classes: (a) Small and very small water bodies (center coordinates: 49.206W and 23.514S); (b) Couro do Boi reservoir (center coordinates: 51.164W and 23.132S); and Piraju hydroelectric dam (center coordinates: 49.334W and 23.173S).

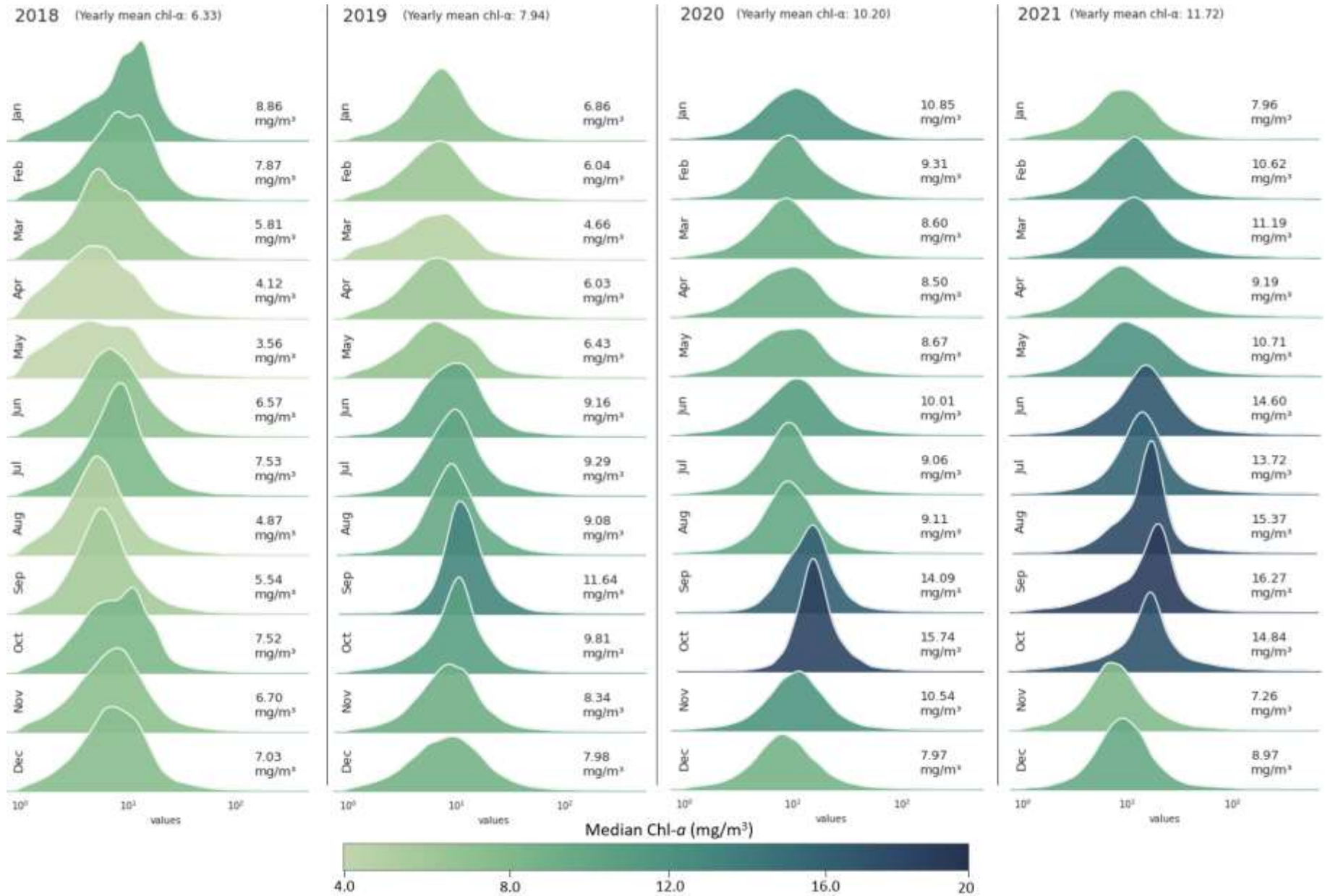


Figure III-24: Chl-a histograms for the Paranapanema region of interest, without Barra Bonita reservoir. Annotated values and color scale indicate the median Chl-a value in the month.

Turbidity Analysis

Turbidity retrievals showed better accuracy than Chl-*a* when validated with field measurements, evidencing that NIR bands have a good sensibility for this parameter. Analyzing the turbidity histograms presented in Figure III-30, it is possible to observe the opposite behavior compared to Chl-*a*. Higher turbidity levels appear in the wet season, with a peak in January 2018 (13.92 FNU), and lower levels in the dry period, reaching a minimum level in May 2021 (2.18 FNU) during the drought period. Rain and water retention time seem to play a role in turbidity as they do in Chl-*a*, but with different outcomes. During raining season, more sediment matter is carried to rivers through erosion. Besides that, lower WRT and higher river discharges induce resuspension in bed sediments.

Another observation is that turbidity histograms tend to be multimodal during high median values, while Chl-*a* are all unimodal after the correction for Barra Bonita. This may be evidence that turbidity does not affect the water bodies uniformly but shows spatially localized patterns with concentrated turbidity, typically in reservoirs' intakes, as seen in the crops from January 2018 (Figure III-36).

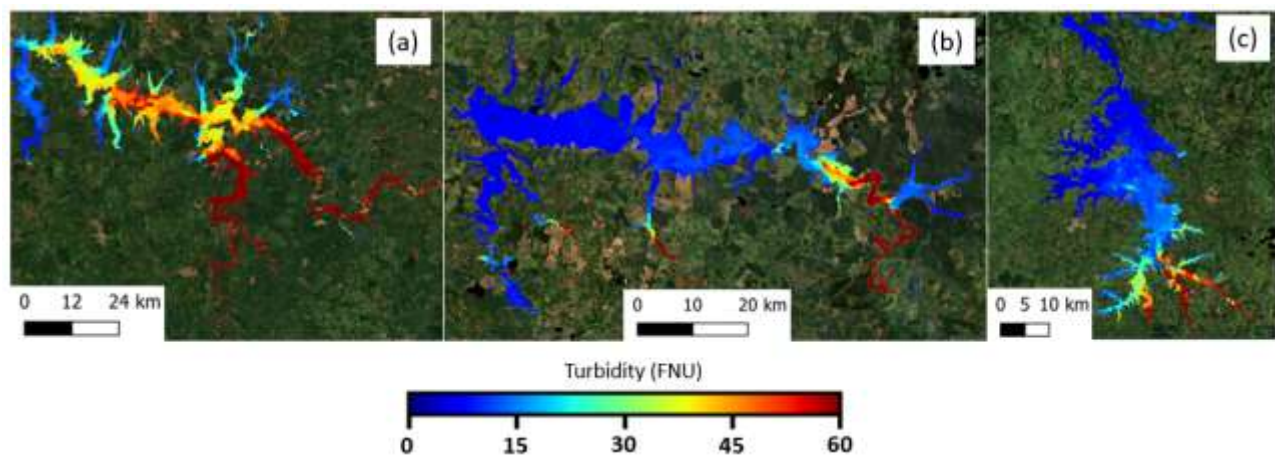


Figure III-25: Turbidity levels for Capivara (a), Jurumirim (b) and Chavantes (c) reservoirs.

Additionally, the same stratified analysis has been performed for turbidity, but January 2018 and May 2021 have been chosen as they have the highest and lowest median turbidity, respectively. The stratified histograms are presented in Figure III-33b, d.

Similar to the observed for Chl-*a*, large water bodies with surface area greater than 100ha were more affected by turbidity than smaller bodies. Median turbidity in large water bodies varied from 14.28 FNU in January 2018 to 2.16 FNU in May 2021. In the same period, variation in very small, small, and medium water bodies was marginal, from 4.81, 4.90, 4.15 FNU to 3.62, 3.49, 3.04 FNU. This apparent lower variability may be, in fact, due to the monthly time scale that has been used for this study. Sediment transport and resuspension in small water bodies may be too much rapid to be analyzed at this time scale as the hydrological processes are much more dynamic in small catchments. Further studies are necessary to confirm these findings and explain their mechanisms.

2.6 Conclusion

Continuous water availability and quality monitoring are essential for proper water resource management, especially during critical periods such as a significant drought. However, the dynamics of water quality and even volume accumulation in a basin as big as the Parana is complex and challenging to be analyzed by field measurements that are spatially punctual and sparse in time.

Considering the increasing availability of sensors with a more satisfactory spatial and temporal resolution, the use of satellite imagery can serve as a complement to field surveys. However, a comprehensive methodology is still lacking for operational use on a regional scale.

This study proposed to assess the effects of the 2021s drought that hit Brazil's southeast region, notably the Parana hydrographic region, through Sentinel-2 imagery. For that, a

methodology has been developed to analyze the water surface and quality dynamics in space and time. The Waterdetect software (Cordeiro et al., 2021) has been used to create water surface maps and the quality parameters were retrieved through models available in the literature. The region of interest covers approximately 320,000 km² and has more than 22,000 water bodies with more than 1ha.

Water surface retrieved by Sentinel-2's 10m resolution optical sensor presented good correlation to water accumulation for big reservoirs ($R^2=0.955$) and stood out as a feasible solution to monitor large surface areas. The overall water extents assessment results showed that the basin was already at critical levels the preceding year (2020), and the alert could have been raised earlier. Moreover, the spatial assessment showed that the drought affected the basin differently. Compared to the dry period of the previous year, most areas remained stable while the depletion was concentrated in the middle-west of the basin, especially around the reservoirs of Ilha Solteira and Três Irmãos.

In the water quality domain, turbidity and Chl-*a* parameters were retrieved for all water pixels through models from the literature. Eutrophication levels are directly correlated with the WRT, being higher in dry periods (lower water inflow) and lower in wet seasons (higher water inflow). As expected, higher Chl-*a* levels for the whole area were observed during the 2021s drought and dry months. On average, Chl-*a* increased 84% in 2021 (11.72 mg/m³) compared to 2018 (6.33 mg/m³) levels. However, the stratified analysis demonstrates that smaller water bodies (below 100 ha) were less affected by the water dynamics in the basin, maintaining their trophic state stable throughout the period. Further study is necessary to confirm these findings.

Concerning turbidity levels, we observed the opposite phenomenon. Turbidity was substantially lower in dry periods and during the 2021s drought, evidencing that different

mechanisms are involved. Turbidity peaks during the wet season suggests it is related to an increase in material eroded to water bodies during rain or resuspension from river bed caused by higher discharges. Additionally, higher WRT in the dry period contributes to the deposition of sediment matter on the river bed.

In summary, regional-scale water monitoring is still a big challenge for decision-makers and authorities. Despite the instability on smaller water bodies, water surface monitoring from satellite images has proved to be a viable option and provides a good estimation for great areas. It is also worth mentioning that as technology advances, orbital sensors will provide us higher resolutions allowing for more precise measurements, regardless of water bodies size. Concerning water quality, the biggest challenge is ensuring a good accuracy from inversion models, which makes room for further research in this subject. Even so, the results were promising and presented us with great potential for large-scale monitoring and possible identification of trends and anomalies.

Conclusion and Prospects

The significant increase in the number of optical satellite sensors available and the ongoing improvement of their spatial, temporal, and spectral resolution pave the way for its application in integrated water resources management.

In recent years, more and more scientific publications have been devoted to the use of RS in the context of inland water bodies for water quality monitoring. However, in contrast to products available for land or oceanic applications (e.g., land use, sea surface temperature, and others), there are neither regional nor global products available for operational monitoring of inland water bodies. The reason for this lack of robust products lies in the complex interactions among water optically active constituents, such as organic and inorganic matter, and the interference caused by atmospheric conditions that represent major challenges that need to be overcome for a global use of RS images for lakes, reservoirs, and rivers.

During the present thesis, several methodologies have been developed towards robustness, scale, and automation of inland water assessment, through the extensive use of data science techniques, such as machine learning and big data. In this context, three main axes have been proposed, as depicted in Figure IV-1, and are represented by boxes in orange, blue and green colors.

Moreover, orchestrating all the steps required for proper inland water assessment, from imagery acquisition, preprocessing and final analysis, represents an additional layer of complexity to the study.

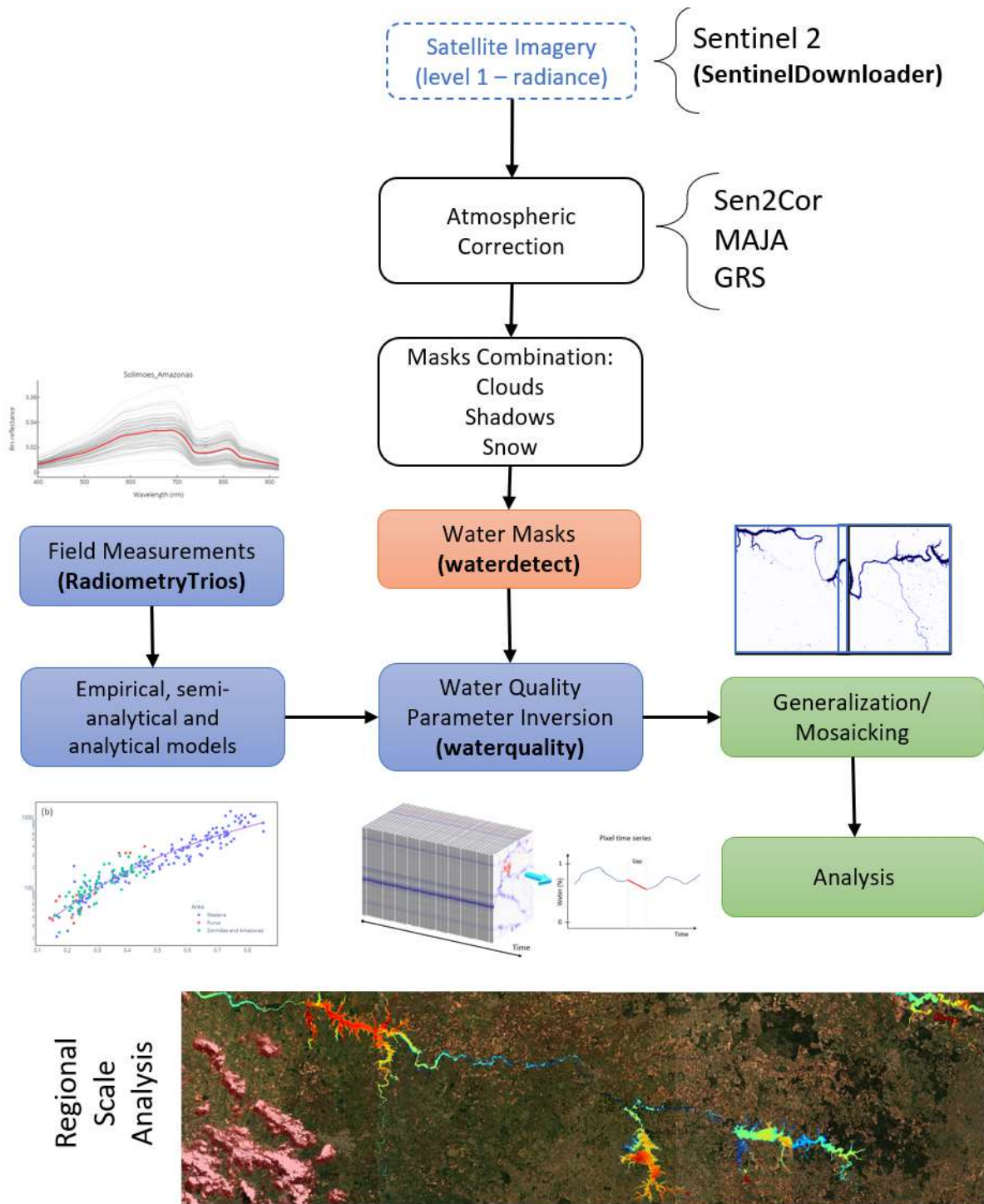


Figure IV-1: Schematic view of the steps involved in assessing inland water quality from satellite remote sensed imagery. Boxes in orange represent steps covered in axis 1 (water detection), boxes in blue represent axis 2 (water classification) and green boxes represent axis 3 (generalization and analysis). Names inside parentheses identify python packages developed in the context of the thesis.

In the first axis, represented by the orange box in Figure IV-1, the water detection problem has been addressed through the development of an unsupervised clustering algorithm called **waterdetect** (Cordeiro et al., 2021). The method, available as an open-source Python package at www.github.com/cordmaur/waterdetect, uses a multidimensional clustering approach based on reflectance values and water indices to identify water pixels without previous knowledge of the scene or other ancillary data and threshold calibration. To guarantee high-performance and low memory consumption for high resolution imagery, this process is coupled with random sampling and a subsequent machine learning classifier to reconstruct the solution to the whole scene.

The proposed method is independent of the sensor and the coverage area being analyzed and can be applied to various conditions. Compared to the most common thresholding approaches and to the three major Level 2 processors available for Sentinel-2, namely, MAJA, Sen2Cor, and FMask, the results showed that the **waterdetect** yielded the best accuracy for all body size ranges analyzed. Furthermore, its performance, robustness, and simplicity make it ideal for operational use with optical imagery.

During the thesis, further studies have been conducted in this field and the original algorithm has been much improved since its conception. Additionally, I participated as co-author in three more papers that addressed the water detection subject in diverse situations. In Peña-Luque et al. (2021), we confirmed the advantage of using multidimensional clustering over other traditional water detection methods (Canny-edge and Hue, Saturation, Value - HSV) when classifying water at both local (i.e., reservoir) and regional levels (Annex I).

In Tottrup et al. (2022), the **waterdetect** algorithm was compared with other water detection methods and tested in several challenging conditions (Annex II). Results indicated that **waterdetect** had higher average accuracy than other tested methods that used only optical input.

In the third paper (Tayer et al., 2022, in review), we compared the performance of Sentinel-2, PlanetScope and sharpened imagery to detect an intermittent river in western Australia. The relevance of the SWIR band for water detection was also assessed.

The second axis, represented by the blue boxes in Figure IV-1, focused on employing water classification schemes to improve the accuracy of water quality parameters inversion. Suspended sediment matter (SPM) has been chosen for the study as it is the first driver of the water leaving radiance over inland waters. In contrast to previous studies with optical water types (OWT), our objective was to define water classes based on their reflectance spectra, that could maximize the inversion accuracy rather than provide a comprehensive qualitative classification schema of water types.

Once again, we used unsupervised machine learning (agglomerative clustering) to classify the spectra from more than 1,000 field measurements into several groups. Clustering performance was analyzed by error metrics in terms of SPM retrieval.

The results show that calibrating SPM models for each optical water type (OWT) can increase the overall retrieval accuracy without needing to produce one model for each river, basin, or geographic region. The conclusion is that four classes are enough, and it matches the RMSLE accuracy achieved by calibrating one model for each of the 13 areas included in the study. To the best of our knowledge, this study is the first that focused on optimizing the classes for SPM retrieval, specifically over inland waters.

The third axis (green boxes in Figure IV-1) presented the methodologies developed toward the large-scale generalization and regional analysis of inland water surface and water quality parameters. Concepts of Big Data such as cloud computing and parallel processing were employed to enable the processing of thousands of images that would not be feasible with conventional tools.

Furthermore, we proposed an assessment methodology, presented as a case study, to assess the effects of the major drought that hit Brazil's southeast in 2021 on the water quality of large, medium and small reservoirs.

In the case study, remote sensing-retrieved water surface was used to analyze the drought effects on reservoirs of different sizes and its spatial dynamics. Initially, the data validation showed a good agreement between RS estimated surface and ground truth data ($R^2 = 0.92$). Additionally, a linear relationship could be established between water surface and the accumulated volume in big reservoirs with a determination coefficient R^2 of 0.96, demonstrating the feasibility of RS to monitor accurately the variability of reservoir surfaces.

The stratified analysis presented little difference between small and large water bodies in terms of total depletion during the drought period. The spatial analysis showed that the effects of the drought were more prominent in the center-west of the basin, while other areas remained at normal levels. Considering that rain levels remained below average in the region, most of this resilience can be credited to the restrictions applied by regulatory authorities.

To assess the water quality parameters (Chl-*a* and turbidity), monthly histograms have been elaborated using the pixel values within all water bodies with surface sizes greater than 1ha and presented as Ridge's graphs. Results for Chl-*a* showed that eutrophication is significantly higher in the dry seasons when water retention time tends to be increase. The overall median Chl-*a* increased 84% from 2018 (6.33 mg/m³) to 2021 (11.72 mg/m³), during the drought, when water storage accumulation reached its minimum levels due to the strong drought that hit the region. In the stratified analysis, the results showed that smaller water bodies below 100ha were less affected by the overall water dynamics in the basin, remaining with a higher eutrophication level throughout the period of analysis.

In terms of turbidity, the opposite phenomenon was observed, with more transparent waters during the dry season, evidencing that different mechanisms are involved. Turbidity peaks during wet season suggests a direct relationship with suspended sediment transport from the local catchments during rainy events or from resuspension processes from river bed caused by higher water discharges.

In summary, large-scale inland water monitoring through RS remains a challenge due to the complexities involved and requires further scientific research. However, the current work permitted to address several challenges related to the use of RS for inland water bodies assessment. More specifically, in the water detection context (Orange box – Figure IV-1), the proposed method provided an easy-to-use package that can be applied in distinct conditions, as it does not require pre-training or threshold calibration and it is not constrained to a specific satellite sensor. Additionally, thanks to its high performance, scientists can increase the area of analysis without the need to degrade its spatial resolution.

Another scientific question that this thesis contributed to answer concerns the difficulty to identify the best inversion algorithm to be applied to a certain water condition (Blue boxes – Figure IV-1). The proposed classification scheme allows the development of a concise number of tailored models that can increase retrieval accuracy without the need to develop one model for each river/basin.

Orchestrating all the steps presented in Figure IV-1 into a fully operational workflow represents another major challenge. One of the reasons is the lack of ready-to-use packages to perform specific steps. To fill some of these gaps, besides the **waterdetect** algorithm, other three packages have been developed under the context of the thesis and were made available to the public. These packages are identified in parentheses in Figure IV-1. To automate the retrieval of

Sentinel-2 imagery, the **SentinelDownloader** package (<https://github.com/cordmaur/SentinelDownloader>) provides a simple Python interface to download Sentinel imagery from the Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). Additionally, the **RadiometryTrios** package (<https://github.com/cordmaur/RadiometryTrios>) provides an interface to manipulate hyperspectral radiometry measurements obtained from field campaigns. This pre-processing tool involves data-cleaning, radiometric interpolation, reflectance generation and others and are requisite for developing retrieval algorithms based on hyperspectral measurements. Finally, the **waterquality** package (<https://github.com/cordmaur/WaterQuality>) extends the functionalities of the **waterdetect** to calculate water quality parameters through the application of inversion algorithms to the satellite reflectances.

Although a fully integrated and automated workflow is still a goal to be pursued, the current thesis proposed a comprehensive methodology for large scale processing and analysis, as detailed in Chapter III. Overall, the current work highlighted the feasibility of employing data science techniques to leverage remote sensing as a viable tool for complementing sparse field measurements and assessing water bodies that are otherwise impossible or unviable to monitor.

Prospects

The current thesis paves the way for future applied research projects on inland water quality. Although the methodologies have been applied successfully in a case study on a regional scale, more research is still necessary to orchestrate all the steps involved to deliver a complete operational framework for inland water assessment.

Some domains that are required to complete the whole task still need improvements. Cloud detection is one example. Current atmospheric correction processors deliver coarse (low spatial resolution) cloud masks, and third-party tools, such as the S2Cloudless, do not take into account

the corresponding shadows. In addition, performance is not optimal, with an overall accuracy below 85% in more challenging scenes (Baetens et al., 2019), and there are many classification issues between water surface and cloud shadows.

Although feasible for regional studies and trending analysis, water quality parameters inversion is also a domain that requires additional research. The water classification scheme introduced in Chapter II represents a step forward as it proposes classes based on the pixels' reflectance that aims to minimize inversion error. Furthermore, this water classification enables the application of tailored algorithms and improves general retrieval accuracy.

However, when validated against punctual field measurements, accuracy is below optimal, especially for low concentrations (when results are noisier), evidencing that the work is being done within the limits of the satellite sensibility (radiometric resolution). Moreover, atmospheric effects can broadly impact the results, and more atmospheric correction algorithms specialized in inland waters should be considered. In this regard, the correction of the sunglint done by GRS is fundamental, as its presence imposes two challenges. First, during water pixel extraction, strong sunglint can prevent current algorithms from correctly identifying the water surface, reducing the accuracy of quantitative analysis. Even if this behavior is treated during the water detection step (e.g., the **waterdetect** has a sunglint detection processor to overcome this issue), a minor presence of sunglint during water quality inversion can lead to erroneous results.

From an end-user perspective, it is cumbersome to put all the pieces to work together and generate final water-quality products for several reasons. First, the existing tools are oriented to the data acquisition structure (i.e., satellite tiles), and they cannot be applied directly to the area of interest, requiring multiple tiles processing, and mosaicking, among other pre-processing steps. This reality tends to evolve as cloud platforms that provide satellite data “on-demand”, such as the

Google Earth Engine, become more popular. Another gap for end users is the complexity involved in running scientific software. Usually, the packages necessary for the required steps (e.g., masking, inversion, and others) are designed for experimental purposes and require scientific knowledge that is specific for each domain. Besides, they lack proper software design and documentation, making them difficult to be used.

Another point that has been overcome in the case study but needs improvement for a better end-use experience regards the processing of large volumes of data. In the methodology presented in Chapter 3, distributed processing has been performed using Dask. Through this technology, the main task is broken into multiple tasks which are sent to distributed workers. This approach can solve memory and processing restrictions, but the setup and use of such tools must be done programmatically and normally require technical knowledge.

Finally, besides the demonstrative character of using recent satellite data, like the optical imagery from Sentinel-2, for water quality assessment, the tools and methods developed in the present thesis can be adapted to other sensors (space- or air-borne), parameters and even to address subjects other than water. In this regard, sensors with higher spectral resolution, such as the one onboard Sentinel-3 mission, could be employed to improve the retrieval accuracy of water quality parameters in larger water bodies. Alternatively, higher spatial resolution, such as the 3 meters Planetscope sensor, could be applied to assess changes in narrow or intermittent river channels (Tayer et al., 2022, in review).

In summary, the use of RS in the water resources domain is vast and requires further development. In this context, the current work represents a step forward, highlighting the feasibility of employing data science techniques to leverage RS as a viable tool for complementing sparse

field measurements and assessing water bodies which are otherwise impossible or unviable to monitor.

Conclusion et Perspectives

L'augmentation significative du nombre de capteurs optiques satellite disponibles et l'amélioration continue de leurs résolutions spatiale, temporelle et spectrale ouvrent la voie à leur application dans la gestion intégrée des ressources en eaux.

Ces dernières années, de plus en plus de publications scientifiques ont été consacrées à l'utilisation de la télédétection dans le contexte des masses d'eau continentales pour le suivi de la qualité des eaux. Cependant, contrairement aux produits disponibles pour des applications terrestres ou océaniques (par exemple, le classement des types de couvertures des sols, la température de surface de la mer, et autres), il n'existe pas de produits régionaux ou mondiaux disponibles pour la surveillance opérationnelle des eaux continentales. La raison de ce manque de produits robustes réside notamment dans les interactions complexes entre les constituants de l'eau optiquement actifs, tels que les matières organiques et inorganiques, et les interférences causées par les conditions atmosphériques qui représentent des défis majeurs à surmonter pour une utilisation globale et robuste des images par télédétection concernant des lacs, rivières et réservoirs.

Au cours de la présente thèse, plusieurs méthodologies ont été développées afin d'améliorer la robustesse, le changement d'échelle et l'automatisation de l'évaluation des eaux continentales, grâce à l'utilisation extensive de techniques de science des données, telles que le *Machine Learning* et le *Big Data*. Dans ce contexte, trois axes principaux ont été proposés, comme le montre la Figure IV-2, et sont représentés par des cases de couleurs orange, bleue et verte.

En outre, orchestrer toutes les étapes nécessaires à une évaluation correcte des eaux continentales depuis l'acquisition d'images, le prétraitement et l'analyse finale, a représenté une couche de complexité supplémentaire pour l'étude.

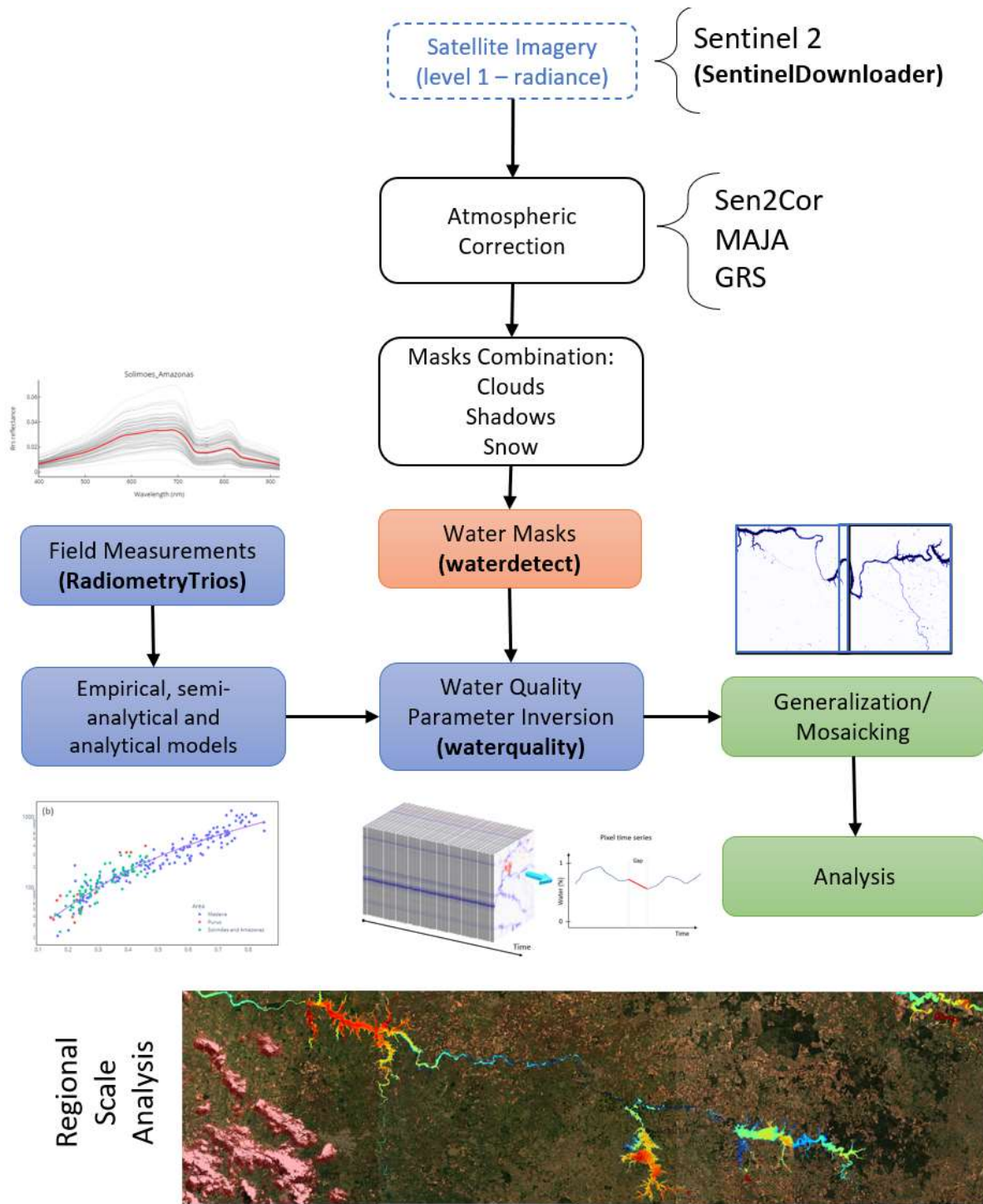


Figure IV-2: Vue schématique des étapes de l'évaluation de la qualité des eaux intérieures à partir d'images de télédétection par satellite. Les cases en orange représentent les étapes couvertes par l'axe 1 (détection de l'eau), les cases en bleu représentent l'axe 2 (classification de l'eau) et les cases vertes représentent l'axe 3 (généralisation et analyse). Les noms entre parenthèses identifient les paquets python développés dans le cadre de la thèse.

Dans le premier axe, représenté par la boîte orange de la Figure IV-2, le problème de la détection de l'eau a été abordé par le développement d'un algorithme de clustering non supervisé appelé **waterdetect** (Cordeiro et al., 2021). La méthode, disponible sous forme d'un paquet Python open-source à l'adresse www.github.com/cordmaur/waterdetect, utilise une approche de clustering multidimensionnel basée sur les valeurs de réflectance et des indices spectraux dédiés à l'eau pour identifier les pixels d'eaux sans la nécessité d'une connaissance préalable sur la scène ou d'autres données auxiliaires et sans calibration de seuils. Pour garantir des performances élevées et une faible consommation de mémoire pour les images à haute résolution, ce processus est associé à un échantillonnage aléatoire et à un classificateur d'apprentissage automatique subséquent pour reconstruire la solution à l'ensemble de la scène.

La méthode proposée est indépendante du capteur et de la zone de couverture analysée et peut être appliquée à diverses conditions. Comparée aux approches de seuillage les plus courantes et aux trois principaux processeurs de niveau L2 disponibles pour Sentinel-2, à savoir MAJA, Sen2Cor et FMask, les résultats ont montré que la détection des eaux avec **waterdetect** donnait la meilleure précision pour toutes les gammes de tailles de pièces d'eaux analysées à partir de 0,5 ha. Les performances, la robustesse et la simplicité de **waterdetect** en font un outil idéal pour une utilisation opérationnelle avec l'imagerie optique.

Pendant la thèse, d'autres études ont été menées dans ce domaine et l'algorithme original a été beaucoup amélioré depuis sa conception. Ainsi, j'ai participé en tant que co-auteur à trois autres articles qui ont abordé le sujet de la détection de l'eau dans diverses situations. Dans Peña-Luque et al. (2021), nous avons confirmé l'avantage de l'utilisation du clustering multidimensionnel par rapport aux autres méthodes traditionnelles de détection de l'eau (Canny-edge et Hue, Saturation,

Value - HSV) lors de la classification de l'eau aux niveaux local (c'est-à-dire le réservoir individuel) et régional (Annexe I).

Dans Tottrup et al. (2022), l'algorithme **waterdetect** a été comparé à d'autres méthodes de détection de l'eau et testé dans des conditions environnementales extrêmes depuis le Groenland jusqu'au Gabon (Annexe II). Les résultats indiquent que **waterdetect** a une précision moyenne plus élevée que les autres méthodes testées basées exclusivement sur les données optique.

Dans le troisième article (Tayer et al., 2022, en cours de révision), nous avons comparé les performances de Sentinel-2, Planetscope et de l'imagerie très haute résolution pour détecter une rivière intermittente en Australie occidentale. La pertinence de la bande SWIR pour la détection de l'eau a également été évaluée dans un contexte semi-aride.

Le deuxième axe, représenté par les cases bleues de la Figure IV-2, se concentre sur l'utilisation d'algorithme de classification de l'eau pour améliorer la précision de l'inversion des paramètres de qualité de l'eau. Les matières en suspension (MES ou SPM en anglais) ont été choisie pour l'étude car elle constitue le premier facteur de contrôle de la radiance de sortie de l'eau dans les eaux intérieures. Contrairement aux études précédentes sur les types d'eau optiques (OWT), notre objectif était de définir des classes d'eau basées sur leurs spectres de réflectance, qui pourraient maximiser la précision de l'inversion plutôt que de fournir un schéma de classification qualitative complet des types d'eau.

Une fois encore, nous avons utilisé l'apprentissage automatique non supervisé (clustering agglomératif) pour classer les spectres de plus de 1 000 mesures sur le terrain en plusieurs groupes. La performance du clustering a été analysée en quantifiant l'erreur d'inversion de la SPM.

Le troisième axe (cases vertes de la Figure IV-2) présente les méthodologies développées pour la généralisation à grande échelle et l'analyse régionale des paramètres de quantité (surface

en eaux) et de qualité de l'eau des eaux intérieures. Des concepts de *Big Data* tels que le *Cloud Computing* et le traitement parallèle ont été employés pour permettre le traitement de milliers d'images qui ne serait pas réalisable avec des outils conventionnels. En outre, nous avons proposé une méthodologie d'évaluation, présentée sous forme d'étude de cas, pour évaluer les effets de la grande sécheresse qui a frappé le sud-est du Brésil en 2021 sur la qualité de l'eau des lacs naturels et de barrages (« réservoirs ») de grande, moyenne et petite taille.

Le troisième axe (cases vertes de la Figure IV-2) présente les méthodologies développées pour la généralisation à grande échelle et l'analyse régionale des paramètres de surface et de qualité de l'eau des eaux intérieures. Des concepts de *Big Data* tels que le *Cloud Computing* et le traitement parallèle ont été employés pour permettre le traitement de milliers d'images qui ne seraient pas réalisables avec des outils conventionnels. En outre, nous avons proposé une méthodologie d'évaluation, présentée sous forme d'étude de cas, pour évaluer les effets de la grande sécheresse qui a frappé le sud-est du Brésil en 2021 sur la qualité de l'eau des réservoirs de grande, moyenne et petite taille.

Dans cette étude de cas, la surface de l'eau estimée par télédétection a été utilisée pour analyser les effets de la sécheresse sur des réservoirs de différentes tailles et sa dynamique spatiale. Initialement, la validation des données a montré un bon accord entre la surface estimée par télédétection et les données de vérité au sol ($R^2 = 0,92$). De plus, une relation linéaire a pu être établie entre la surface de l'eau et le volume accumulé dans les grands réservoirs avec un coefficient de détermination R^2 de 0,96, démontrant la faisabilité de la télédétection pour surveiller avec précision la variabilité des surfaces des réservoirs.

L'analyse stratifiée a présenté peu de différence entre les petites et les grandes masses d'eau en termes de pertes d'eaux pendant la période de sécheresse. L'analyse spatiale a montré que les

effets de la sécheresse étaient plus importants dans le centre-ouest du bassin, alors que les autres zones sont restées à des niveaux normaux. Si l'on considère que les niveaux de pluie sont restés inférieurs à la moyenne dans la région, la majeure partie de cette résilience peut être attribuée aux restrictions appliquées par les autorités réglementaires.

Pour évaluer les paramètres de qualité de l'eau (Chl-*a* et turbidité), des histogrammes mensuels ont été élaborés en utilisant les valeurs des pixels dans tous les plans d'eau dont la surface est supérieure à 1ha et présentés sous forme de graphiques de Ridge. Les résultats pour le Chl-*a* ont montré que l'eutrophisation est significativement plus élevée pendant les saisons sèches lorsque le temps de rétention de l'eau tend à augmenter. La médiane globale de Chl-*a* a augmenté de 84 % entre 2018 (6,33 mg/m³) et 2021 (11,72 mg/m³), pendant la sécheresse, lorsque l'accumulation du stockage de l'eau a atteint ses niveaux minimums en raison de la forte sécheresse qui a frappé la région. Dans l'analyse stratifiée, les résultats ont montré que les plans d'eau plus petits, inférieurs à 100ha, ont été moins affectés par la dynamique globale de l'eau dans le bassin, restant avec un niveau d'eutrophisation plus élevé tout au long de la période d'analyse.

En termes de turbidité, le phénomène inverse a été observé, avec des eaux plus transparentes pendant la saison sèche, ce qui prouve que différents mécanismes sont impliqués. Les pics de turbidité pendant la saison des pluies suggèrent une relation directe avec le transport de sédiments en suspension provenant des bassins versants locaux pendant les événements pluvieux ou des processus de remise en suspension du lit de la rivière causés par des débits d'eau plus élevés.

En résumé, la surveillance à grande échelle des eaux intérieures par le biais de la télédétection reste un défi en raison de sa complexité et nécessite des recherches scientifiques supplémentaires. Cependant, le travail actuel a permis de relever plusieurs défis liés à l'utilisation de la télédétection pour l'évaluation des masses d'eau intérieures. Plus précisément, dans le contexte

de la détection des eaux (cases orange - Figure IV-2), la méthode proposée a fourni un ensemble facile à utiliser qui peut être appliqué dans des conditions distinctes, car elle ne nécessite pas de pré-entraînement ou de calibrage de seuil et qu'elle n'est pas limitée à un capteur satellite spécifique. De plus, grâce à sa haute performance, les scientifiques peuvent augmenter la zone d'analyse sans avoir à dégrader la résolution spatiale des résultats.

Une autre question scientifique à laquelle cette thèse a contribué à répondre concerne la difficulté d'identifier le meilleur algorithme d'inversion à appliquer à une certaine condition d'eau (cases bleues - Figure IV-2). Le schéma de classification proposé permet le développement d'un nombre concis de modèles adaptés qui peuvent augmenter la précision de l'extraction sans avoir besoin de développer un modèle pour chaque rivière/bassin.

L'orchestration de toutes les étapes présentées dans la Figure IV-2 en un flux de travail pleinement opérationnel représente un autre défi majeur. L'une des raisons en est le manque de paquets prêts à l'emploi pour réaliser des étapes spécifiques. Pour combler certaines de ces lacunes, outre l'algorithme **waterdetect**, trois autres paquets ont été développés dans le cadre de la thèse et ont été mis à la disposition du public. Ces paquets sont identifiés entre parenthèses dans la Figure IV-2. Pour automatiser la récupération de l'imagerie Sentinel-2, le paquet **SentinelDownloader** (<https://github.com/cordmaur/SentinelDownloader>) fournit une interface Python simple pour télécharger l'imagerie Sentinel depuis le Copernicus Open Access Hub (<https://scihub.copernicus.eu/>). En outre, le paquet **RadiometryTrios** (<https://github.com/cordmaur/RadiometryTrios>) fournit une interface permettant de manipuler les mesures radiométriques hyper-spectrales obtenues lors de campagnes sur le terrain. Ce prétraitement comprend le nettoyage des données, l'interpolation radiométrique, la génération de la réflectance et d'autres éléments nécessaires au développement d'algorithmes d'extraction basés

sur des mesures hyper-spectrales. Enfin, le paquet de qualité de l'eau (<https://github.com/cordmaur/WaterQuality>) étend les fonctionnalités de **waterdetect** pour calculer les paramètres de qualité de l'eau par l'application d'algorithmes d'inversion aux réflectances du satellite.

Bien qu'un flux de travail entièrement intégré et automatisé reste un objectif à poursuivre, la thèse actuelle a proposé une méthodologie complète pour le traitement et l'analyse à grande échelle, comme détaillé dans le Chapitre III. Dans l'ensemble, le travail actuel a mis en évidence la faisabilité de l'utilisation de techniques de science des données pour exploiter la télédétection comme un outil viable pour compléter les mesures de terrain éparses et évaluer les masses d'eau qu'il serait autrement impossible ou non viable de surveiller.

Perspectives

La présente thèse ouvre la voie à de futurs projets de recherche appliquée sur la qualité des eaux intérieures. Bien que les méthodologies aient été appliquées avec succès dans une étude de cas à l'échelle régionale, des recherches supplémentaires sont encore nécessaires pour harmoniser toutes les étapes impliquées afin de fournir un cadre opérationnel complet pour l'évaluation des eaux intérieures.

Certains domaines nécessaires pour mener à bien l'ensemble de la tâche doivent encore être améliorés. La détection des nuages en est un exemple. Les processeurs de correction atmosphérique actuels fournissent des masques de nuages assez grossiers (à faible résolution spatiale), et les outils tiers, tels que le S2Cloudless, ne prennent pas en compte les ombres correspondantes. En outre, les performances ne sont pas optimales, avec une précision globale inférieure à 85 % dans des scènes plus difficiles (Baetens et al., 2019), et il existe de nombreux problèmes de classification entre les ombres de nuages à la surface de l'eau et des nuages.

Bien que réalisable pour les études régionales et l'analyse des tendances, l'inversion des paramètres de qualité de l'eau est également un domaine qui nécessite des recherches supplémentaires. Le schéma de classification de l'eau présenté au chapitre II représente un pas en avant car il propose des classes basées sur la réflectance des pixels qui visent à minimiser l'erreur d'inversion. En outre, cette classification de l'eau permet l'application d'algorithmes adaptés et améliore la précision générale de l'extraction.

Cependant, lorsqu'elle est validée par rapport à des mesures ponctuelles sur le terrain, la précision est inférieure à l'optimum, en particulier pour les faibles concentrations (lorsque les résultats sont plus bruités) lorsque l'inversion est effectuée dans les limites de sensibilité du satellite (résolution radiométrique). De plus, les effets atmosphériques peuvent avoir un impact important sur les résultats, et il est important de considérer des algorithmes de correction atmosphérique spécifiques aux eaux intérieures. À cet égard, la correction du sunglint, notamment effectuée par la GRS, est fondamentale car sa présence impose deux défis. Tout d'abord, lors de l'extraction des pixels d'eau, la réflexion directe de la lumière solaire peut empêcher les algorithmes actuels d'identifier correctement la surface de l'eau, ce qui réduit la précision de l'analyse quantitative. Même si ce comportement est traité au cours de l'étape de détection de l'eau (par exemple, **waterdetect** dispose d'un processeur de détection de la teinte solaire pour surmonter ce problème), une présence mineure de la réflexion solaire au cours de l'inversion de la qualité de l'eau peut conduire à des résultats erronés.

Du point de vue de l'utilisateur final, il est actuellement difficile d'assembler tous les éléments et de générer des produits finaux sur la qualité de l'eau pour plusieurs raisons. Premièrement, les outils existants sont orientés vers la structure d'acquisition des données (c'est-à-dire les « tuiles » de satellite) et ne peuvent pas être appliqués directement à la zone d'intérêt d'un

utilisateur, ce qui nécessite le traitement de plusieurs tuiles et le mosaïquage, entre autres étapes de prétraitement. Cette réalité tend à évoluer au fur et à mesure que les plateformes *cloud* qui fournissent des données satellitaires "à la demande", comme le Google Earth Engine, deviennent plus populaires. Une autre lacune pour les utilisateurs finaux est la complexité liée à l'utilisation de logiciels scientifiques. Habituellement, les paquets nécessaires aux étapes requises (par exemple, le masquage, l'inversion, et autres) sont conçus à des fins expérimentales et nécessitent des connaissances scientifiques spécifiques à chaque domaine. En outre, ils ne sont pas conçus et documentés correctement, ce qui rend leur utilisation difficile.

Un autre point qui a été résolu dans l'étude de cas, mais qui doit être amélioré pour une expérience d'utilisation finale optimale, concerne le traitement de grands volumes de données. Dans la méthodologie présentée au chapitre 3, le traitement distribué a été effectué à l'aide du paquet Dask. Grâce à cette technologie, la tâche principale est divisée en plusieurs tâches qui sont envoyées à des travailleurs distribués. Cette approche peut résoudre les problèmes de mémoire et de traitement, mais la configuration et l'utilisation de ces outils doivent être programmées et nécessitent des connaissances techniques assez poussées.

Enfin, outre le caractère démonstratif de l'utilisation de données satellitaires récentes, comme l'imagerie optique de Sentinel-2, pour l'évaluation de la qualité de l'eau, les outils et méthodes développés dans la présente thèse peuvent être adaptés à d'autres capteurs (spatiaux ou aériens) et à d'autres paramètres. À cet égard, des capteurs à plus haute résolution spectrale, tels que ceux de la mission Sentinel-3, pourraient être utilisés pour améliorer la précision de l'extraction des paramètres de qualité de l'eau dans les grandes masses d'eau. De même, une résolution spatiale plus élevée, comme le capteur Planetscope de 3 mètres, pourrait être appliquée pour évaluer les

changements dans les canaux de rivière étroits ou intermittents (Tayer et al., 2022, en cours de révision).

En résumé, l'utilisation de la télédétection dans le domaine des ressources en eau est vaste et doit être davantage développé. Dans ce contexte, le travail actuel représente un pas en avant, soulignant la faisabilité de l'utilisation de techniques de science des données pour tirer parti de la télédétection en tant qu'outil viable pour le complément de mesures de terrain éparses et l'évaluation de masses d'eau qu'il serait impossible ou non viable de surveiller conventionnellement.

References

- Abrahart, R.J., Kneale, P.E., See, L.M., 2004. Neural networks for hydrological modelling. Balkema Publ, Leiden.
- ANA, 2021a. Brazilian Water Resources Report - 2021. National Water and Sanitation Agency.
- ANA, 2021b. Subsídios técnicos à proposta de Declaração de situação crítica de escassez quantitativa dos recursos hídricos na Região Hidrográfica do Paraná, até 30 de novembro de 2021.
- ANA, 2019a. Sistema Nacional de Informações sobre Recursos Hídricos (SNIRH) [WWW Document]. URL <http://snirh-web-hml.azurewebsites.net/> (accessed 6.8.22).
- ANA, 2019b. Brazilian Water Bodies [WWW Document]. URL <https://metadados.snirh.gov.br/geonetwork/srv/eng/catalog.search#/metadata/7d054e5a-8cc9-403c-9f1a-085fd933610c> (accessed 3.7.22).
- ANA, 2018. Brazilian Water Resources Report - 2017 - Full Report. National Water and Sanitation Agency, Brasília.
- ANEEL, ANA, 2010. RESOLUÇÃO CONJUNTA Nº 3. Estabelecer as condições e os procedimentos a serem observados pelos concessionários e autorizados de geração de energia hidrelétrica para a instalação, operação e manutenção de estações hidrométricas visando ao monitoramento pluviométrico, limnimétrico, fluviométrico, sedimentométrico e de qualidade da água associado a aproveitamentos hidrelétricos. [WWW Document]. URL https://arquivos.ana.gov.br/infohidrologicas/cadastro/ResolucaoConjunta_n_003-2010.pdf (accessed 9.7.22).
- Ansper, A., Alikas, K., 2018. Retrieval of chlorophyll-a from Sentinel-2 MSI data for the European Union Water Framework Directive reporting purposes. *Remote Sensing* 11, 64. <https://doi.org/10.3390/rs11010064>
- Baetens, L., Desjardins, C., Hagolle, O., 2019. Validation of Copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. *Remote Sensing* 11, 433. <https://doi.org/10.3390/rs11040433>
- Baghdadi, N., Zribi, M. (Eds.), 2016. Optical remote sensing of land surfaces: techniques and methods, Remote sensing observations of continental surfaces set. ISTE Press Ltd. ; Elsevier Ltd, London : Oxford.
- Bangira, T., Alfieri, S.M., Menenti, M., van Niekerk, A., 2019. Comparing thresholding with machine learning classifiers for mapping complex water. *Remote Sensing* 11, 1351. <https://doi.org/10.3390/rs11111351>
- Barbosa, C.C.F., Novo, E.M.L. de M., Martins, V.S., 2019. Introdução Ao Sensoriamento Remoto De Sistemas Aquáticos: Princípios E Aplicações. Instituto Nacional de Pesquisas Espaciais.
- Bi, S., Li, Y., Xu, Jie, Liu, G., Song, K., Mu, M., Lyu, H., Miao, S., Xu, Jiafeng, 2019. Optical classification of inland waters based on an improved Fuzzy C-Means method. *Opt. Express* 27, 34838. <https://doi.org/10.1364/OE.27.034838>
- Buma, W.G., Lee, S.-I., Seo, J.Y., 2018. Recent surface water extent of Lake Chad from multispectral sensors and grace. *Sensors* 18, 2082. <https://doi.org/10.3390/s18072082>
- Cairo, C., Barbosa, C., Lobo, F., Novo, E., Carlos, F., Maciel, D., Flores Júnior, R., Silva, E., Curtarelli, V., 2020. Hybrid Chlorophyll-a Algorithm for Assessing Trophic States of a

- Tropical Brazilian Reservoir Based on MSI/Sentinel-2 Data. *Remote Sensing* 12, 40. <https://doi.org/10.3390/rs12010040>
- Carlson, R., Simpson, J., 1996. North American Lake Management Society February, 1996 92.
- Chami, M., Lafrance, B., Fougnie, B., Chowdhary, J., Harmel, T., Waquet, F., 2015. OSOAA: a vector radiative transfer model of coupled atmosphere-ocean system for a rough sea surface application to the estimates of the directional variations of the water leaving reflectance to better process multi-angular satellite sensors data over the ocean. *Opt. Express* 23, 27829. <https://doi.org/10.1364/OE.23.027829>
- Chen, S., Han, L., Chen, X., Li, D., Sun, L., Li, Y., 2015. Estimating wide range Total Suspended Solids concentrations from MODIS 250-m imageries: An improved method. *ISPRS Journal of Photogrammetry and Remote Sensing* 99, 58–69. <https://doi.org/10.1016/j.isprsjprs.2014.10.006>
- Condé, R. de C., Martinez, J.-M., Pessotto, M.A., Villar, R., Cochonneau, G., Henry, R., Lopes, W., Nogueira, M., 2019. Indirect assessment of sedimentation in hydropower dams using MODIS remote sensing images. *Remote Sensing* 11, 314. <https://doi.org/10.3390/rs11030314>
- Cordeiro, M.C.R., Martinez, J.-M., Peña-Luque, S., 2021. Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors. *Remote Sensing of Environment* 253, 112209. <https://doi.org/10.1016/j.rse.2020.112209>
- D. N. Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, T. L. Veith, 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE* 50, 885–900. <https://doi.org/10.13031/2013.23153>
- Delegido, J., Tenjo, C., Ruiz-Verdú, A., Peña, R., Moreno, J., 2014. Modelo empírico para la determinación de clorofila-a en aguas continentales a partir de los futuros Sentinel-2 y 3. Validación con imágenes HICO. *Rev. Teledetec.* 0, 37. <https://doi.org/10.4995/raet.2014.2295>
- Dogliotti, A.I., Ruddick, K.G., Nechad, B., Doxaran, D., Knaeps, E., 2015. A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment* 156, 157–168. <https://doi.org/10.1016/j.rse.2014.09.020>
- Doxaran, D., Froidefond, J.-M., Lavender, S., Castaing, P., 2002. Spectral signature of highly turbid waters Application with SPOT data to quantify suspended particulate matter concentrations. *Remote Sensing of Environment* 13.
- Espinoza Villar, R., Martinez, J.-M., Le Texier, M., Guyot, J.-L., Fraizy, P., Meneses, P.R., Oliveira, E. de, 2013. A study of sediment transport in the Madeira River, Brazil, using MODIS remote-sensing images. *Journal of South American Earth Sciences, Hydrology, Geochemistry and Dynamic of South American Great River Systems* 44, 45–54. <https://doi.org/10.1016/j.jsames.2012.11.006>
- Feyisa, G.L., Meilby, H., Fensholt, R., Proud, S.R., 2014. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sensing of Environment* 140, 23–35. <https://doi.org/10.1016/j.rse.2013.08.029>
- Frampton, W.J., Dash, J., Watmough, G., Milton, E.J., 2013. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing* 82, 83–92. <https://doi.org/10.1016/j.isprsjprs.2013.04.007>

- Géron, A., 2017. Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems, First edition. ed. O'Reilly Media, Beijing ; Boston.
- Giglio, L., Desloîtres, J., Justice, C.O., Kaufman, Y.J., 2003. An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment* 87, 273–282. [https://doi.org/10.1016/S0034-4257\(03\)00184-6](https://doi.org/10.1016/S0034-4257(03)00184-6)
- Gleick, P.H., Pacific Institute for Studies in Development, Environment, and Security, Stockholm Environment Institute (Eds.), 1993. *Water in crisis: a guide to the world's fresh water resources*. Oxford University Press, New York.
- Google Developers, n.d. Sentinel-2 Cloud Masking with s2cloudless | Google Earth Engine [WWW Document]. Google Developers. URL <https://developers.google.com/earth-engine/tutorials/community/sentinel-2-s2cloudless> (accessed 9.6.22).
- Guo, H., Goodchild, M.F., Annoni, A. (Eds.), 2020. *Manual of Digital Earth*. Springer Singapore, Singapore. <https://doi.org/10.1007/978-981-32-9915-3>
- Hagolle, O., Huc, M., Pascual, D.V., Dedieu, G., 2010. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VEN μ S, LANDSAT and SENTINEL-2 images. *Remote Sensing of Environment* 114, 1747–1755. <https://doi.org/10.1016/j.rse.2010.03.002>
- Han, B., Loisel, H., Vantrepotte, V., Mériaux, X., Bryère, P., Ouillon, S., Dessailly, D., Xing, Q., Zhu, J., 2016. Development of a Semi-Analytical Algorithm for the Retrieval of Suspended Particulate Matter from Remote Sensing over Clear to Very Turbid Waters. *Remote Sensing* 8, 211. <https://doi.org/10.3390/rs8030211>
- Harmel, T., Chami, M., Tormos, T., Reynaud, N., Danis, P.-A., 2018. Sun glint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. *Remote Sensing of Environment* 204, 308–321. <https://doi.org/10.1016/j.rse.2017.10.022>
- Harmel, T., Gilerson, A., Tonizzo, A., Chowdhary, J., Weidemann, A., Arnone, R., Ahmed, S., 2012. Polarization impacts on the water-leaving radiance retrieval from above-water radiometric measurements. *Appl. Opt., AO* 51, 8324–8340. <https://doi.org/10.1364/AO.51.008324>
- Hoeser, T., Kuenzer, C., 2020. Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review-Part I: Evolution and Recent Trends. *Remote Sensing* 12, 1667. <https://doi.org/10.3390/rs12101667>
- Huffman, G.J., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J., Xie, P., David T Bolvin, Braithwaite, D., 2019. *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG)*.
- Jackson, T., Sathyendranath, S., Mélin, F., 2017. An improved optical classification scheme for the Ocean Colour Essential Climate Variable and its applications. *Remote Sensing of Environment, Earth Observation of Essential Climate Variables* 203, 152–161. <https://doi.org/10.1016/j.rse.2017.03.036>
- Jia, T., Zhang, Y., Dong, R., 2021. A Universal Fuzzy Logic Optical Water Type Scheme for the Global Oceans. *Remote Sensing* 13, 4018. <https://doi.org/10.3390/rs13194018>
- Kairu, E.N., 1982. An introduction to remote sensing. *GeoJournal* 6. <https://doi.org/10.1007/BF00210657>
- Khandelwal, A., Karpatne, A., Marlier, M.E., Kim, J., Lettenmaier, D.P., Kumar, V., 2017. An approach for global monitoring of surface water extent variations in reservoirs using MODIS data. *Remote Sensing of Environment, Big Remotely Sensed Data: tools, applications and experiences* 202, 113–128. <https://doi.org/10.1016/j.rse.2017.05.039>

- Kirk, J.T.O., 1994. *Light and Photosynthesis in Aquatic Ecosystems*, 2nd ed. Cambridge University Press, Cambridge. <https://doi.org/10.1017/CBO9780511623370>
- Kordelas, G.A., Manakos, I., Aragonés, D., Díaz-Delgado, R., Bustamante, J., 2018. Fast and automatic data-driven thresholding for inundation mapping with Sentinel-2 data. *Remote Sensing* 10, 910. <https://doi.org/10.3390/rs10060910>
- Kordelas, G.A., Manakos, I., Lefebvre, G., Poulin, B., 2019. Automatic inundation mapping using Sentinel-2 data applicable to both Camargue and Doñana biosphere reserves. *Remote Sensing* 11, 2251. <https://doi.org/10.3390/rs11192251>
- Lary, D.J., Alavi, A.H., Gandomi, A.H., Walker, A.L., 2016. Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, Special Issue: Progress of Machine Learning in Geosciences 7, 3–10. <https://doi.org/10.1016/j.gsf.2015.07.003>
- Lima, T.A., Beuchle, R., Langner, A., Grecchi, R.C., Griess, V.C., Achard, F., 2019. Comparing Sentinel-2 MSI and Landsat 8 OLI Imagery for Monitoring Selective Logging in the Brazilian Amazon. *Remote Sensing* 11, 961. <https://doi.org/10.3390/rs11080961>
- Lins, R.C., Martinez, J.-M., Motta Marques, D.D., Cirilo, J.A., Fragoso, C.R., 2017. Assessment of chlorophyll-a remote sensing algorithms in a productive tropical estuarine-lagoon system. *Remote Sensing* 9, 516. <https://doi.org/10.3390/rs9060516>
- Long, C.M., Pavelsky, T.M., 2013. Remote sensing of suspended sediment concentration and hydrologic connectivity in a complex wetland environment. *Remote Sensing of Environment* 129, 197–209. <https://doi.org/10.1016/j.rse.2012.10.019>
- Lubac, B., Loisel, H., 2007. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Remote Sensing of Environment* 110, 45–58. <https://doi.org/10.1016/j.rse.2007.02.012>
- Makarigakis, A., Jimenez-Cisneros, B., 2019. UNESCO's Contribution to Face Global Water Challenges. *Water* 11, 388. <https://doi.org/10.3390/w11020388>
- Makarigakis, A.K., Jimenez-Cisneros, B.E., 2019. UNESCO's contribution to face global water challenges. *Water* 11, 388. <https://doi.org/10.3390/w11020388>
- Markert, K.N., Chishtie, F., Anderson, E.R., Saah, D., Griffin, R.E., 2018. On the merging of optical and SAR satellite imagery for surface water mapping applications. *Results in Physics* 9, 275–277. <https://doi.org/10.1016/j.rinp.2018.02.054>
- Martinez, J.M., Guyot, J.L., Filizola, N., Sondag, F., 2009. Increase in suspended sediment discharge of the Amazon River assessed by monitoring network and satellite data. *CATENA* 79, 257–264. <https://doi.org/10.1016/j.catena.2009.05.011>
- Martinis, S., Twele, A., Voigt, S., 2011. Unsupervised extraction of flood-induced backscatter changes in sar data using Markov image modeling on irregular graphs. *IEEE Transactions on Geoscience and Remote Sensing* 49, 251–263. <https://doi.org/10.1109/TGRS.2010.2052816>
- Matsushita, B., Yang, W., Yu, G., Oyama, Y., Yoshimura, K., Fukushima, T., 2015. A hybrid algorithm for estimating the chlorophyll-a concentration across different trophic states in Asian inland waters. *ISPRS Journal of Photogrammetry and Remote Sensing* 102, 28–37. <https://doi.org/10.1016/j.isprsjprs.2014.12.022>
- Mélin, F., Vantrepotte, V., Clerici, M., D'Alimonte, D., Zibordi, G., Berthon, J.-F., Canuti, E., 2011. Multi-sensor satellite time series of optical properties and chlorophyll-a concentration in the Adriatic Sea. *Progress in Oceanography* 91, 229–244. <https://doi.org/10.1016/j.pocean.2010.12.001>
- Melo, D. de C.D., Scanlon, B.R., Zhang, Z., Wendland, E., Yin, L., 2016. Reservoir storage and hydrologic responses to droughts in the Paraná River basin, south-eastern Brazil. *Hydrol. Earth Syst. Sci.* 20, 4673–4688. <https://doi.org/10.5194/hess-20-4673-2016>

- Mercan, S., Alam, M.S., 2011. Anomaly detection in hyperspectral imagery using stable distribution, in: Sadjadi, F.A., Mahalanobis, A. (Eds.), . Presented at the SPIE Defense, Security, and Sensing, Orlando, Florida, United States, p. 80490V. <https://doi.org/10.1117/12.884913>
- Mitchell, T.M., 1997. Machine Learning, McGraw-Hill series in computer science. McGraw-Hill, New York.
- Mobley, C.D., 1999. Estimation of the remote-sensing reflectance from above-surface measurements. *Appl. Opt.* 38, 7442. <https://doi.org/10.1364/AO.38.007442>
- Moore, G.F., Aiken, J., Lavender, S.J., 1999. The atmospheric correction of water colour and the quantitative retrieval of suspended particulate matter in Case II waters: Application to MERIS. *International Journal of Remote Sensing* 20, 1713–1733. <https://doi.org/10.1080/014311699212434>
- Moore, T.S., Campbell, J.W., Dowell, M.D., 2009. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sensing of Environment* 113, 2424–2430. <https://doi.org/10.1016/j.rse.2009.07.016>
- Moore, T.S., Dowell, M.D., Bradt, S., Ruiz Verdu, A., 2014. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Remote Sensing of Environment* 143, 97–111. <https://doi.org/10.1016/j.rse.2013.11.021>
- Morel, A., Prieur, L., 1977. Analysis of variations in ocean color1. *Limnology and Oceanography* 22, 709–722. <https://doi.org/10.4319/lo.1977.22.4.0709>
- Mueller-Wilm, U., Devignot, O., Pessiot, L., 2019. Sen2cor configuration and user manual.
- Nechad, B., Ruddick, K.G., Park, Y., 2010. Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment* 114, 854–866. <https://doi.org/10.1016/j.rse.2009.11.022>
- Neil, C., Spyrakos, E., Hunter, P.D., Tyler, A.N., 2019. A global approach for chlorophyll-a retrieval across optically complex inland waters based on optical water types. *Remote Sensing of Environment* 229, 159–178. <https://doi.org/10.1016/j.rse.2019.04.027>
- Novoa, S., Doxaran, D., Ody, A., Vanhellemont, Q., Lafon, V., Lubac, B., Gernez, P., 2017. Atmospheric Corrections and Multi-Conditional Algorithm for Multi-Sensor Remote Sensing of Suspended Particulate Matter in Low-to-High Turbidity Levels Coastal Waters. *Remote Sensing* 9, 61. <https://doi.org/10.3390/rs9010061>
- Olsson, F., Mackay, E.B., Barker, P., Davies, S., Hall, R., Spears, B., Exley, G., Thackeray, S.J., Jones, I.D., 2022. Can reductions in water residence time be used to disrupt seasonal stratification and control internal loading in a eutrophic monomictic lake? *Journal of Environmental Management* 304, 114169. <https://doi.org/10.1016/j.jenvman.2021.114169>
- Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M.K., Ondrusek, M., Oppelt, N., Stumpf, R., 2020. Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. *Remote Sensing of Environment* 240, 111604. <https://doi.org/10.1016/j.rse.2019.111604>
- Pekel, J.-F., Cottam, A., Gorelick, N., Belward, A.S., 2016. High-resolution mapping of global surface water and its long-term changes. *Nature* 540, 418–422. <https://doi.org/10.1038/nature20584>
- Peña-Luque, S., Ferrant, S., Cordeiro, M., Ledauphin, T., Maxant, J., Martinez, J., 2021. Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level. *Remote Sensing* 13, 3279. <https://doi.org/10.3390/rs13163279>
- Pinet, S., Martinez, J.-M., Ouillon, S., Lartiges, B., Villar, R.E., 2017. Variability of apparent and inherent optical properties of sediment-laden waters in large river basins – lessons

- from in situ measurements and bio-optical modeling. *Opt. Express*, OE 25, A283–A310. <https://doi.org/10.1364/OE.25.00A283>
- Qiu, S., Zhu, Z., He, B., 2019. Fmask 4.0: Improved cloud and cloud shadow detection in Landsats 4–8 and Sentinel-2 imagery. *Remote Sensing of Environment* 231, 111205. <https://doi.org/10.1016/j.rse.2019.05.024>
- Rocha, P.C., de Araújo, A.P., 2011. O REGIME HIDROLÓGICO NA BACIA DO RIO PARANAPANEMA: variabilidade interanual e espacial 14.
- Rozante, J.R., Moreira, D.S., Goncalves, L.G.G. de, Vila, D.A., 2010. Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America. *Weather and Forecasting* 25, 885–894. <https://doi.org/10.1175/2010WAF2222325.1>
- Santos, A.L.M.R. dos, Martinez, J.M., Filizola, N.P., Armijos, E., Alves, L.G.S., 2018. Purus River suspended sediment variability and contributions to the Amazon River from satellite data (2000–2015). *Comptes Rendus Geoscience* 350, 13–19. <https://doi.org/10.1016/j.crte.2017.05.004>
- Smith, M.E., Robertson Lain, L., Bernard, S., 2018. An optimized Chlorophyll a switching algorithm for MERIS and OLCI in phytoplankton-dominated waters. *Remote Sensing of Environment* 215, 217–227. <https://doi.org/10.1016/j.rse.2018.06.002>
- Souza, C., Kirchhoff, F., Oliveira, B., Ribeiro, J., Sales, M., 2019. Long-Term Annual Surface Water Change in the Brazilian Amazon Biome: Potential Links with Deforestation, Infrastructure Development and Climate Change. *Water* 11, 566. <https://doi.org/10.3390/w11030566>
- SPR/ANA, 2020. Atualização da Base de Dados Nacional de Referência de Massas d'Água.
- Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M.W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y., Tyler, A.N., 2018. Optical types of inland and coastal waters. *Limnology and Oceanography* 63, 846–870. <https://doi.org/10.1002/lno.10674>
- SRHU-MMA, 2006. Caderno da Região Hidrográfica do Paraná. MMA, Brasília.
- Suhet, Hoersch, B., 2015. Sentinel-2 User Handbook.
- Tayer, T.C., Douglas, M.M., Cordeiro, M.C.R., Tayer, A.D.N., Callow, J.N., Beesley, L., McFarlane, D., 2022. Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, Planetscope and Sharpened imagery [Submitted for publication] *GIScience & Remote Sensing*, 27.
- Toming, K., Kutser, T., Laas, A., Sepp, M., Paavel, B., Nõges, T., 2016. First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery. *Remote Sensing* 8, 640. <https://doi.org/10.3390/rs8080640>
- Topp, S.N., Pavelsky, T.M., Jensen, D., Simard, M., Ross, M.R.V., 2020. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. *Water* 12, 169. <https://doi.org/10.3390/w12010169>
- Tottrup, C., Druce, D., Meyer, R.P., Christensen, M., Riffler, M., Dulleck, B., Rastner, P., Jupova, K., Sokoup, T., Haag, A., Cordeiro, M.C.R., Martinez, J.-M., Franke, J., Schwarz, M., Vanthof, V., Liu, S., Zhou, H., Marzi, D., Rudyanto, R., Thompson, M., Hiestermann, J., Alemohammad, H., Masse, A., Sannier, C., Wangchuk, S., Schumann, G., Giustarini, L., Hallowes, J., Markert, K., Paganini, M., 2022. Surface Water Dynamics from Space: A Round Robin Intercomparison of Using Optical and SAR High-Resolution Satellite

- Observations for Regional Surface Water Detection. *Remote Sensing* 14, 2410. <https://doi.org/10.3390/rs14102410>
- UN, U.N., 2018. World Urbanization Prospects [WWW Document]. World Urbanization Prospects. URL <https://population.un.org/wup/Download/> (accessed 7.6.22).
- UNDP, 2022a. New threats to human security in the Anthropocene. United Nations Development Programme, New York, NY.
- UNDP, 2022b. Sustainable Development Goals [WWW Document]. UNDP. URL <https://www.undp.org/sustainable-development-goals> (accessed 5.12.22).
- UNDP (Ed.), 2020. The next frontier: human development and the Anthropocene, Human development report. United Nations Development Programme, New York, NY.
- UNESCO-iWSSM, 2019. Water security and the sustainable development goals (Series I), Global Water Security Issues (GWSI) Series. UNESCO Publishing, Paris.
- UNESCO-WWAP, 2021. World Water Development Report 2021: Valuing Water. UNITED NATIONS EDUCATIONAL, S.I.
- Uudeberg, K., Aavaste, A., Kõks, K.-L., Ansper, A., Uusõue, M., Kangro, K., Ansko, I., Ligi, M., Toming, K., Reinart, A., 2020. Optical Water Type Guided Approach to Estimate Optical Water Quality Parameters. *Remote Sensing* 12, 931. <https://doi.org/10.3390/rs12060931>
- Vanderplas, J.T., 2016. Python data science handbook: essential tools for working with data, First edition. ed. O'Reilly Media, Inc, Sebastopol, CA.
- Vantrepotte, V., Loisel, H., Dessailly, D., Mériaux, X., 2012. Optical classification of contrasted coastal waters. *Remote Sensing of Environment* 123, 306–323. <https://doi.org/10.1016/j.rse.2012.03.004>
- Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, İ., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., SciPy 1.0 Contributors, 2020. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Wieland, M., Martinis, S., 2019. A modular processing chain for automated flood monitoring from multi-spectral satellite data. *Remote Sensing* 11, 2330. <https://doi.org/10.3390/rs11192330>
- Woźniak, S.B., Stramski, D., 2004. Modeling the optical properties of mineral particles suspended in seawater and their influence on ocean reflectance and chlorophyll estimation from remote sensing algorithms. *Appl. Opt.*, AO 43, 3489–3503. <https://doi.org/10.1364/AO.43.003489>
- Xue, K., Ma, R., Wang, D., Shen, M., 2019. Optical Classification of the Remote Sensing Reflectance and Its Application in Deriving the Specific Phytoplankton Absorption in Optically Complex Lakes. *Remote Sensing* 11, 184. <https://doi.org/10.3390/rs11020184>
- Yadav, S., Yamashiki, Y., Susaki, J., Yamashita, Y., Ishikawa, K., 2019. Chlorophyll estimation of lake water and coastal water using Landsat-8 and Sentinel-2A satellite. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLII-3/W7, 77–82. <https://doi.org/10.5194/isprs-archives-XLII-3-W7-77-2019>
- Yepez, S., Laraque, A., Martinez, J.-M., De Sa, J., Carrera, J.M., Castellanos, B., Gallay, M., Lopez, J.L., 2018. Retrieval of suspended sediment concentrations using Landsat-8 OLI satellite images in the Orinoco River (Venezuela). *Comptes Rendus Geoscience, Rivers of the Andes and the Amazon Basin: Deciphering global change from the hydroclimatic variability in the critical zone* 350, 20–30. <https://doi.org/10.1016/j.crte.2017.08.004>

- Zhu, Z., Woodcock, C.E., 2012. Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment* 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>
- Zupanc, A., 2020. Improving Cloud Detection with Machine Learning. Sentinel Hub Blog. URL <https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13> (accessed 3.8.22).

List of Figures

Figure 1: (a) Remote sensing schema. (b) Spectral signatures for distinct natural targets. Extracted from Mercan and Alam (2011).	18
Figure 2: The mean spectral profiles of several rivers/basins considered in the present study measured with field radiometers.	20
Figure 3: Spectral response functions for Sentinel-2/MSI, Visual-NIR, and SWIR bands and their corresponding spatial resolutions.	22
Figure 4: Schematic comparison of traditional programming approach (a) and machine learning approach (b). Adapted from Géron (2017).	25
Figure 5: Scatter plot of the Normalized Difference Water Index (NDWI, Y-axis) and SWIR reflectance (X-axis) in sr^{-1} for several pixels in a Sentinel-2 scene, showing clustering results in different colors.	27
Figure 6: Schematic view of the steps involved in assessing inland water quality from satellite remote sensed imagery that were developed in this thesis. Boxes in orange represent steps covered in axis 1 (water detection), boxes in blue represent axis 2 (water classification) and green boxes represent axis 3 (generalization and analysis).	32
Figure 7: (a) Principe général de la télédétection passive, (b) Signatures spectrales de différentes cibles naturelles. Extrait de Mercan et Alam (2011).....	46
Figure 8: Les profils spectraux moyens de plusieurs rivières/bassins considérés dans la présente étude mesurés avec des radiomètres de terrain.	48
Figure 9: Fonctions de réponse spectrale pour les bandes Sentinel-2/MSI, Visual-NIR et SWIR et leurs résolutions spatiales correspondantes.	50

Figure 10: Comparaison schématique de l'approche de programmation traditionnelle (a) et de l'approche d'apprentissage automatique (b). Adapté de Géron (2017).....	54
Figure 11: Diagramme de dispersion de l'indice de différence normalisé de l'eau (NDWI, axe Y) et de la réflectance SWIR (axe X) en sr^{-1} pour les pixels d'une scène Sentinel-2, montrant les résultats du regroupement avec différentes couleurs.	56
Figure 12: Vue schématique des étapes de l'évaluation de la qualité des eaux intérieures à partir d'images de télédétection par satellite qui ont été développées dans cette thèse. Les cases en orange représentent les étapes couvertes par l'axe 1 (détection de l'eau), les cases en bleu représentent l'axe 2 (classification de l'eau) et les cases vertes représentent l'axe 3 (généralisation et analyse).....	61
Figure I-1: Example of study conducted using the waterdetect algorithm on the lake Chad region. (a) monthly water surface in km^2 for each considered tile individually; (b) lake Chad water mask.	74
Figure I-2: Overview of the main sites where the waterdetect algorithm has been successfully applied. Background image: “World Map Blue” licensed under Creative Commons license.....	75
Figure II-1: Map of the locations that were sampled for both radiometric above water measurements and surface water collection.	100
Figure II-2: SPM concentration range box plots displayed as a function of the watershed where radiometric and water samples were collected.	101
Figure II-3: Schematic representation of the above-water reflectance measurements that was used for all the sampling points analyzed in this study.....	102
Figure II-4: Spectra curves and mean reflectance for each Area within the dataset. Red lines represent the mean reflectance in sr^{-1}	104

Figure II-5: Example of spectra normalization in 3 reflectance spectra from the Madeira river. Original spectra on the left and normalized spectra on the right. SPM concentration were of 4mg/l (blue), 75.8mg/l (red) and 286mg/l (green). 106

Figure II-6: Spectral response functions for Sentinel-2 Visual-NIR and SWIR bands. ..109

Figure II-7: Schematic view of the steps involved in the OWT definition assessment for different scenarios considering different input features (blue box) and model inversion types (yellow box). 114

Figure II-8: R_{RS} spectra used in this study classified by Optical Water Type (OWT) as proposed by Spyarakos et al. (2019). Red lines represent cluster means for each OWT. 116

Figure II-9: Overall RMSLE errors for hybrid model calibrated for high and low concentrations, according to different “proxy” bands and threshold values for three different spectral bands: 665, 740 e 865 nm. Results obtained from the calibration dataset..... 118

Figure II-10: Baseline retrieval model: 2-Class Optical Water Types (Low/High SPM range) (a) Mean reflectance spectra for each group; (b) SPM range for each group; (c) Actual SPM (X axis) vs Predicted SPM (Y axis) graph in mg/l. The colors for panels a, b, and c represent the corresponding the low SPM (Red) or high SPM (Blue) concentration group. 123

Figure II-11: Baseline retrieval model: Watershed-based OWTs (a) Mean raw reflectance for each considered watershed; (b) Mean normalized reflectance for each considered watershed; (c) Actual vs prediction SPM concentrations in mg/l considering one algorithm for each watershed. The colors represent different watersheds..... 124

Figure II-12: Baseline retrieval model: Spyarakos’ OWTs (a) Mean raw reflectance for each considered OWT; (b) SPM range for each OWT; (c) Actual vs prediction SPM concentrations in mg/l considering one algorithm for each OWT. The colors represent different classes. 130

Figure II-13: Performance of the assignment accuracy (Euclidean distance) for different input features offering decreasing spectral resolutions. Cluster definition was realized using k-means and full resolution hyperspectral Rrs dataset.132

Figure II-14: Performance of the assignment accuracy as a function of distance metric used (Euclidean and Mahalanobis distances) for different input features offering decreasing spectral resolutions. Cluster definition was achieved using k-means applied to Sentinel-2/MSI spectral bands and for a varying number of clusters, from 1 to 10.133

Figure II-15: Performance of the assignment accuracy as a function of distance metric used (Euclidean and Mahalanobis distances) for different input features offering decreasing spectral resolutions. Cluster definition was achieved using Agglomerative Clustering applied to Sentinel-2/MSI spectral bands and for a varying number of clusters, from 1 to 10.134

Figure II-16: Impact of the clustering step on SPM retrieval as a function of the number of OWTs and for three different algorithms: K-means, FCM and Agglomerative clustering. Retrieval performance is assessed using RMSLE on the R_{RS} validation dataset. The input features were R_{RS} sampled to Sentinel2 MSI spectral bands for both clustering and assignment (Mahalanobis distance).135

Figure II-17: Impact of the clustering step on SPM retrieval as a function of the number of OWTs and for three different algorithms: K-means, Fuzzy C-means and Agglomerative clustering. Retrieval performance is assessed using RMSLE on the R_{RS} validation dataset. The input features were normalized R_{RS} sampled to Sentinel2 MSI spectral bands for both clustering and assignment (Mahalanobis distance).136

Figure II-18: Impact of the clustering step on SPM retrieval as a function of the number of OWTs and for R_{RS} input features sampled at decreasing spectral resolution: 10nm, 20nm, 30nm, Sentinel-2 and Sentinel-3 spectral bands. For this test, clustering used K-means algorithm while

the assignment step was based on Euclidean distance and Sentinel-2 spectral resolution as input features. 136

Figure II-19: optimal 4-OWT clustering for SPM concentration retrieval (a) Mean Rrs for each considered OWT; (b) SPM range for each OWT; (c) Actual vs prediction SPM concentrations in mg/l considering one algorithm for each OWT (Table II-9). The colors represent the four different classes. 141

Figure II-20: Suspended sediment particle diameter (D50) distribution for the four optimal OWTs configuration, expressed in μm 144

Figure II-21: Geographical distribution of the OWTs (“clusters”) accordingly to the 13 hydrological regions. OWT frequency was normalized to 100 % for each region. 145

Figure III-1: Schematic diagram of the water detection process. Blue boxes represent different threads. Gray boxes group the threads that run in parallel. The orange box represents a process that runs entirely in the cloud. 151

Figure III-2: Examples of the reports saved for each Sentinel-2 images during the water detection step. (a) Scene overviews with true color (RGB) image, final water mask, no/masked data, sunglint probability and the MNDWI and NDWI indices. (b) NDWI x MNDWI and MNDWI x B12 scatter plots with final clustering results. 152

Figure III-3: Schematic representation of angles and geometry between light source, target and viewing sensor. 153

Figure III-4: Effect of the sunglint in a Sentinel-2 image on the surface of the Lake of Cuitzeo, Mexico. 154

Figure III-5: Representation of the glint angles in the 5km grid cells. The lower the sunglint angle (yellow color), the higher probability of sunglint in the region. Angle values higher than 25° are considered exempt from sunglint (dark colors)..... 156

Figure III-6: Representation of the cloud shadow matching algorithm used to identify shadows on the ground.	158
Figure III-7: (a) Schematic view of the monthly water masks process creation for each tile. (b) Schematic view of the monthly masks combination into a single mosaic.	160
Figure III-8: Water surface extent mosaic for January 2018. In red, the pixels that remained covered by clouds or without valid data during the whole month.	161
Figure III-9: Monthly water surface estimation in the area of interest and the percentage of invalid pixels due to cloud persistence.	162
Figure III-10: Schematic representation of the pixel level interpolation along the time axis.	163
Figure III-11: (a) Mean value for each coarse pixel, calculated through time dimension, representing percentage of surface water over 40,000ha area; (b) standard deviation for each coarse pixel.	165
Figure III-12: (a) In the left panel, the hydrographic network of the area of interest. (b) In the right panel, the anomaly map produced for November 2021.	165
Figure III-13: Example of the stratified analysis for an unnamed water body. (a) Shape file with its area and coordinates; (b) high resolution imagery; and (c) overlaid water mask for June 2020.	167
Figure III-14: Reservoirs' statistics (quantity and total area) grouped by category for the area of interest.	167
Figure III-15: (a) Parana Hydrographic Region in the national context; (b) Regions' water bodies and reservoirs being monitored by ANA. The black grid represents the region of interest selected for the current study and the Sentinel-2 tiles considered.	173

Figure III-16: Schematic flowchart of the main steps, data and methods used in the current study.	175
Figure III-17: Schematic diagram of the water detection process. Blue boxes represent different threads. Gray boxes group the threads that run in parallel. The orange box represents a process that runs entirely in the cloud.	179
Figure III-18: Schematic view of the surface water estimation process steps. (a) Monthly water masks combination; (b) monthly masks combination into mosaics; (c) time interpolation.	181
Figure III-19: (a) Monthly water surface area time-series obtained from field measurements (red line) and from the satellite imagery (blue line) for the 16 selected reservoirs, total volume as measured in the field (green line) and errors between field and satellite measurements (purple bars); (b) Correlation between actual field measured area and estimated remote sensing area.	186
Figure III-20: Overall water surface extents (lower panel) in the study area. Red dot indicates when ANA raised the drought alert in the basin. In the upper panel, red bars indicate rainfall below average and blue bars indicate rainfall above average, according to MERGE product. The dashed line represents the average rainfall computed between 2000-2019.	190
Figure III-21: Water surface anomaly maps, represented in percentage of change. (a) April's 2020 anomaly compared to the mean; (b) April's 2021 anomaly compared to the mean of 2018 and 2019; (c) November 2020 compared to April 2020; (d) November 2021 compared to November 2020.	191
Figure III-23: Overall water surface extents (lower panel) in the Paranapanema basin. In the upper panel, red bars indicate rainfall below average and blue bars indicate rainfall above average, according to MERGE product. The dashed line represents the average rainfall computed between 2000-2019.	193

Figure III-22: Water surface variation, measured in percentage, in comparison to the maximum extents for the water bodies grouped by size.	193
Figure III-24: Comparison between remote sensing derived turbidity (blue lines) and field measurements (red dots) for Jurumirim, Taciba and Ourinhos stations in the Paranapanema basin.	194
Figure III-25: (a) Comparison between field turbidity (x axis) and retrieved remote sensing turbidity (y axis), for Jurumirim, Taciba and Ourinhos stations; (a) Comparison between field Chl-a (x axis) and retrieved remote sensing Chl-a (y axis), for Jurumirim, Canoas I and Canoas II stations. Mashup performed considering a maximum data lag of 10 days.	195
Figure III-26: Comparison between remote sensing derived Chl-a (blue lines) and field measurements (red dots) for Jurumirim, Canoas I and Canoas II stations in the Paranapanema basin.	196
Figure III-27: Chl-a concentration maps for the Paranapanema region. (a) monthly composite for 04/2018; and (b) monthly composite for 11/2020.....	198
Figure III-28: Turbidity concentration maps for the Paranapanema region. (a) monthly composite for 04/2018; and (b) monthly composite for 11/2020.....	199
Figure III-29: Chl-a histograms for the Paranapanema region of interest. Annotated values and color scale indicate the median Chl-a value in the month.....	201
Figure III-30: Turbidity histograms for the Paranapanema region of interest. Annotated values and color scale indicate the median turbidity (FNU) value in the month.	202
Figure III-31: Crops from the Ourinhos station, on dates surrounding the turbidity peak value occurred on June 2, 2016. Top images represent L2A RGB crops and bottom images GRS corrected water bodies. Red dot marks the retrieval position.	207

Figure III-32: Crops from the Canoas I station, on dates surrounding the Chl-a peak value occurred on October 14, 2020. Top images represent L2A RGB crops and bottom images GRS corrected water bodies. Red dot marks the retrieval position.	207
Figure III-33: (a, c) Chl-a density histograms and (b, c) turbidity density histograms for the Paranapanema basin.	210
Figure III-34: Evolution in Chl-a during 2019s dry season for the following size classes: (a) Small and very small water bodies (center coordinates: 49.206W and 23.514S); (b) Couro do Boi reservoir (center coordinates: 51.164W and 23.132S); and Piraju hydroelectric dam (center coordinates: 49.334W and 23.173S).	211
Figure III-35: Chl-a histograms for the Paranapanema region of interest, without Barra Bonita reservoir. Annotated values and color scale indicate the median Chl-a value in the month.	212
Figure III-36: Turbidity levels for Capivara (a), Jurumirim (b) and Chavantes (c) reservoirs.	213
Figure IV-1: Schematic view of the steps involved in assessing inland water quality from satellite remote sensed imagery. Boxes in orange represent steps covered in axis 1 (water detection), boxes in blue represent axis 2 (water classification) and green boxes represent axis 3 (generalization and analysis). Names inside parentheses identify python packages developed in the context of the thesis.	218
Figure IV-2: Vue schématique des étapes de l'évaluation de la qualité des eaux intérieures à partir d'images de télédétection par satellite. Les cases en orange représentent les étapes couvertes par l'axe 1 (détection de l'eau), les cases en bleu représentent l'axe 2 (classification de l'eau) et les cases vertes représentent l'axe 3 (généralisation et analyse). Les noms entre parenthèses identifient les paquets python développés dans le cadre de la thèse.	228

List of Tables

Table II-1: Summary of data stored in the database and its water quality parameters. ...	101
Table II-2: Sentinel-2 spectral bands used in models' calibration.	108
Table II-3: Best fits for each group of measurements (low and high SPM concentrations). The combined metrics were calculated on the validation dataset.	120
Table II-4: Models for the each watershed. The metrics are calculated considering the points in the validation dataset.	125
Table II-5: Models for the Spyrakos' OWT classes. The metrics are calculated considering the points in the validation dataset.	126
Table II-6: SPM retrieval performance as a function of the number of OWTs. Accuracy was assessed on the validation dataset. OWT classification was realized using K-means algorithm and non-normalized Rrs, with spectral input features set to Sentinel-2 MSI spectral bands. Assignment used Euclidean distance and Sentinel-2 MSI spectral bands as input features.	137
Table II-7: R_{RS} values, expressed as mean and standard deviation and for each Sentinel-2 MSI spectral bands, for the four OWTs that represents the optimal water type classification for inland water SPM concentration retrieval.	139
Table II-8: Model type and accuracy for the 4-OWT classification retained as the most efficient spectral clustering for SPM retrieval. Models were calibrated using the training dataset and the accuracy metrics were calculated using the validation dataset.	139
Table II-9: Comparison of different classification schemas.	140
Table III-1: Reservoirs with area greater than 10,000ha, considered for the data validation. Reservoirs without stage curve were excluded from the analysis.	178

Table III-2: Stations used in water quality parameters validation.	179
Table III-3: Reservoirs' size classes considered for the stratified analysis.....	208

Annexes

- I. Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level

Article

Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level

Santiago Peña-Luque ^{1,*}, Sylvain Ferrant ², Maurício C. R. Cordeiro ^{3,4}, Thomas Ledauphin ⁵, Jerome Maxant ⁵ and Jean-Michel Martinez ⁴¹ Centre National d'Études Spatiales (CNES), 18 Avenue Edouard Belin, 31400 Toulouse, France² Centre d'Étude Spatiale de la Biosphère (CESBIO), UPS-Université de Toulouse, CNRS, CNES, BRD, INRAE, 18 Av. Edouard Belin, bpi 2801, 31401 Toulouse, France; sylvain.ferrant@cesbio.cnrs.fr³ Agência Nacional de Águas (ANA), Setor Policial Sul, Área 5, Quadra 3, Brasília 70610-200, Brazil; mauricio@ana.gov.br⁴ Géosciences Environnement Toulouse (GET), Unité Mixte de Recherche 5563, BRD/CNRS/Université, 31400 Toulouse, France; martinez@ird.fr⁵ ICUBE-SERTIT, Université de Strasbourg, 67412 Illkirch Graffenstaden, France; tledauphin@unistra.fr (T.L.); maxant@unistra.fr (J.M.)

* Correspondence: santiago.penalouque@cnrs.fr

Citation: Peña-Luque, S.; Ferrant, S.; Cordeiro, M.C.R.; Ledauphin, T.; Maxant, J.; Martinez, J.-M. Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level. *Remote Sens.* **2021**, *13*, 3279. <https://doi.org/10.3390/rs13163279>

Academic Editor: Carl J. Legleiter

Received: 10 June 2021

Accepted: 29 July 2021

Published: 19 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: Water stock monitoring is a major issue for society on a local and global scale. Sentinel-1&2 satellites provide frequent acquisitions to track water surface dynamics; proxy variables to enable water surface volume monitoring. How do we combine such observations along time for each sensor? What advantages and disadvantages of single-date, monthly or time-windowed estimations? In this context, we analysed the impact of merging information through different types and lengths of time-windows. Satellite observations were processed separately on optical (Sentinel-2) and radar (Sentinel-1) water detectors at 10m resolution. The analysis has been applied at two scales. First, validating with 26 large scenes (110 × 110 km) in different climatic zones in France, time-windows yielded an improvement on radar detection (F1-score improved from 0.72 to 0.8 for 30 days on average logic) while optical performances remained stable (F1-score 0.89). Second, validating reservoir area estimations with 29 instrumented reservoirs (20–1250 ha), time-windows presented in all cases an improvement on both optical and radar error for any window length (5–30 days). The mean relative absolute error in optical area detection improved from 16.9% on single measurements to 12.9% using 15 days time-windows, and from 22.15% to 15.1% in radar detection). Regarding reservoir filling rates, we identified an increased negative bias for both sensors when the reservoir is nearly full. This work helped to compare accuracies of separate optical and radar capabilities, where optical statistically outperforms radar at both local and large scale to the detriment of less frequent measurements. Furthermore, we propose a geomorphological indicator of reservoirs to predict the quality of radar area monitoring ($R^2 = 0.58$). In conclusion, we suggest the use of time-windows on operational water mapping or reservoir monitoring systems, using 10–20 days time-windows with average logic, providing more frequent and faster information to water managers in periods of crisis (e.g., water shortage) compared to monthly estimations.

Keywords: water cycle; water surfaces; reservoirs; multi-temporal; water detection; radar imaging; optical imaging; area monitoring

1. Introduction

Surface water dynamics are key variables to monitor the continental water stocks, themselves crucial to human societies at a regional and global scale. Large reservoirs are known to have a strong impact on hydrology, decreasing by around 2% the river discharge to oceans [1]. They also have considerable importance for agriculture, domestic,

and industrial water uses, considering that their annual storage capacity is equivalent to 10% of the annual soil water storage at the global scale [2]. In much larger numbers, smaller surface reservoirs ($<0.1 \text{ km}^2$) range from 0 to 10 reservoirs per km^2 depending on the region. Their densities increase with average precipitations, from 10 to $10^6 \text{ m}^3/\text{km}^2$. Their importance has been analysed in terms of stream discharge reduction, especially important during driest conditions, and seepage rates that appear to be higher than the evaporation rate in existing studies [3]. Their positive socio economic impacts are reported worldwide for the local populations [4,5]. In that context, there is a need to improve knowledge of large and small reservoir characteristics.

High-resolution remote sensing data are identified to be highly appropriate to map and quantify the properties of large and smaller reservoirs over vast areas. Recent releases of surface water changes derived from high-resolution satellites have opened new insights into global surface water dynamics. For instance, the global surface water database [6], based on Landsat program observations since the 80s, allows investigating into the long term and seasonal evolution of surface water area extents around the world, providing quantitative evidence of human or natural induced changes over the 3 last decades. It was used at the global scale, together with altimeter global datasets, to analyse the water volume variations of 137 large reservoirs and lakes around the world, validated with in situ data from 18 lakes, with the limitation that some regions like Africa and southwestern Europe suffer from a lack of Landsat observations [7]. In that context, the global coverage of Sentinel-1&2 observations highly improves systematic radar and optical satellite observation capacities, with a higher spatial resolution (10 to 20 m) and revisit frequency (5 to 12 days). These operational missions are crucial to base new near real time surface water dynamics monitoring systems at a global scale. Numerous studies have already proposed and compared methods to extract water area extents from their optical and radar data.

Multispectral Sentinel-2 data are used mainly through the application of a threshold on Spectral Indices (SI) that have been proposed worldwide: the Normalized Difference Water Index (NDWI [8]), Modified NDWI (MNDWI), Automated Water Extraction Index (AWEI), Water Index 2015 (WIs₂₀₁₅) to classify water pixels [9–14] among others. A calibration of those Spectral Indices (SI) thresholds is often crucial for each acquisition date and area covered. Automated thresholding procedures have been proposed, such as Otsu's thresholding, to split SI bimodal histogram values into water and non-water classes [15]. False detections of water bodies in urban areas, bare soil, clouds, snow/ice, and shadow areas can be mitigated with filters and combined methods. Nevertheless, it is impossible to correct all false detections on all scenes, especially shadows in low illumination in mountainous areas. Water bodies with low (high depth, black seaweed, dark bottom, shadow area) or high (high turbidity, shallow waters with bright bottom) reflectance might not be detected using those methods. A second method has been proposed to convert the RGB color space (SWIR2, NIR, RED bands from Sentinel-2 respectively) into the HSV (Hue, Saturation, and Value) space where the chromaticity (H and S) and the brightness (V) components are decoupled [16–18]. To cope with the false detection obtained with traditional water mapping methods using water indices in Mediterranean lakes and wetlands, Genetic Programming algorithms have been used to demonstrate their better accuracy especially for temporary lakes and wetlands [19]. Other methods based on classification or object segmentation or even deep learning have been used successfully [20–23]. Their comparison shows that SI based methods are less robust than classifiers especially in complex waterbodies (turbidity and aquatic vegetation for instance), but the requirement of external data (training samples) makes classifiers complex to set up [24]. A promising approach has been tested, to combine multiple SIs and raw reflectance in unsupervised multidimensional hierarchical clustering [25]. This last method aims at optimizing at a large scale the use of all the information contained in a Sentinel-2 scene without the use of any auxiliary data or time-series mosaicking. This software outperformed the existing programs and SIs most frequently used methods, such as Sen2Cor [26], Multi-

sensor Atmospheric Correction and Cloud Screening-Atmospheric Correction Joint Algorithm (MAJA) [27], FMask [28], or MNDWI thresholding techniques, especially for small surface water (<0.5 ha).

Water detection in synthetic aperture radar (SAR) data has been performed at a global scale (SWBD from SRTM mission), continental level [29,30], and local scale with coarse to high-resolution data, for both annual/seasonal water monitoring and flood events mapping [31–38]. Various methods have been used within the literature to delineate water from SAR data, either as a singular process or in combination. These include histogram thresholding [31,39–42], fuzzy classification [43,44], region growing [32,43] and texture analysis [45]. Thresholding techniques aim at separating low back-scatters from surface water using a threshold, which becomes especially difficult for mixed pixels or when the water back-scatter is affected by wind-induced roughness, floating vegetation, or when the swath of the image is large enough to have important incidence angle amplitude. In those cases, the threshold needs to be modified on a scene per scene basis [31,41]. To be successful, the histogram of the image values shall be bi-modal [46]. Thresholding technics have been combined with texture information [45] or region-growing segmentation algorithms used to increase the accuracy of water mapping [29,36,39,42,43,47]. Change detection and threshold techniques are also used in flood mapping [34,48,49]. Those mixed techniques give good detection accuracy (>95%) even if false detection corresponding to smooth dry terrains and radar shadows needs to be corrected with ancillary data. Few scientific references present a fully automatic water detection processor for surface water mapping from Sentinel-1 imagery [34,44,50]. Those methods rely on an initial classification using automatic thresholding, coupled with a fuzzy-logic-based classification refinement, and a final classification including auxiliary data. Other automatic methods rely on ancillary data (optical water maps, optical images) to train machine learning classifiers [24,35]. And among all classifiers, Random Forest (RF) method is known to be the fastest and most robust, i.e., small effect of RF parameters and wrong labelled samples on classification accuracy, with a small training time [51].

Besides the sensor physics (optical, radar) and its water detection algorithm, there is an interesting dimension to study the water dynamics: the observation time window. Thanks to frequent optical and radar observations of Sentinels 1&2, water detection algorithms may combine different observations (inside a time window) to produce the most accurate map for a given date. The size of the observation window might have a different impact on the quality of the resulting water maps depending on the validation criteria (i.e., quantification of global water surface changes, estimation of single lake surfaces, or river widths). The longer the window, the higher accuracies of water maps are expected, in detriment of missing rapid changes, such as flood events or changes in reservoirs related to infilling or draw off managements actions. For example, current water maps services are generated from monthly observations (GSW [17], Water bodies product from Copernicus Land Cover [52]). New products based on shorter observation windows (from 5 to 20 days) would provide faster and more frequent information on reservoir storage to water managers during water crises such as for water shortage events. An analysis of the overall quality effect is then needed to understand the use of time windows for water detection applications.

The aim of this article is to characterize the effect of the time observation window on water detection simultaneously at two levels: regional scale (water maps) and water bodies scale (reservoirs area time series). First, several state-of-the-art water detection methods on single images for optical and radar sensors are characterized and compared at large scale evaluation. Then one single-date detection method of each sensor is selected in the following. Different multi-temporal time-window logics are evaluated and compared to the selected single-date observation method for each sensor at a large scale and reservoir area monitoring. This evaluation allows also the comparison of water detection capabilities of Sentinels 1&2 (radar and optical). Finally, an analysis of the impact of reservoirs water level status and their geomorphological characteristics is presented to have a better

understanding of the factors that affect water detection/water body area extent-accuracy at the reservoir level, especially between radar and optical data.

2. Materials and Methods

2.1. Data

Optical and radar images are the main sources for water detection, which contain information of surface water extents at different acquisition dates. Optical images are obtained from the Copernicus Sentinel-2 constellation, with a revisit time of five days [53] and an operational life-time planned for the next decade [54]. Due to observation swath overlaps between adjacent orbits in high latitudes, some zones in Europe have a higher frequency of observations [55]. Ground resolution of optical images from Sentinel-2 is 10 m for visible bands and 20 m for SWIR bands. To better characterize the top of canopy level of images and correct atmospheric effects, Sentinel-2 images have been processed by MAJA algorithm [56], which corrects reflectance levels and detects clouds and shadows (produced by relief or clouds) [57]. S2-L2A images processed by MAJA are publicly available in THEIA datacenter [58].

For radar images, Copernicus Sentinel-1 observations have been chosen for the same reasons of continuity as Sentinel-2: similar revisit time (six days in Europe, 12 days in global) and similar resolution. On the main acquisition mode overland (Interferometric Wide Swath), Ground Range Detected (GRD) product at high-resolution level-1 has a resolution of 20×22 m and pixel spacing of 10×10 m on range and azimuth axes [59]. Two polarizations are available: VV and VH.

Auxiliary data are used to automatically select learning samples or apply corrections over water masks derived from radar and optical images. First, HAND (Height above nearest drainage) [60] derived from MERIT DEM [61]. This product, since highly correlated to the depth of the water table in case of theoretical inundations, is used to identify flood free areas and we use it to exclude zones from the classification process. These zones will always be detected as “non-water”. Second, the Global Surface Water [17] occurrence dataset has been chosen to identify globally permanent waters, used as training data for radar water detections by pixel level sampling. The resolution of the GSW product (30 m) seems appropriate for input data (S1 raw resolution = 20×22 m) in order to have a wide variety of permanent water samples.

2.2. Water Detection Methods

The process to generate water masks is divided into two different fluxes for radar and optical images. Since Sentinel-1 and Sentinel-2 have similar resolutions, intermediary and output data are generated on the same grid (Military Grid Reference System-MGRS) and resolution, which facilitates comparing the results. For this work, all data are resampled at 10 m resolution. The following Figure 1 depicts the workflow for the water masks generation, which includes the multiple-date processing of single-date water masks.

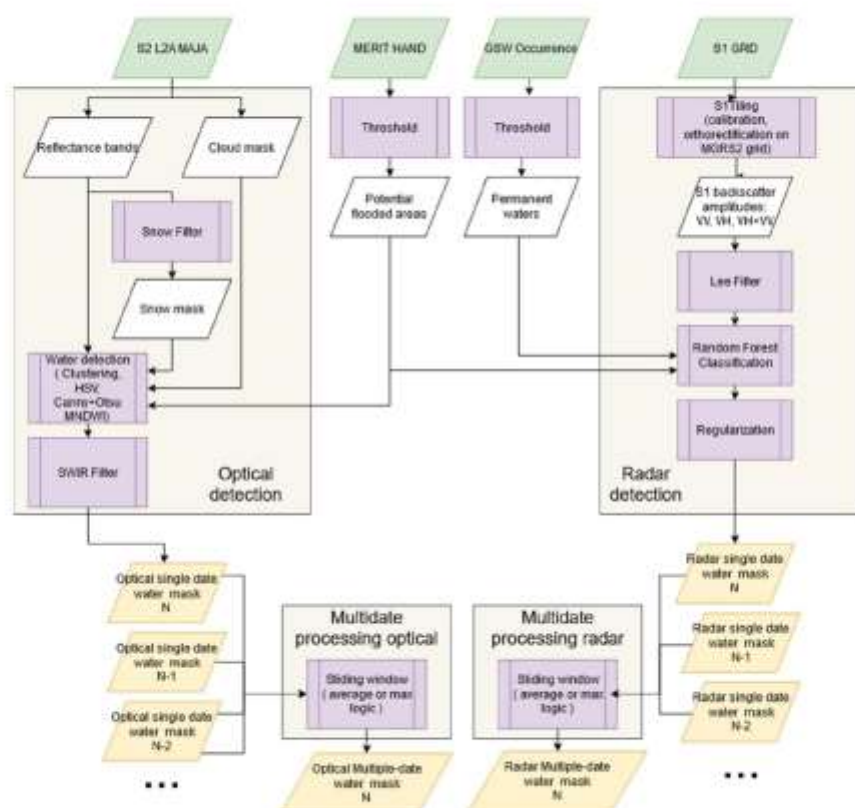


Figure 1. Water detection workflow on Sentinel1 and Sentinel2 observations, based on single and multiple date processes.

On the optical part, optical reflectances and cloud/shadow masks are retrieved from the S2L2A MAJA product. Cloud/cloud-shadow/relief-shadow masks are interpreted as a “No Data” layer. Second, a snow filter is used. Some of the water detectors that are evaluated in this work use the Modified Normalized Difference Water Index (MNDWI), which has the same spectral definition as the Normalized Difference Snow Index (NDSI) [62], quite sensible for snow detection. Therefore, it is important to separate snowy regions first, before classifying water regions. Since MAJA snow masks have 240 m resolution, a finer snow detector has been developed for this study. To avoid commission errors with turbid waters, a new snow filter is proposed based on reflectance values: Red > 1200, Blue > 1500, Red/Blue < 1.5. This filter is based on the fact that many turbid waters are prone to red values and shadowed snow regions prone to bluish tones.

To limit commission errors on mountain ridges or shadowed zones in steep valley regions, the MERIT HAND product is used to identify only floodable zones. In our case, regions placed higher than 25 m above their direct drainage zone are not likely to be flooded and they are excluded from the water detection process. This level has been calculated by comparing GSW maximum extent with the resulting floodable area from HAND.

Concerning the optical water detection algorithms, three state of the art classifiers have been integrated for the evaluation: (1) Edge filter Canny and Otsu thresholding on

MNDWI [9]; (2) Binary clustering on Hue/Saturation/Value dimensions issued from "PyIntertidalDEM" algorithm[63]; (3) Agglomerative Clustering issued from "WaterDetect" algorithm [25] in three variants: using two channels (SWIR band, NDWI), using three channels (SWIR band, NDWI, MNDWI) and using four channels (SWIR band, NDWI, MNDWI, MBWI [64]). Lastly, a SWIR filter is placed after the water detection classifiers based on values lower than 800.

The radar detection module first projects the input Sentinel-1 data to the Sentinel-2 grid, in order to work in equivalent layers. This process is handled by the SITiling chain [65], which does the calibration, projection, and concatenation of images in the MGRS2 grid. Multitemporal speckle filtering has not been applied, but a Lee filter applied on single date images is more appropriate as it accounts for edges. Multiple window sizes have been evaluated. On radar detection, a supervised method has been implemented based on Random Forest (100 estimators, max depth = 3, test size = 20%). Global Surface Water occurrence map is used as training data, where 2000 water samples (pixels) are obtained from permanent waters (occurrence > 90%) and 10,000 land samples from areas where water has been never detected (occurrence = 0%). Finally, a regularization process is done, to remove isolated "as water detected" pixels, through the "ClassificationMapRegularization" application from OrpheoToolBox [66], evaluated at different window sizes, and a single-date water mask is obtained.

Every optical and radar observation runs a water detection process to obtain a single-date water mask. Then, single-date water masks are combined together with a multiple-date process for a time-window/sampling period. Five time-windows were tested in this study: 5, 10, 15, 20, and 30 days. Two multiple-date merging methods have been developed. The "Average" method classifies a pixel of the mask as water if at least 50% of the observations were classified as water by the single date classifiers. The "Max" method classifies a pixel of the mask as water if it has been classified as water at least once by the single date classifiers. This last method is equivalent to the maximum surface water extent as observed during the time-window period by Sentinel-1&2.

2.3. Region of Interest Extraction Process and Surface Water Area Estimation

A geographical layer of water bodies area extent found within the studied MGRS2 tiles is created from an Optical Water Occurrence mask (OWO). The OWO is computed for each Sentinel-2 pixel as the ratio between the number of water detection over the number of valid observations for optical data. The water bodies area extent is built from the dilation by 50meters of the OWO exceeding 15%. Every targeted water body, localized by its coordinates, is associated with the closest water body area extent. The water area extent of a water body for a single date or a time-window water mask is computed as the sum of water pixels' area extent (10 × 10 m) contained in this area.

2.4. Reference Data-Large Scenes Water Masks

In order to validate water masks for large scenes in different seasons, a reference dataset has been developed using Active Learning for Cloud Detection (ALCD) software exploiting Sentinel-2 images. The motivation behind the use of ALCD for validating results rather than a hydrological database comes from the fact that water bodies are highly subject to change in time, depending on the season, meteorological conditions, management policies, etc.

The ALCD software was developed initially to generate reference cloud masks which may be used to validate Sentinel-2 cloud masks, such as those generated operationally by MAJA. This approach has been applied to generate reference water masks based on Sentinel-2 LIC [67] imagery.

The reference water masks are generated using an iterative active learning procedure which enables the creation of an accurate reference mask in less than 3–5 h for a 110 × 110 km scene. For this, it is necessary to manually create reference points on the image, train the model (based on Random Forest of OTB [66]) and predict with ALCD, then add new

reference points for the most problematic areas, repeat training/predictions cycles as many times as necessary (usually 5–7 iterations). The final step consists of a manual correction of persisting errors. For the training process, the following Sentinel-2 bands were used, B2: Blue, B3: Green, B4: Red, B8: NIR, B11: SWIR1, B12: SWIR2. In addition, derived indices such as MNDWI and slope information derived from SRTM [68] were also exploited.

The generated validation dataset has been shared in the ZENODO platform in open access [69]. A total of 14 sites have been covered on different eco-climatic zones (oceanic, Mediterranean, mountainous, continental) that are presented in Figure 2. Then, 26 scenes have been completely labelled on these sites on different seasons and conditions (snow, flood event, turbid waters, wetlands, urban areas, dry scenes), as described in Table 1.



Figure 2. 26 Large scene reference masks (110 × 110 km) have been generated over 14 sites in France.

Since the aim is to evaluate inland waters, coastal waters have been excluded using the GSHHG world shorelines database [70], applying a dilation of 400 m towards the land. The total surface of labelled data is 238,126 km².

Table 1. Description of the scenes used as reference with corresponding continental water surface in km² and percentage.

Region	Tile	Dates	Water Area Km ² (%)	Scene Content
Alpes 1	31TCK	24 February 2019	37.85 (0.31%)	Snow, Mountain
Alpes 2	31TGL	28 August 2018	125.04 (1.03%)	Mountain
Alsace	32ULU	12 September 2018	83.78 (0.69%)	Lowlands, Slopes
		21 Mars 2019	100.14 (0.83%)	Snow, Lowlands, Slopes
Ardèche	31TFL	17 February 2019	122.75 (1.01%)	Snow, Lowland, Slopes
		24 Mars 2019	119.75 (0.98%)	Lowlands, Slopes
		20 September 2019	109.15 (0.90%)	-
Ariège	30TCH	23 October 2018	39.94 (0.34%)	Mountains
		22 Mars 2019	31.42 (0.26%)	Snow, Mountains
Bordeaux	30TXQ	11 September 2018	187.36 (2.51%)	Coast, turbid/clear water
		23 February 2019	184.40 (2.47%)	-
Bretagne	30UXU	23 February 2019	54.93 (0.46%)	Wetlands, small bodies
		08 July 2018	39.94 (0.34%)	-
Camargue	31TFJ	27 September 2018	415.6 (4.14%)	Coast, large water bodies
		31 Mars 2019	489.11 (4.71%)	-
Chateauroux	31TCM	19 August 2018	133.5 (1.1%)	Lowlands
		25 February 2019	121.24 (1.0%)	-
Gironde	30TXR	23 February 2019	92.63 (1.37%)	Delta, turbid/clear water
Havre	30UYV	24 Mars 2020	85.91 (8.17%)	Delta, lowlands
Marnande	30TYQ	22 February 2019	95.33 (0.79%)	Wetlands, small bodies

Der Lake region	31UFP	10 July 2019	173.5 (0.85%)	Lowlands/gentle slopes
		04 December 2019	45.38 (0.37%)	-
		29 December 2019	121.86 (1.01%)	Flood
Orient Lake region	31UEP	17 July 2019	106.47 (0.88%)	Lowlands/gentle slopes
		04 December 2019	41.27 (0.37%)	-
		29 December 2019	81.65 (0.67%)	Flood
Total			3239.90 (1.36%) – 238,126 total surface	

¹ "–": same as previous line.

Concerning ALCD scenes accuracy, ALCD scenes have been evaluated over three parts of France covering heterogeneous water bodies. In the Ardeche region (Tile 31TFL), these results have been compared to a detailed static water database issued from watershed managers (3-Rivières watershed). In the Normandy (Tile 30UYV) and the Sologne regions (Tile 31TDV), the results have been compared to the TOPAGE static water database [71]. In each case, the size of un-detected water bodies was checked with their size distribution being around 85%, 10%, and 1% respectively for 0–0.1 ha, 0.1–0.2 ha, and 0.2–0.5 ha classes. Another result is that more than 90% of the water bodies larger than 0.5 ha were correctly detected. Accordingly, it is estimated that the water bodies above 0.5 ha are also similarly well detected over the other ALCD tiles.

2.5. Reference Data-Reservoirs Area

Time Series of in situ daily observations of 29 reservoirs in the Occitanie region, in the south of France, has been gathered to qualify area monitoring from 2017 to 2019. Given the water surface elevation $Z(t)$ and volume $V(t)$ series provided by reservoir managers, which contains bathymetric information on their relationship, the area $S(t)$ time series has been calculated as follows. First, the derivative values of $V(Z)$ function of each reservoir are calculated, resulting in $S(Z)$. Hence, $S(Z)$ is used to transform $Z(t)$ time series to $S(t)$ time series. Assuming $Z(t)$ uncertainty of ± 2 cm on in-situ measurements as a worst-case scenario during high filling rate periods, mean $S(t)$ uncertainty has been computed as ± 0.14 hectares or $\pm 0.12\%$ of the area in our reservoirs dataset. Such worst-case uncertainty is small compared to $S(t)$ errors obtained from satellite estimations (see Section 3.2.3). As shown in Figure 3, the reference reservoir maximum sizes were distributed from 1200 ha to 38 ha. The geographical distribution of the reference reservoirs in the south of France may be also noticed in Figure 3.

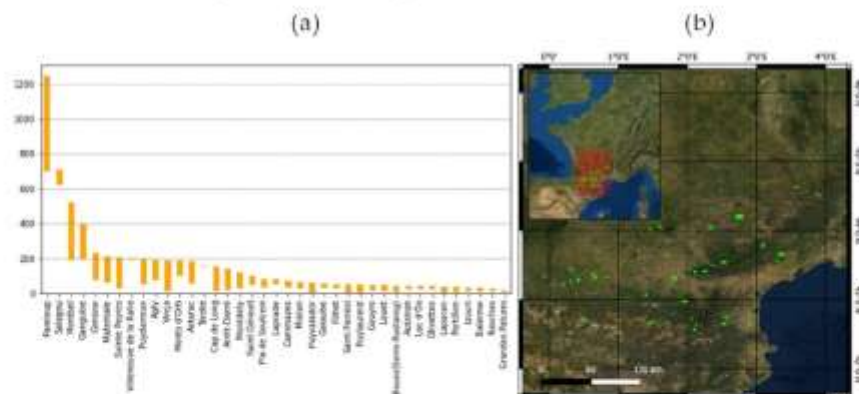


Figure 3. Reference reservoirs: (a) Surface dynamics over 3 years of in situ data (2017–2019), (b) Location of Reference Reservoirs in France (green).

To have a better understanding of the impact of the reservoirs' state on the area estimations, all the measurements have been associated with a reservoir state: high filling rate (area of reservoir >80% quantile of the in-situ area time series), low filling rate (area of reservoir <20% quantile of the in-situ area time series), increasing filling rate (if the area is greater than the previous day and outside high/low range), decreasing filling rate (if the area is lower than the previous day and outside high/low range).

2.6. Large Scene Water Masks Assessment

The assessment of large scene water masks has been developed in two stages.

First, optical and radar classifiers have been evaluated on a single-date basis. This assessment included the estimation of the following indicators: accuracy, F1-score, precision, and recall. As well, we give the number of comparisons with F1-score lower than 0.5 shown as "failures", to provide an insight into the instability of the water detection method.

Second, the best optical and radar classifiers from the first stage have been selected based on their median F1-Score. Then, different multiple-date methods have been assessed based on the same indicators of the first stage.

2.7. Reservoirs Area Assessment

Since reference reservoirs have a wide range of maximum water surface areas (from 20 ha to 1200 ha), error metrics chosen for this study should consider reservoirs' size disparity. Thus, error metrics based on relative and absolute relative error have been used for this assessment, using the following definitions:

$$\text{relative error} = \frac{X - X_{ref}}{X_{ref}} \quad (1)$$

$$\text{absolute of relative error} = \text{abs} \left(\frac{X - X_{ref}}{X_{ref}} \right) \quad (2)$$

where X is the reservoir area measured by satellite at a given date and X_{ref} refers to the reservoir area obtained by in-situ measurements.

2.8. Reservoir Characteristics and Geomorphological Indicators

Reservoir characteristics might be also related to the quality of area monitoring. In this study, we propose to compare maximum area extent, altitude, and four geomorphological indexes to their results on Absolute Relative Error on area estimation.

The two first geomorphological indexes are the "Eroded Area" index, defined as the eroded surface area (maximum extent surface buffered by 50 m) divided by the maximum surface area, and the "Eroded Perimeter" index, which uses perimeter instead of area. Both indexes penalize reservoirs with narrow arms (widths <100 m will disappear), and yield higher values for large surfaces, where the loss of eroded surface is relatively lower compared to the maximum surface.

The second two morphological operations are the "Convex-Hull Area" index and the "Convex-Hull Perimeter" which use the Convex-Hull shape instead of the Eroded one. Low Convex-Hull Area or Perimeter index denotes a complex shape with sinuous boundaries.

Figure 4 shows examples of Maximum surface, Convex-Hull surface, and Eroded surface and their geomorphological values on four different reservoirs. Table 2 present the values of the proposed geomorphological indexes, where maximum values are represented with bold style.

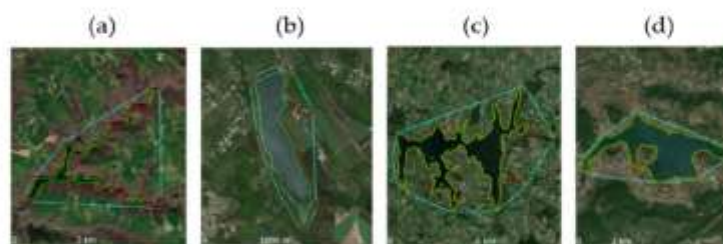


Figure 4. Representation of maximum extent (red), Convex hull (blue) and Eroded surface (green) for the following reservoirs: (a) Saint Géraud, (b) Tordre, (c) Pareloup, (d) Salagou.

Table 2. Geomorphological indexes derived from maximum extent surfaces for 4 different reservoirs. Maximum values are presented in bold style for each index (row).

Geomorphological Index	Saint Géraud	Tordre	Pareloup	Salagou
Eroded Area Index	0.35	0.66	0.78	0.85
Eroded Perimeter Index	0.68	0.81	0.88	0.95
Convex Area Index	0.21	0.72	0.35	0.619
Convex Perimeter Index	0.62	0.79	0.40	0.69

3. Results

The proposed water detection methods and the multi-temporal approaches have been evaluated by two different assessment procedures. The first part addresses the quality of large scene water masks by comparing the resulting water maps to reference masks for specific dates. The second part addresses the quality assessment of area estimations for individual water bodies using in-situ data from 29 dams (two years' time series). For both sections, single-date and multiple-date water detection results are compared separately for the optical and radar observations.

3.1. Large Scene Water Masks Evaluation

Every ground truth mask generated with ALCD has been compared with all the optical and radar water maps available in the range $[D - 5, D + 5]$ days, where "D" is the acquisition date of the ALCD input mask. Radar and optical water detection methods have been evaluated independently. Optical input images with partial (cloudy) observations have been included in the evaluation; nevertheless, the areas marked as "cloud" or "shadow" on the S2-L2A input products are excluded from the evaluation. Thus, the intersection of all valid pixels from ground truth and classifier outputs are included in the evaluation (no sampling).

3.1.1. Optical Water Masks Evaluation at Single-Date Observation

Three different algorithms have been evaluated for water detection on Sentinel-2-L2A images: (1) Edge filter Canny and Otsu thresholding on MNDWI; (2) Binary clustering on Hue/Saturation/Value dimensions extracted from PyIntertidalDEM algorithm; (3) Agglomerative clustering based on the "WaterDetect" algorithm in three variants: two features (SWIR band, NDWI), three features (two first features + MNDWI) and using four features (three first features + MBWI). Furthermore, an additional filter has been evaluated on top of the basic algorithms: a threshold filter at 0.8 on the SWIR band. A total of 74

images have been evaluated and compared to the 26 ALCD ground truth masks using indicators described in Section 2.6. Results are presented in Table 3:

Table 3. Median Scores for the Optical methods assessment and total number of failures (scenes with $F1 < 0.5$) for 74 scene evaluations. Best values are presented in bold style for each score (column).

Optical Method	SWIR Filter	F1 Score	Precision	Recall	Accuracy	Failures
Canny-Otsu MNDWI	-	0.814496	0.916422	0.826836	0.99697	7
Canny-Otsu MNDWI	Yes	0.832646	0.92836	0.826021	0.997206	5
HSV	-	0.7884	0.91694	0.875777	0.996111	12
HSV	Yes	0.810096	0.941263	0.856648	0.996487	7
Clustering 2 channels	Yes	0.888889	0.944292	0.888526	0.997996	6
Clustering 3 channels	-	0.868414	0.864012	0.933613	0.997416	7
Clustering 3 channels	Yes	0.890239	0.893692	0.917815	0.997814	4
Clustering 4 channels	Yes	0.886854	0.925346	0.886162	0.997939	6

Agglomerative clustering methods (“Clustering”) achieved the best F1 median results over the 74 comparisons, providing the best scores on recall. Also, the “Clustering 3 channels” method with SWIR filter provided the best median F1 Score and the lowest number of failed detections (three images of $F1 < 0.5$ over 74 comparisons). The rest of the optical methods (Canny-Otsu MNDWI and HSV) have better Precision than Recall scores, SWIR filter has shown an improvement of median precision scores or number of failures at this evaluation.

3.1.2. Radar Water Masks Evaluation at Single-Date Observation

The radar water detection method, based on Random Forest trained on historic optical water detections, has been assessed through combinations of different window sizes of Lee speckle filters (unit: 10 m pixels) and output regularization filters based on ball structure (radius unit: 10 m pixels). A total of 123 images have been evaluated and compared to the 26 ALCD ground truth masks providing the following results, presented in Table 4:

Table 4. Median Scores for the Radar methods assessment and total number of failed comparisons ($F1 < 0.5$) for 123 scene evaluations. Best values are presented in bold style for each score (column).

Radar Method	Lee Filter Size	Regularization Ball Radius	F1 Score	Precision	Recall	Accuracy	Failures
Random Forest	No	1	0.681	0.667	0.716	0.995	14
Random Forest	3 × 3	1	0.689	0.674	0.723	0.995	20
Random Forest	No	2	0.727	0.787	0.710	0.996	8

The effect of larger preprocessing Lee filtering windows slightly improves the median recall score in detriment of a higher number of failures. The effect of a larger Regularisation window improves the median precision and the number of failed scenes. Regularisation has a cleaning effect on small, isolated detections or water omissions, produced by speckle and surface roughness irregularities present in crop bare soils or windy water surfaces.

3.1.3. Multiple-Date Water Masks Evaluation

The multiple-date water masks assessment considers different time window lengths and different methods to process single-date water masks.

The selected optical water detection algorithm, based on the best F1-score, is the “Clustering 3 channels” with a SWIR filter; the selected radar water detection algorithm

is based on Random Forest with a regularisation size equal to two. Like in previous assessments, 74 optical evaluations and 104 radar evaluations have been performed. Results are shown on Table 5 their median values and number of failures:

Table 5. Median Scores for Optical and Radar assessments on different multi-temporal window methods and lengths and number of failed scenes ($F1 < 50\%$) over 74 optical scenes and 104 radar scenes. Best values are presented in bold style for each score (column).

Method	Average					Max				
	F1_Score	Precision	Recall	Accuracy	Failures	F1_Score	Precision	Recall	Accuracy	Failures
Radar 1 day	0.727	0.787	0.710	0.9962	8	0.727	0.787	0.710	0.996	8
Radar 5 days	0.744	0.789	0.756	0.996	1	0.705	0.669	0.784	0.995	3
Radar 10 days	0.795	0.887	0.754	0.9970	3	0.672	0.584	0.810	0.994	9
Radar 15 days	0.793	0.890	0.755	0.9970	3	0.588	0.492	0.844	0.993	23
Radar 20 days	0.797	0.894	0.752	0.9972	2	0.565	0.424	0.862	0.992	33
Radar 30 days	0.802	0.907	0.746	0.9973	2	0.508	0.586	0.869	0.990	46
Optical 1 day	0.890	0.893	0.917	0.9978	4	0.890	0.893	0.917	0.9978	4
Optical 5 days	0.896	0.906	0.895	0.9979	9	0.894	0.870	0.930	0.9978	7
Optical 10 days	0.899	0.889	0.903	0.9980	10	0.883	0.834	0.936	0.9975	9
Optical 15 days	0.901	0.903	0.904	0.9978	10	0.857	0.786	0.942	0.9966	10
Optical 20 days	0.886	0.885	0.915	0.9977	7	0.846	0.767	0.960	0.9963	18
Optical 30 days	0.889	0.899	0.894	0.9981	7	0.824	0.740	0.953	0.9964	14

Regarding radar results, “average” methods improve precision, and hence, F1-Score with longer time-windows. Recall remains constant even with increasing time-window length, whereas failures drop considerably if a time window is set. “Max” methods, on the contrary, worsen the performance when increasing the length of the time-window in precision and number of failures.

About optical results, “average” multiple-date methods do not improve scores for the general scene evaluation, and failures are increased. On the other hand, “Max” multiple-date methods decrease performance for longer windows.

3.2. Area Monitoring Evaluation on Reservoirs

This evaluation is based on the comparison of reservoirs’ areas issued from satellite observations with daily in-situ measurements from 29 reservoirs in the South of France. Time series of two years (2018, 2019) have been used for this analysis. Figure 5 shows an example of reference area time-series and satellite area estimations for one reservoir.

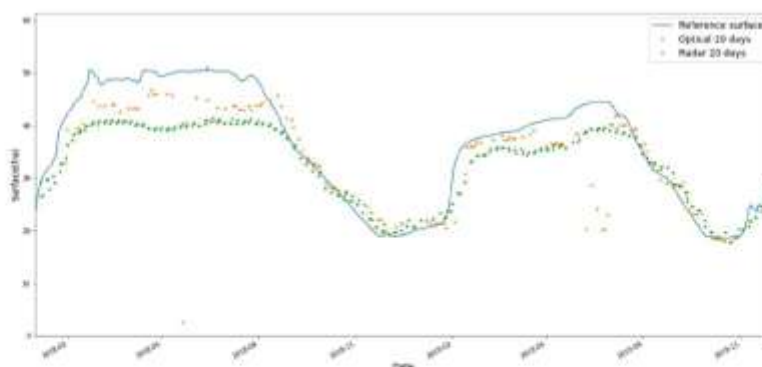


Figure 5. Satellite area estimations of Louet reservoir (Optical in orange, Radar in green) using 20 days time-window compared with the reference in-situ areas (blue line).

3.2.1. Error Assessment of Reservoirs Area Monitoring

The surface area of each water body is computed from a large-scene water mask, as described in Section 2.3, and error metrics are then processed, described in Section 2.7. The proposed reservoir monitoring methods are classified according to how large-scene water masks are generated using the following criteria:

- Sensor: optical measurements ("MO") and radar measurements ("MR"), based on the same water classification methods evaluated in Section 3.1.3.
- Single/Multi-date method: They are categorized in three classes; Single-date methods are based on just one satellite observation ("MO1", "MR1"), multiple-date methods apply a backwards time-window ("MO2", "MR2") and the last methods calculate the sum of all the surfaces detected as water at least once during a natural month ("MO3", "MR3"). Multiple-date methods ("MO2", "MR2") present two possible logics: average ("_avg") and maximum pixel wise surface ("_max"), as described in Section 2.2. Also, multiple-date methods with time-windows have a suffix ("_WN"), where N is the time-window size in days.

In order to understand the general behavior of each method, all measurements for 29 reservoirs for two years are shown in Figure 6 by boxplots. Optical methods have been represented in red tone boxes and radar methods in blue tone boxes. Multiple-date methods based on "maximum" logic are represented with more intense colors. Green horizontal lines inside the boxes show median values. The box is bounded at 25% and 75% quantiles and whiskers are set on 10% and 90% quantiles.

Regarding median values for this global visualization, it may be noticed that all the methods have a generalized negative bias on reservoir area estimation. The optical method's underestimations are less accused than the corresponding radar methods. For example, MO2_W15_avg has a median value of -5.49%, while the radar version MR2_W15_avg has a median value of -13.8%. Regarding the effect of time-windows, it may be observed that longer windows are associated with a reduced underestimation. For instance, median values on the optical values improve from -7.23% in MO1 to -5.28% in MO2_W20_avg. Radar median values improve from -16.04% in MR1 to -4.1% in MR2_W20_max. Among the time-window methods, those applying the "maximum" multi-date logic have smaller bias values than the equivalent "average" methods for the same window length. MO3 and MR3 methods may be considered equivalent to MO2_W30_max or MR2_W30_max methods where estimations are produced only once per month (only natural month data are considered). Their error boxplots are quite similar as expected. Respecting median values, MO2_W30_max and MR2_W30_max achieve better accuracies, with 0.3% and 1.3% respectively.

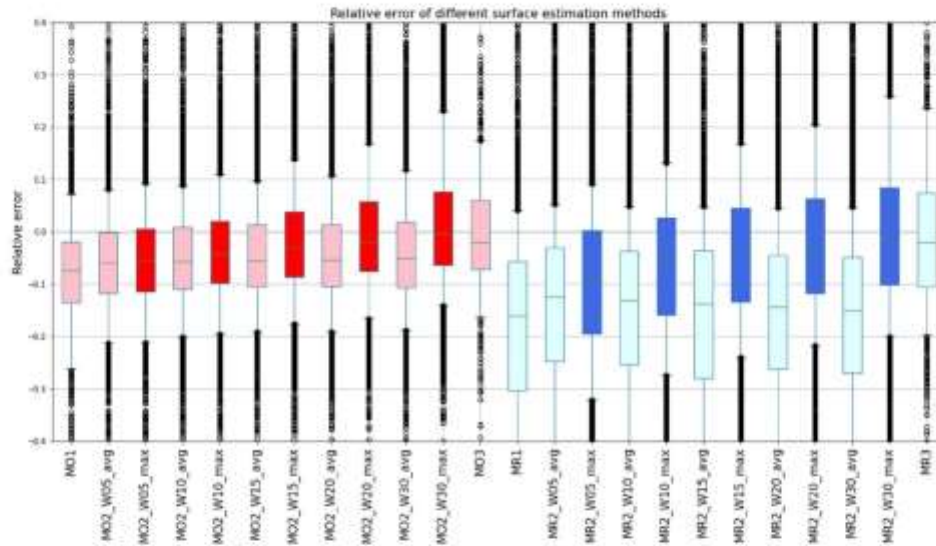


Figure 6. Relative error of Reservoir area estimation using different combinations of optical and radar masks for 2 years of measurements at 29 instrumented reservoirs. Naming convention: MO = Optical, MR = Radar; MO1, MR1 = single date masks; MO2, MR2 = multiple dates masks; MO3, MR3 = maximum extent during the natural month; WN = N days of time-window length; avg = average logic; max = maximum logic. Optical methods have been represented in red tone boxes and radar methods in blue tone boxes. Multiple-date methods based on “maximum” logic are represented with more intense colors. Green horizontal lines inside the boxes show median values. The box is bounded at 25% and 75% quantiles and whiskers are set on 10% and 90% quantiles. Extreme values are represented as black circles.

3.2.2. Influence of Reservoir Filling Rate

The dynamics of water stored in reservoirs have an impact on the water surface characteristics (area, nature of boundary water/land areas) and thus may affect the water mask quality for each reservoir. We define two contrasted hydrological conditions by subset the water area extent: high filling rate and low filling rate (see Section 2.5)

Figure 7 shows the same relative error of each method presented in Figure 6 but restricted to dates with high filling rate status. Two main differences appear. First, the general negative bias of area estimations is higher than in the global case. For example, median relative errors of “MO_W20_avg” change from -5.33% in all reservoirs status to -7.96% in the high filling rate. Second, this negative bias decreases less while increasing the time-window compared to the global case.

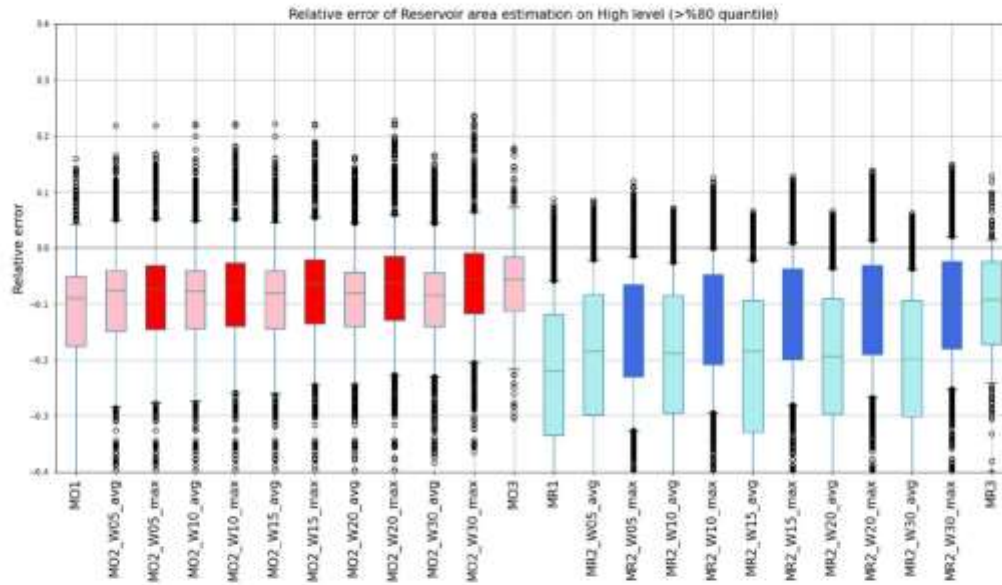


Figure 7. Relative error of Area estimation for High level reservoirs (>80% area quantile) measurements. See description at Figure 6 caption.

This generalised underestimation of the area extent detected for high filling rate periods might be explained by the dense vegetation often covering the rarely flooded shores and upstream parts of the reservoirs. Water detection algorithms are very affected in areas with dense vegetation in optical images [17] and SAR [72], as shown in Figure 8. Shrubs and small trees also affect radar detection, to a larger extent than optical data detection, as shown in water occurrence maps in Section 4.1. Confusions between sand and water are also a source of detection failures with radar data, as shown in the inner part of the lake in Figure 8.

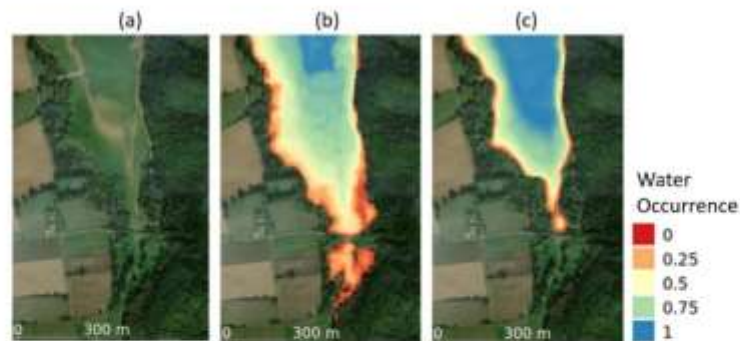


Figure 8. Effect of dense vegetation and sandy surfaces on occurrence maps on the southern area of Miélan reservoir, where radar water detections are quite limited due to vegetation. (a) and background: ESRI aerial image; (b): water occurrence map from 2 year of optical observations; (c): water occurrence map from 2 year of radar observations. Zero occurrence has been displayed as transparent.

The same analysis made for low filling rate periods shows different behaviors, as shown in Figure 9. In hydrologically drier conditions, bias increases with increasing time window. Time-window methods using “maximum” logic turn negative bias into high positive bias. The “average” method over 20 days leads to removing the negative bias observed with the optical data detection methods whereas radar detection biases are not improved with time window. This overestimation is mainly due to the integration of higher water level observations within the time window.

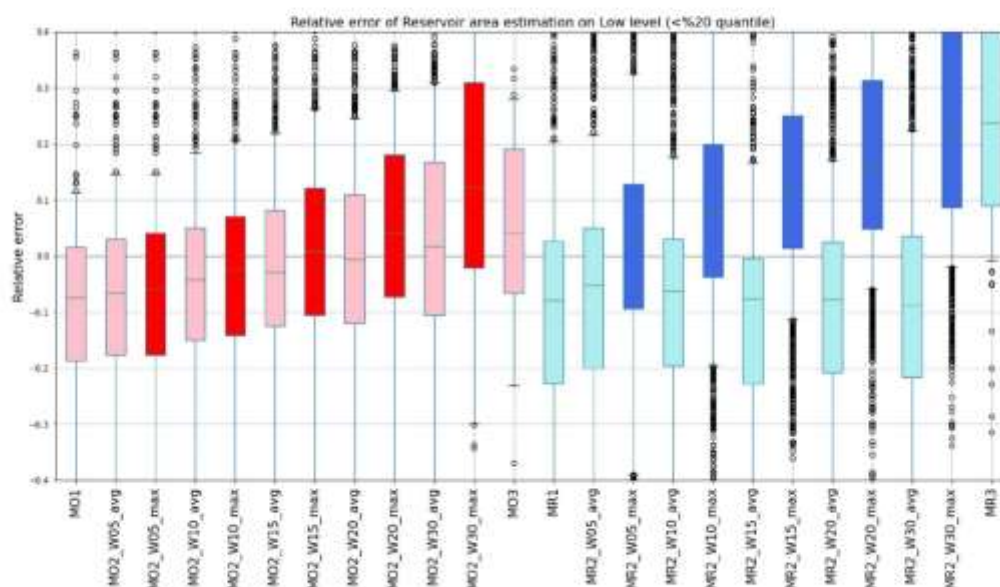


Figure 9. Relative error of Area estimation for Low level reservoir measurements. See description at Figure 6 caption.

3.2.3. Analysis of Absolute of Relative Error on Reservoir Area Estimation

To provide a more comprehensive study on reservoir area estimation error and its dispersion, we propose to analyze absolute values of relative area errors from all the reservoir area estimations altogether. Working with absolute values, median or quantile error values better reflect all kinds of error, positive and negative. In Figure 10, we have sorted the quantile values (50%, 90%) of absolute of relative error of area estimation of each method for all reservoirs as a whole, and the following results may be observed:

- On the optical methods, MO3 (maximum on natural month) provides the best results on quantile 50% (median value), but it performs worse on quantile 90% compared with other optical methods. Time-window methods based on “maximum” logic perform well on 50% quantiles, but they are not the best on 90%. Time-window methods based on “average” logic with 10–15–20 days perform well on quantile 90%. For example, the mean absolute relative error in MO1 is 16.9%, whereas MO2_W15_avg is 12.9%. In conclusion, “average” and “max” time-windows have similar positive results compared to single-date methods.
- On the radar methods, MR3 (maximum on natural month) provides the best performance on both quantiles 50% and 90%; on the time-window methods, methods based on “maximum” outperform the “average” methods for both quantiles. For example, the mean absolute relative error in MR1 is 22.7%, whereas MR2_W10_avg is 19.5% and MR2_W10_max is 15.1%. For “maximum” methods, relatively long windows (20,

30 days) perform better than short windows (5, 10 days). For “average” methods, quantiles 50% are better with short windows (5, 10 days) but quantiles 90% are better with middle (10, 20 days) windows.

- Any multi-temporal method (“MO2”, “MO3”, “MR2”, “MR3”) outperforms the other methods based on single observations (“MO1”, “MR1”) on quantile 50% and 90% of absolute of relative error on the area estimation.
- Any optical method (“MO”) outperforms the other methods based on radar observations (“MR”) on quantile 50% and 90% of absolute of relative error on the area estimation.

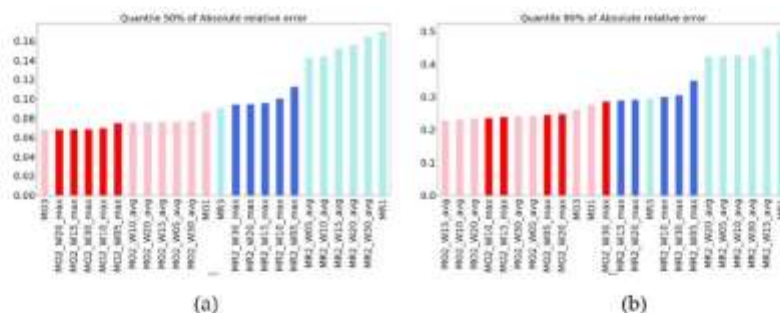


Figure 10. Quantiles of absolute of relative error on area estimation on each method. (a) Quantile 50%; (b): Quantile 90%.

3.2.4. Geomorphological Influence on Area Estimation Quality

To have a better understanding, an analysis of optical and radar area monitoring error (median error) has been plotted against the maximum area, altitude, and four geomorphological indicators of all reservoirs in Figure 11. Separately for optical and radar methods, a linear regression function has been fitted for each ensemble of Area Monitoring Error and geomorphological information of each reservoir. For the optical method (MO2_W20_avg), individual geomorphological indicators show a poor correlation with area monitoring quality, with R^2 values < 0.2 . Using all morphological indices in one single linear model for optical, it yields a 0.42 R^2 score, which is low. For Radar Methods (MR2_W20_avg), Eroded Perimeter and Area indexes have the most relevant information to predict the area monitoring quality, with R^2 scores of 0.44 and 0.58. Results corroborate the idea that radar water detection is limited for narrow water areas (width < 100 m) by the lower spatial resolution and the effect of bright adjacent land zones comparatively to Sentinel-2. This impacts significantly the total area extent detection accuracy. Using all morphological indices in one single linear model for radar yields a 0.67 R^2 value. Hence, the presented geomorphological indices would better predict the quality of area estimation of each reservoir on radar methods than in optical ones. With respect to a maximum surface dataset of 104 reservoirs in the Occitanie region, we appreciate that 37.5% of reservoirs have Eroded Perimeter index values under 0.8, which suggests that they will have bad area estimation quality using radar methods (median absolute of relative error higher than 20%).

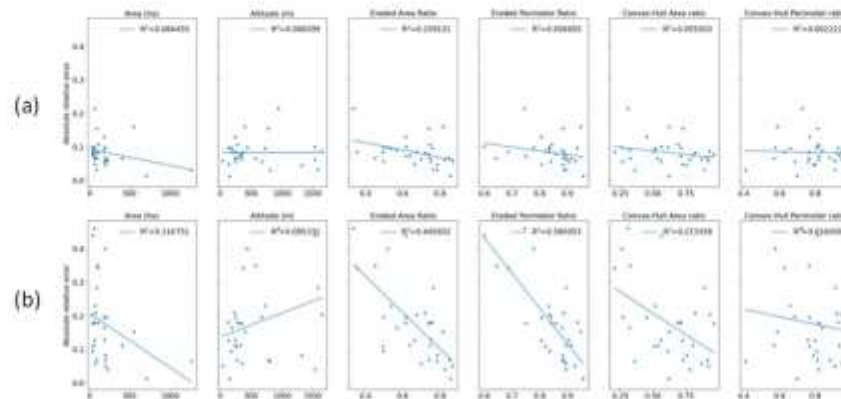


Figure 11. Median Absolute of relative error of area estimations of reservoirs depending on their geomorphological indexes: maximum area, altitude, Eroded Area ratio, Convex-Hull Area ratio, Convex-Hull perimeter ratio and Eroded Perimeter ratio. (a) Results based in optical method “MO2_W20_avg” (b) Results based in radar method “MR_W20_avg”.

4. Discussion

4.1. Multitemporal Impact on Optical and Radar Detections

We presented a methodology to generate large-scale water maps and process them to monitor reservoir areas. In comparison with other works on reservoir monitoring based on optical surface observations [73,74] or radar observations [75], the presented surface and area estimation method proposes a multi-temporal approach. This strategy of merging multiple date information corrects occasional noisy events, acting as a temporal (not to confuse with spatial filter) low-pass filter on the time series. On the other hand, detection of rapid changes will be penalized, which affects reservoir area estimation (e.g., coastal reservoirs that change with the tide). Regarding end-to-end performance on reservoir area monitoring presented in Section 3.2.3, results show a general positive effect of multi-temporal approach on reservoir area estimation, for both radar and optical methods and for any kind of merging logic (average, max). Nevertheless, there are differences between radar and optical performances.

First, there are constant factors affecting water mask quality which are different for optical and radar observations. On SAR images from Sentinel-1, water detection is based on spotting zones with low backscattering. When a water body is contiguous to high radar reflectance elements (trees, artificial structures as bridges, or walls), the SAR impulse response of these bright elements affects the adjacent low signal zones on water surfaces. Being Sentinel-1 resolution equal to 20×22 m and affecting at least one adjacent pixel, it may represent a loss of 40 m when measuring river widths, or erosion of 20 m inside the perimeter of a lake or reservoir. Detection problems by trees are aggravated by Sentinel-1 high incidence angle range (29.1° – 46°) compared to Sentinel-2 (0° – 10.1°). There are other constant factors that challenge the radar water detection at boundaries, such as sand, which may have very low backscatter for certain angles and certain moisture levels [44], or the effect of vegetation at the surface water. Other relatively constant factors (which duration is longer than several weeks) also affect optical detection, like a tree or urban shadows [76] or sunglint [77]. As a result, water boundaries zones with sand and vegetation are often better detected with optical measurements. Figures 8 and 12 show the systematic drastic drop of radar occurrence near water bodies shores, especially for forested surroundings. On the contrary, optical occurrence remains stable and drops closer to the shore.

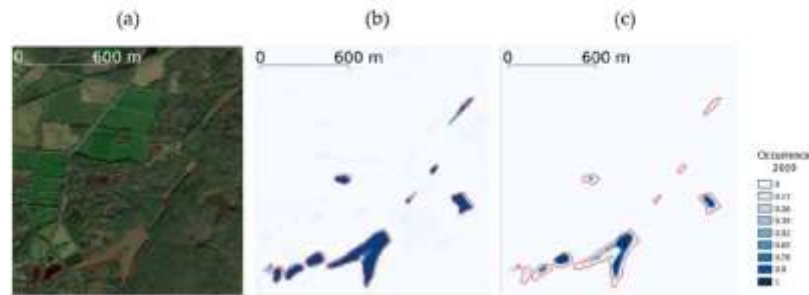


Figure 12. Representation of the reference water mask on a specific date (red polygons) with ESRI Satellite background (a), compared with water occurrence map based on Optical measurements for 1 year of observations (b) and the water occurrence map based on Radar measurements for 1 year of observations (c).

Second, the quality of the masks is affected by variable observation conditions depending on the method (optical, radar). The multi-temporal approach allows to correct and erase false detections inside the time-window: clouds or cloud shadows perturbations for optical detection; speckle, certain bare soil conditions (roughness, soil moisture), and water roughness variations due to wind for radar detection [78]. In this study, clouds or cloud shadows are generally flagged and removed by the MAJA preprocessing chain. Some remaining artifacts (cloud or shadow not detected) may sporadically affect detection quality. On the contrary, scene variable conditions that perturb the radar detection quality, such as speckle, water/soil roughness, or soil moisture, are always present in radar images and cannot be flagged. Hence, radar water masks quality might be more frequently affected by undetected changes of these variables.

Multiple-date methods statistically improve more unflagged radar detections than flagged optical detections. This should come to a reduction of omission errors in rough waters or commission errors in bare soils, and compensates for speckle effects in omission and commission, as illustrated in Figure 13. Also, multiple-date methods like maximum extent logic combine ascending and descending radar views of reservoirs surrounded by trees, reducing the shadowed area and improving area estimation. This is an important result for regions that have a permanent cloud coverage in tropical or polar areas, for which the Sentinel-1 data time-series could be more complete/accurate than cloud-free composite Sentinel-2.

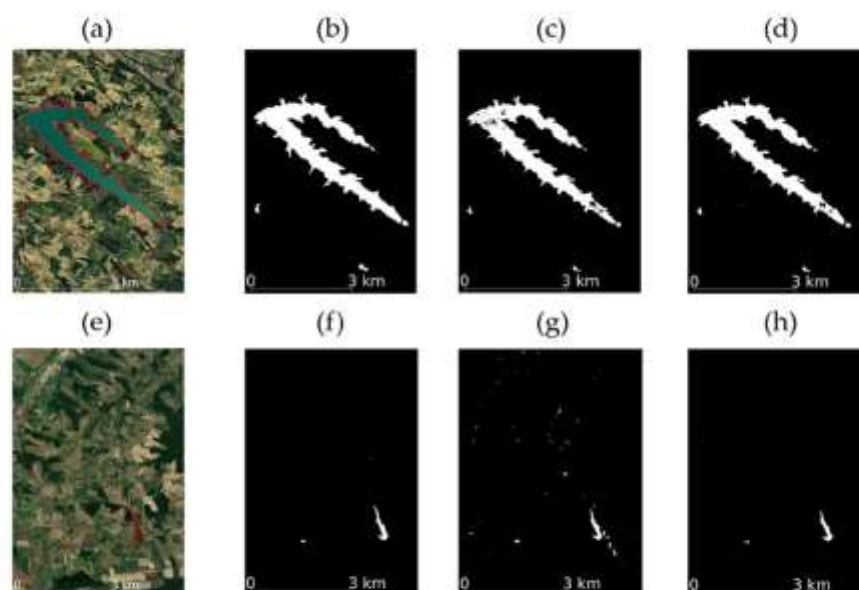


Figure 13. Impact of multi-temporal water detection on radar images. Top images show the improvement on water detection on rough waters on a large reservoir. Bottom images show the improvement on bare soils commission error. (a,e): GoogleMaps Satellite Image with reference water mask contour (red); (b,f): Reference water mask; (c,g): single-date radar water mask; (d,h): Multiple-date radar water masks (15 days) with average logic.

Given the cited differences between optical and radar methods, which are affected by different constant and variable factors, we have kept a separate analysis for reservoir monitoring evaluation. In this context, it should be taken into account that water boundaries detection will be different on optical and radar sensors for thin water bodies (<50 m), sandy shores, artificial structures, and some vegetation types above water. Nevertheless, merging both outputs together is possible and complementary (better accuracies on optical methods and assured systematic radar observations). On complex zones like wetlands, the benefits of merging both sensor measurements have been already shown [79] or in flood duration analysis [74].

4.2. Observation Frequency

The presented work has been evaluated in metropolitan France, where Sentinel-1 and Sentinel-2 observation conditions might not be completely generalizable. Sentinel-1 observations are more frequent (2–6 days) than other zones in the world (12 days) [55,80]. Also, Sentinel-2 observations in tropical and arctic zones are very limited due to frequent clouds. With regard to our reference reservoirs, there is a notorious difference in frequency of reservoir observations due to swaths overlap or local climatological conditions for optical measurements. Figure 14 shows how time-window with average logic using 20 days changes the quality of the area estimations depending on the mean duration between clear observations. This effect is expressed as the difference of the mean absolute of relative error multiple date estimations (MO2_W20_avg, MR2_W20_avg) compared to single date observations (MO1, MR1). In view of the results, where the estimations are improved in general, it cannot be stated that more frequent clear observations increase

accuracy for either optical or radar methods. In other words, even if the time-window captures a small number of observations, quality improvement on area estimation has been observed. This fact could justify longer time-windows in zones with less frequent Sentinel-1 observations, with durations like 12 or 24 days in order to include at least one or two observations.

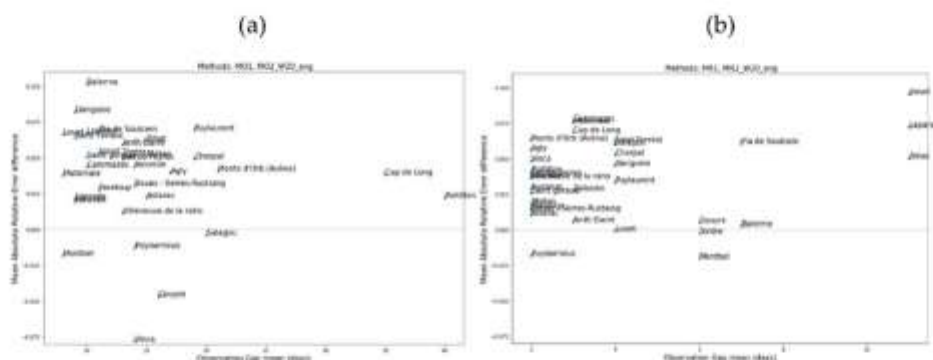


Figure 14. Improvement of mean absolute of relative error between single date measurements (MOI, MRI) and multiple date (MO2_W20_avg, MR2_W20_avg) compared with the Number of observations on each reservoir for optical (a) and radar (b) measurements.

5. Conclusions

We argued at the beginning of the article that multiple-date approaches had to be considered and evaluated for the improvement of water maps (large scenes) and the accuracy of reservoir area estimation. Through the evaluation of 26 reference water maps (110 km × 110 km at 10 m resolution), and the analysis of 29 reservoirs in France for 2 years, we have shown the improvement of accuracy for optical (Sentinel-2) and radar (Sentinel-1) measurements in two ways. First, large scene water maps scores have improved significantly on radar multi-date approaches (an increase of 0.08 in F-Score with average logic at 30 days), while optical maps just yield similar results to multiple-date. On the reservoir area assessment, both multiple dates approaches (average and maximum logics) improved results for both radar and optical approaches for any time-window length. For example, the mean absolute relative error on area estimation from single date radar measurements improved from 22% to 14% using the Maximum logic on 15 days. For optical data, even if large scene quality did not apparently improve (similar F1-score), the mean absolute of relative error on reservoir area estimation decreases from 16.9% to 12.9% using the Average logic on 15 days. Table 6 summarizes the impacts of multiple-date approach compared to single-date results.

Table 6. Effect of multiple-dates method compared to its corresponding single-date method on different evaluations.

		Large Scene Evaluation		Reservoir Area Evaluation	
		Optical	Radar	Optical	Radar
Multiple-dates Method Effect	“Average” logic	Neutral	Very Positive	Positive	Positive
	“Maximum” logic	Negative	Negative	Positive	Very Positive

Additionally, the present work yields a quality evaluation and comparison of radar and optical water maps, using state of the art methods. While similar in the precision score

(0.9), optical water maps had better median recall scores (0.9) than radar (0.75), which may be explained by difficulties of radar on smaller water bodies like thin rivers or small ponds. Concerning reservoir area monitoring, optical estimations outperform radar. Optical approaches yield a median absolute of relative error of 6.8% (30 days with “maximum” logic) while radars yield 9.5% (30 days with maximum logic). It is important to remark that both methods have a tendency to underestimate area (negative bias), being more significant on the radar. Considering the water level status of reservoirs, our methods have shown better estimations at low filling rate conditions than at high filling rates. This observation is important for application purposes, as low filling rates occur during dry conditions when water restrictions and water shortage happens. Lastly, in order to predict the radar performance on reservoir area monitoring, we propose the analysis of the “Eroded Area Ratio”, showing a considerable correlation with area estimation quality (mean of absolute of relative error).

As a result of conducting this research, we propose two improvements to future water maps and reservoir monitoring systems based on satellite measurements. First, future water maps products should take advantage of multiple-date approaches to improve mapping accuracy, especially in radar observations. We propose to use time windows between 10–20 days applying “average” logic which yields significant improvements on large-scale water maps and reservoirs area estimation. In zones with less frequency of Sentinel-1 observations (mainly outside Europe), time-windows of 12–24 days would be recommended. Second, reservoir filling rate may influence the quality of area estimations, being critical on full reservoirs surrounded by dense vegetation or artificial structures affecting radar. Some solutions are proposed for future investigations: flag high-filled reservoir measurements surrounded by such conditions by means of land cover maps; reconstruct area dynamics just focusing in zones not surrounded by vegetation or artificial structures [75] or correct area dynamics using auxiliary measurements, such as altimetry, digital elevation models or in situ measurements. New altimetry systems providing a larger number of tracked water bodies in a global approach, like IceSat2 based on Lidar [81], altimeters Sentinel-3 [82], Sentinel-6 [83], and the future SWOT mission [84], will provide essential information to be combined with the proposed area estimations. Such a combination will help to better determine the volume dynamics of reservoirs and lakes globally. As a final idea and with regard to the results of this work, we suggest that operational reservoir monitoring systems should combine both sensors: optical measurements, with better accuracies, and radar measurements, with more observations independent from cloud conditions, to provide a complete as possible insight on water dynamics.

Author Contributions: Conceptualization, S.P.-L.; methodology, S.P.-L. and S.F.; software, S.P.-L., M.C.R.C.; validation, S.P.-L.; formal analysis, S.P.-L.; investigation, S.P.-L., S.F. and J.-M.M.; resources, S.P.-L.; data curation, S.P.-L., T.L. and J.M.; writing—original draft preparation, S.P.-L., S.F.; writing—review and editing, S.P.-L., S.F., M.C.R.C., J.-M.M., T.L. and J.M.; visualization, S.P.-L.; supervision, S.P.-L.; project administration, S.P.-L.; All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Supported by the French Future Investment Programme (Programme d’Investissement d’Avenir, PIA) through the SWOT Downstream program for hydrological applications and CNES agency.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy of the reservoirs reference data.

Acknowledgments: Reference data from French reservoirs have been collected by M.Blanc (Regional direction DREAL Occitanie, France) and processed by P.Kosuth (Ministry of Ecological Tran-

sition, France). Large-scene reference data has been enriched and curated by UNISTRA-Icube (J.Briant, M.Studer, H.Yesou) and C-S (G.Nicolas). Data processing and validation processes have been optimized by C-S (A.Bricier). The project relies heavily on open-source software. All programming was done in OTB 7.2, Python 3.7, and associated libraries, including Numpy, Pandas and Matplotlib.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Biemans, H.; Haddeland, I.; Kabat, P.; Ludwig, F.; Hutjes, R.W.A.; Heinke, J.; Von Bloh, W.; Gerten, D. Impact of reservoirs on river discharge and irrigation water supply during the 20th century. *Water Resour. Res.* **2011**, *47*, doi:10.1029/2009wr008929.
2. Zhou, T.; Nijssen, B.; Gao, H.; Lettenmaier, D.P. The Contribution of Reservoirs to Global Land Surface Water Storage Variations. *J. Hydrometeorol.* **2016**, *17*, 309–325, doi:10.1175/jhm-d-15-0002.1.
3. Habets, F.; Molénat, J.; Carluet, N.; Douez, O.; Leenhardt, D. The cumulative impacts of small reservoirs on hydrology: A review. *Sci. Total. Environ.* **2018**, *643*, 850–867, doi:10.1016/j.scitotenv.2018.06.188.
4. Blanc, E.; Strobl, E. Is Small Better? A Comparison of the Effect of Large and Small Dams on Cropland Productivity in South Africa. *World Bank Econ. Rev.* **2013**, *28*, 545. Available online: <https://ssrn.com/abstract=2309848> (accessed on 9 August 2021).
5. Acheampong, E.N.; Ozor, N.; Sekyi-Annan, E. Development of small dams and their impact on livelihoods: Cases from northern Ghana. *Afr. J. Agric. Res.* **2014**, *9*, 1867–1877, doi:10.5897/AJAR2014.8610.
6. Pekel, J.-F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422, doi:10.1038/nature20584.
7. Busker, T.; De Roo, A.; Gelati, F.; Schwatke, C.; Adamovic, M.; Bisselink, B.; Pekel, J.-F.; Cottam, A. A global lake and reservoir volume analysis using a surface water dataset and satellite altimetry. *Hydrol. Earth Syst. Sci.* **2019**, *23*, 669–690, doi:10.5194/hess-23-669-2019.
8. McFeeters, S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *Int. J. Remote Sens.* **1996**, *17*, 1425–1432, doi:10.1080/01431169608948714.
9. Xu, H. Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *Int. J. Remote Sens.* **2006**, *27*, 3025–3033, doi:10.1080/01431160600589179.
10. Du, Y.; Zhang, Y.; Ling, F.; Wang, Q.; Li, W.; Li, X. Water Bodies' Mapping from Sentinel-2 Imagery with Modified Normalized Difference Water Index at 10-m Spatial Resolution Produced by Sharpening the SWIR Band. *Remote Sens.* **2016**, *8*, 354, doi:10.3390/rs8040354.
11. Feyisa, G.L.; Meilby, H.; Fensholt, R.; Proud, S. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. *Remote Sens. Environ.* **2014**, *140*, 23–35, doi:10.1016/j.rse.2013.08.029.
12. Acharya, T.D.; Subedi, A.; Lee, D.H. Evaluation of Water Indices for Surface Water Extraction in a Landsat 8 Scene of Nepal. *Sensors* **2018**, *18*, 2580, doi:10.3390/s18082580.
13. Fisher, A.; Flood, N.; Danaher, T. Comparing Landsat water index methods for automated water classification in eastern Australia. *Remote Sens. Environ.* **2016**, *175*, 167–182, doi:10.1016/j.rse.2015.12.055.
14. Yang, X.; Qin, Q.; Yésou, H.; Ledauphin, T.; Koehl, M.; Grussenmeyer, P.; Zhu, Z. Monthly estimation of the surface water extent in France at a 10-m resolution using Sentinel-2 data. *Remote Sens. Environ.* **2020**, *244*, 111803, doi:10.1016/j.rse.2020.111803.
15. Otsu, N. A Threshold Selection Method from Gray-Level Histograms. *IEEE Trans. Syst. Man Cybern.* **1979**, *9*, 62–66, doi:10.1109/bsmc.1979.4310076.
16. Pekel, J.-F.; Vancutsem, C.; Bastin, L.; Clerici, M.; Vanbogaert, E.; Bartholomé, E.; Defourny, P. A near real-time water surface detection method based on HSV transformation of MODIS multi-spectral time series data. *Remote Sens. Environ.* **2014**, *140*, 704–716, doi:10.1016/j.rse.2013.10.008.
17. Pekel, J.F.; Belward, A.; Gorelick, N. Global Water Surface Dynamics: Toward a Near Real Time Monitoring Using Landsat and Sentinel Data. *AGU Fall Meet. Abstr.* **2017**, *2017*, GC13C-0742.
18. Ngoc, D.D.; Loisel, H.; Jamet, C.; Vantrepotte, V.; Duforêt-Gaurier, I.; Minh, C.D.; Mangin, A. Coastal and inland water pixels extraction algorithm (WiPE) from spectral shape analysis and HSV transformation applied to Landsat 8 OLI and Sentinel-2 MSI. *Remote Sens. Environ.* **2019**, *223*, 208–228, doi:10.1016/j.rse.2019.01.024.
19. Doña, C.; Morant, D.; Picazo, A.; Rocheda, C.; Sánchez, J.; Camacho, A. Estimation of Water Coverage in Permanent and Temporary Shallow Lakes and Wetlands by Combining Remote Sensing Techniques and Genetic Programming. Application to the Mediterranean Basin of the Iberian Peninsula. *Remote Sens.* **2021**, *13*, 652, doi:10.3390/rs13040652.
20. Hollstein, A.; Scgl, K.; Guanter, L.; Brell, M.; Enesco, M. Ready-to-Use Methods for the Detection of Clouds, Cirrus, Snow, Shadow, Water and Clear Sky Pixels in Sentinel-2 MSI Images. *Remote Sens.* **2016**, *8*, 666, doi:10.3390/rs8080666.
21. Yousefi, P.; Jalab, I.L.A.; Ibrahim, R.W.; Noor, N.F.M.; Ayub, M.N.; Ganji, A. Water-Body Segmentation in Satellite Imagery Applying Modified Kernel Kmeans. *Malays. J. Comput. Sci.* **2018**, *31*, 143–154, doi:10.22452/mjcs.vol31no2.4.
22. Kaplan, G.; Avdan, U. Object-based water body extraction model using Sentinel-2 satellite imagery. *Eur. J. Remote Sens.* **2017**, *50*, 137–143, doi:10.1080/22797254.2017.1297340.

23. Wieland, M.; Martinis, S. A Modular Processing Chain for Automated Flood Monitoring from Multi-Spectral Satellite Data. *Remote Sens.* **2019**, *11*, 2330, doi:10.3390/rs11192330.
24. Bangira, T.; Alfieri, S.M.; Menenti, M.; Van Niekerk, A. Comparing Thresholding with Machine Learning Classifiers for Mapping Complex Water. *Remote Sens.* **2019**, *11*, 1351, doi:10.3390/rs11111351.
25. Cordeiro, M.C.; Martinez, J.-M.; Peña-Luque, S. Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors. *Remote Sens. Environ.* **2021**, *253*, 112209, doi:10.1016/j.rse.2020.112209.
26. Main-Knorr, M.; Pflug, B.; Louis, J.; Debaecker, V.; Müller-Wilm, U.; Gascon, F. Sen2Cor for Sentinel. In Proceedings of the Image and Signal Processing for Remote Sensing XXIII, Warsaw, Poland, 11–13 September 2017; Volume 10427.
27. Hagolle, O.; Huc, M.; Pascual, D.V.; Dedieu, G. A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENUS, LANDSAT and SENTINEL-2 images. *Remote Sens. Environ.* **2010**, *114*, 1747–1755, doi:10.1016/j.rse.2010.03.002.
28. Liu, K.; Song, C.; Wang, J.; Ke, L.; Zhu, Y.; Zhu, J.; Ma, R.; Luo, Z. Remote Sensing-Based Modelling of the Bathymetry and Water Storage for Channel-Type Reservoirs Worldwide. *Water Resour. Res.* **2020**, *56*, 027147, doi:10.1029/2020wr027147.
29. Evans, T.L.; Costa, M.; Telmer, K.; Silva, T. Using ALOS/PALSAR and RADARSAT-2 to Map Land Cover and Seasonal Inundation in the Brazilian Pantanal. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2010**, *3*, 560–575, doi:10.1109/jstars.2010.2089042.
30. Chapman, B.; McDonald, K.; Shimada, M.; Rosenqvist, A.; Schroeder, R.; Hess, L. Mapping Regional Inundation with Spaceborne L-Band SAR. *Remote Sens.* **2015**, *7*, 5440, doi:10.3390/rs70505440.
31. Brisco, B.; Short, N.; van der Sanden, J.J.; Landry, R.; Raymond, D. A semi-automated tool for surface water mapping with RADARSAT-Can. *J. Remote Sens.* **2009**, *35*, 336–344, doi:10.5589/m09-025.
32. Matgen, P.; Hostache, R.; Schumann, G.; Pfister, L.; Hoffmann, L.; Savetije, H. Towards an automated SAR-based flood monitoring system: Lessons learned from two case studies. *Phys. Chem. Earth Parts A/B/C* **2011**, *36*, 241–252, doi:10.1016/j.pce.2010.12.009.
33. Pham-Duc, B.; Prigent, C.; Aires, F. Surface Water Monitoring within Cambodia and the Vietnamese Mekong Delta over a Year, with Sentinel-1 SAR Observations. *Water* **2017**, *9*, 366, doi:10.3390/w9060366.
34. Clement, M.; Kilsby, C.; Moore, P. Multi-temporal synthetic aperture radar flood mapping using change detection. *J. Flood Risk Manag.* **2018**, *11*, 152–168, doi:10.1111/jfr3.12303.
35. Huang, W.; De Vries, B.; Huang, C.; Lang, M.W.; Jones, J.W.; Creed, I.F.; Carroll, M.L. Automated Extraction of Surface Water Extent from Sentinel-1 Data. *Remote Sens.* **2018**, *10*, 797, doi:10.3390/rs10050797.
36. Bolanos, S.; Stiff, D.; Brisco, B.; Pietroniro, A. Operational Surface Water Detection and Monitoring Using Radarsat. *Remote Sens.* **2016**, *8*, 285, doi:10.3390/rs8040285.
37. Hardy, A.; Ehrlich, G.; Cross, D.E.; Bunting, P.; Liywalii, F.; Sakala, J.; Sulumesi, A.; Singini, D.; Smith, M.; Willis, T.; et al. Automatic Detection of Open and Vegetated Water Bodies Using Sentinel 1 to Map African Malaria Vector Mosquito Breeding Habitats. *Remote Sens.* **2019**, *11*, 593, doi:10.3390/rs11050593.
38. Huth, J.; Gessner, U.; Klein, I.; Yesou, H.; Lai, X.; Oppelt, N.; Kuenzer, C. Analyzing Water Dynamics Based on Sentinel-1 Time Series—A Study for Dongting Lake Wetlands in China. *Remote Sens.* **2020**, *12*, 1761, doi:10.3390/rs12111761.
39. O'Grady, D.; Leblanc, M.; Gillieson, D. Use of ENVISAT ASAR Global Monitoring Mode to complement optical data in the mapping of rapid broad-scale flooding in Pakistan. *Hydrol. Earth Syst. Sci.* **2011**, *15*, 3475–3494, doi:10.5194/hess-15-3475-2011.
40. Manjushree, P.; Kumar, L.; Bhatt, C.M.; Rao, G.S.; Bhanumurthy, V. Optimization of threshold ranges for rapid flood inundation mapping by evaluating backscatter profiles of high incidence angle SAR images. *Int. J. Disaster Risk Sci.* **2012**, *3*, 113–122, doi:10.1007/s13753-012-0011-5.
41. White, L.; Brisco, B.; Pregitzer, M.; Tedford, B.; Boychuk, L. RADARSAT-2 Beam Mode Selection for Surface Water and Flooded Vegetation Mapping. *Can. J. Remote Sens.* **2014**, *40*, 135–151, doi:10.1080/07038992.2014.943393.
42. Li, J.; Wang, S. An automatic method for mapping inland surface waterbodies with Radarsat-2 imagery. *Int. J. Remote Sens.* **2015**, *36*, 1367–1384, doi:10.1080/01431161.2015.1009653.
43. Martinis, S.; Twele, A.; Voigt, S. Unsupervised Extraction of Flood-Induced Backscatter Changes in SAR Data Using Markov Image Modeling on Irregular Graphs. *IEEE Trans. Geosci. Remote Sens.* **2011**, *49*, 251–263, doi:10.1109/tgrs.2010.2052816.
44. Martinis, S.; Plank, S.; Ćwik, K. The Use of Sentinel-1 Time-Series Data to Improve Flood Monitoring in Arid Areas. *Remote Sens.* **2018**, *10*, 583, doi:10.3390/rs10040583.
45. Bioresita, F.; Puissant, A.; Stumpf, A.; Malet, J.-P. A Method for Automatic and Rapid Mapping of Water Surfaces from Sentinel-1 Imagery. *Remote Sens.* **2018**, *10*, 217, doi:10.3390/rs10020217.
46. Chini, M.; Hostache, R.; Giustarini, L.; Matgen, P. A Hierarchical Split-Based Approach for Parametric Thresholding of SAR Images: Flood Inundation as a Test Case. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6975–6988, doi:10.1109/tgrs.2017.2737664.
47. Senthilnath, J.; Shenoy, H.V.; Rajendra, R.; Omkar, S.N.; Mani, V.; Diwakar, P.G. Integration of speckle de-noising and image segmentation using Synthetic Aperture Radar image for flood extent extraction. *J. Earth Syst. Sci.* **2013**, *122*, 559–572, doi:10.1007/s12040-013-0305-z.
48. Hornacek, M.; Wagner, W.; Sabel, D.; Truong, H.-L.; Snoeij, P.; Hahmann, T.; Diedrich, E.; Doubkova, M. Potential for High Resolution Systematic Global Surface Soil Moisture Retrieval via Change Detection Using Sentinel. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 1303–1311, doi:10.1109/jstars.2012.2190136.
49. Martinez, J.-M.; Le Toan, T. Mapping of flood dynamics and spatial distribution of vegetation in the Amazon floodplain using multitemporal SAR data. *Remote Sens. Environ.* **2007**, *108*, 209–223, doi:10.1016/j.rse.2006.11.012.

50. Twele, A.; Cao, W.; Plank, S.; Martinis, S. Sentinel-1-based flood mapping: A fully automated processing chain. *Int. J. Remote Sens.* **2016**, *37*, 2990–3004, doi:10.1080/01431161.2016.1192304.
51. Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. *Remote Sens. Environ.* **2016**, *187*, 156–168, doi:10.1016/j.rse.2016.10.010.
52. Masse, A. Product User Manual: Water Bodies Sentinel-2 100M v1, Copernicus Global Land Operations—Cryosphere and Water, CGLOPS-2. Available online: https://land.copernicus.eu/global/sites/cgls.vito.be/files/products/CGLOPS2_PUM_WB100m_V1_IL10.pdf (accessed on 9 August 2021).
53. Sentinel-2 User Handbook. Available online: https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook (accessed on 9 August 2021).
54. Sentinel-2 ESA Special Publication. Available online: https://sentinel.esa.int/documents/247904/349490/52_SP-1322_2.pdf (accessed on 9 August 2021).
55. Sentinel-2 ESA—Revisit and Coverage. Available online: <https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/revisit-coverage> (accessed on 9 August 2021).
56. Rouquié, B.; Hagolle, O.; Bréon, F.-M.; Boucher, O.; Desjardins, C.; Rémy, S. Using Copernicus Atmosphere Monitoring Service Products to Constrain the Aerosol Type in the Atmospheric Correction Processor MAJA. *Remote Sens.* **2017**, *9*, 1230, doi:10.3390/rs9121230.
57. Baetens, L.; Desjardins, C.; Hagolle, O. Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure. *Remote Sens.* **2019**, *11*, 433, doi:10.3390/rs11040433.
58. The Theia Data Center. Available online: <https://theia.cnes.fr/atdistrib/rocket/#/search?collection=VENUS> (accessed on 21 September 2020).
59. Sentinel-1 ESA Special Publication. Available online: https://sentinel.esa.int/documents/247904/349449/51_SP-1322_1.pdf (accessed on 9 August 2021).
60. Nobre, A.; Cuartas, L.; Hodnett, M.; Rennó, C.; Rodrigues, G.; Silveira, A.; Waterloo, M.; Saleska, S. Height Above the Nearest Drainage—A hydrologically relevant new terrain model. *J. Hydrol.* **2011**, *404*, 13–29, doi:10.1016/j.jhydrol.2011.03.051.
61. Yamazaki, D.; Koshima, D.; Sosa, J.; Bates, P.D.; Allen, G.H.; Pavelsky, T.M. MERIT Hydro: A High-Resolution Global Hydrography Map Based on Latest Topography Dataset. *Water Resour. Res.* **2019**, *55*, 5053–5073, doi:10.1029/2019wr024873.
62. Dozier, J. Spectral signature of alpine snow cover from the Landsat thematic mapper. *Remote Sens. Environ.* **1989**, *28*, 9–22, doi:10.1016/0034-4257(89)90101-6.
63. Khan, J.U.; Ansary, N.; Durand, F.; Testut, L.; Ishaque, M.; Calmant, S.; Krien, Y.; Islam, A.S.; Papa, F. High-Resolution Intertidal Topography from Sentinel-2 Multi-Spectral Imagery: Synergy between Remote Sensing and Numerical Modeling. *Remote Sens.* **2019**, *11*, 2888, doi:10.3390/rs11242888.
64. Wang, X.; Xie, S.; Zhang, X.; Chen, C.; Guo, H.; Du, J.; Duan, Z. A robust Multi-Band Water Index (MBWI) for automated extraction of surface water from Landsat 8 OLI imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2018**, *68*, 73–91, doi:10.1016/j.jag.2018.01.018.
65. Koleck, T. S1 Tiling Chain—Ortho-Rectification of Sentinel-1 Data on Sentinel-2 Grid. Available online: <https://gitlab.orfeo-toolbox.org/s1-tiling/s1tiling> (accessed on 9 August 2021).
66. Grizonnet, M.; Michel, J.; Poughon, V.; Inglada, J.; Savinaud, M.; Cresson, R. Orfeo ToolBox: Open-source processing of remote sensing images. *Open Geospat. Data, Softw. Stand.* **2017**, *2*, 15, doi:10.1186/s40965-017-0031-6.
67. Gascon, F.; Bouzinac, C.; Thépaut, O.; Jung, M.; Francesconi, B.; Louis, J.; Lonjou, V.; Lafrance, B.; Massera, S.; Gaudel-Vacaresse, A.; et al. Copernicus Sentinel-2A Calibration and Products Validation Status. *Remote Sens.* **2017**, *9*, 584, doi:10.3390/rs9060584.
68. NASA Shuttle Radar Topography Mission. Available online: <http://www2.jpl.nasa.gov/srtm/> (accessed on 9 August 2021).
69. Pena Luque Santiago, 2019. CNES ALCD Open Water Masks (Version 1.1); Zenodo. Available online: <https://zenodo.org/record/3675333/files/YR4dmd8RWUk> (accessed on 9 August 2021).
70. Wessel, P.; Smith, W.H.F. A global, self-consistent, hierarchical, high-resolution shoreline database. *J. Geophys. Res. Space Phys.* **1996**, *101*, 8741–8743, doi:10.1029/96jb00104.
71. Sandre; OFB; IGN. BD Topage: French Hydrographic Database. Available online: <https://geo.data.gouv.fr/en/datasets/237d2617f3377a6b74187a17adc83ee948619b9e> (accessed on 9 August 2021).
72. Catry, T.; Li, Z.; Roux, E.; Herbretreau, V.; Gurgel, H.; Mangeas, M.; Seyler, F.; Dessay, N. Wetlands and Malaria in the Amazon: Guidelines for the Use of Synthetic Aperture Radar Remote-Sensing. *Int. J. Environ. Res. Public Health* **2018**, *15*, 468, doi:10.3390/ijerph15030468.
73. Biswas, N.K.; Hossain, F.; Bonnema, M.; Lee, H.; Chishtie, F. Towards a global Reservoir Assessment Tool for predicting hydrologic impacts and operating patterns of existing and planned reservoirs. *Environ. Model. Softw.* **2021**, *140*, 105043, doi:10.1016/j.envsoft.2021.105043.
74. Rüttich, M.; Martinis, S.; Wieland, M. Automatic Flood Duration Estimation Based on Multi-Sensor Satellite Data. *Remote Sens.* **2020**, *12*, 643, doi:10.3390/rs12040643.
75. Eyler, B. Mekong Dam Monitor: Methods and Processes. Available online: <https://www.stimson.org/2020/mekong-dam-monitor-methods-and-processes/> (accessed on 9 August 2021).

76. Yang, X.; Qin, Q.; Grussenmeyer, P.; Koehl, M. Urban surface water body detection with suppressed built-up noise based on water indices from Sentinel-2 MSI imagery. *Remote Sens. Environ.* **2018**, *218*, 259–270, doi:10.1016/j.rse.2018.09.016.
77. Harmel, T.; Chami, M. Estimation of the sunglint radiance field from optical satellite imagery over open ocean: Multidirectional approach and polarization aspects. *J. Geophys. Res. Ocean.* **2013**, *118*, 76–90, doi:10.1029/2012jc008221.
78. Alsdorf, D.E.; Rodriguez, E.; Lettenmaier, D.P. Measuring surface water from space. *Rev. Geophys.* **2007**, *45*, doi:10.1029/2006eg000197.
79. Slagter, B.; Tsendbazar, N.-E.; Vollrath, A.; Reiche, J. Mapping wetland characteristics using temporally dense Sentinel-1 and Sentinel-2 data: A case study in the St. Lucia wetlands, South Africa. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *86*, 102009, doi:10.1016/j.jag.2019.102009.
80. Alaska Satellite Facility. Sentinel-1 Data Coverage Heat Maps. Available online: <https://asf.alaska.edu/data-sets/sar-data-sets/sentinel-1/sentinel-1-acquisition-maps/> (accessed on 9 August 2021).
81. Cooley, S.W.; Ryan, J.C.; Smith, L.C. Human alteration of global surface water storage variability. *Nat. Cell Biol.* **2021**, *591*, 78–81, doi:10.1038/s41586-021-03262-3.
82. C. Donlon et al., The Global Monitoring for Environment and Security (GMES) Sentinel-3 mission. *Remote Sens. Environ.* **2012**, *120*, 37–57, doi:10.1016/j.rse.2011.07.024.
83. Donlon, C.J.; Cullen, R.; Gialicchi, L.; Vuilleumier, P.; Francis, C.R.; Kuschnerus, M.; Simpson, W.; Bouridah, A.; Caleno, M.; Bertoni, R.; et al. The Copernicus Sentinel-6 mission: Enhanced continuity of satellite sea level measurements from space. *Remote Sens. Environ.* **2021**, *258*, 112395, doi:10.1016/j.rse.2021.112395.
84. Biancamaria, S.; Lettenmaier, D.P.; Pavelsky, T.M. The SWOT Mission and Its Capabilities for Land Hydrology. *Surv. Geophys.* **2016**, *37*, 307–337, doi:10.1007/s10712-015-9346-y.

II. Surface Water Dynamics from Space: A Round Robin Intercomparison of
Using Optical and SAR High-Resolution Satellite Observations for Regional
Surface Water Detection

Article

Surface Water Dynamics from Space: A Round Robin Intercomparison of Using Optical and SAR High-Resolution Satellite Observations for Regional Surface Water Detection

Christian Tottrup ^{1,*}, Daniel Druce ¹, Rasmus Probst Meyer ¹, Mads Christensen ¹, Michael Riffler ², Bjoern Dulleck ², Philipp Rastner ², Katerina Jupova ³, Tomas Sokoup ³, Arjen Haag ^{4,5}, Mauricio C. R. Cordeiro ⁶, Jean-Michel Martinez ⁶, Jonas Franke ⁷, Maximilian Schwarz ⁸, Victoria Vanthof ⁹, Suxia Liu ^{10,11}, Haowei Zhou ^{10,11}, David Marzi ¹², Rudiyanto Rudiyanto ¹³, Mark Thompson ¹⁴, Jens Hiestermann ¹⁴, Hamed Alemohammad ¹⁵, Antoine Masse ¹⁶, Christophe Sannier ¹⁶, Sonam Wangchuk ¹⁷, Guy Schumann ¹⁸, Laura Giustarini ¹⁸, Jason Hallowes ¹⁹, Kel Markert ⁵ and Marc Paganini ²⁰



Citation: Tottrup, C.; Druce, D.; Meyer, R.P.; Christensen, M.; Riffler, M.; Dulleck, B.; Rastner, P.; Jupova, K.; Sokoup, T.; Haag, A.; et al. Surface Water Dynamics from Space: A Round Robin Intercomparison of Using Optical and SAR High-Resolution Satellite Observations for Regional Surface Water Detection. *Remote Sens.* **2022**, *14*, 2410. <https://doi.org/10.3390/rs14102410>

Academic Editor: Vito Pascazio

Received: 7 April 2022

Accepted: 10 May 2022

Published: 17 May 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

- ¹ DHI A/S, 2970 Hørsholm, Denmark; dadr@dhigroup.com (D.D.); rame@dhigroup.com (R.P.M.); mads@dhigroup.com (M.C.)
- ² GeoVille GmbH, 6020 Innsbruck, Austria; riffler@geoville.com (M.R.); dulleck@geoville.com (B.D.); rastner@geoville.com (P.R.)
- ³ Gisat s.r.o., 170 00 Praha, Czech Republic; katerina.jupova@gisat.cz (K.); tomas.sokoup@gisat.cz (T.S.)
- ⁴ Deltares, 2629 HV Delft, The Netherlands; arjen.haag@deltares.nl
- ⁵ SERVIR-Mekong, Bangkok 10400, Thailand; kel.markert@nasa.gov
- ⁶ Géosciences Environnement Toulouse (GET), Unité Mixte de Recherche 5563, IRD/CNRS/Université, 31400 Toulouse, France; mauricio@ana.gov.br (M.C.R.C.); jean-michel.martinez@ird.fr (J.-M.M.)
- ⁷ Remote Sensing Solutions GmbH, 81673 München, Germany; franke@rsggmbh.de
- ⁸ Department of Biology, Ludwig-Maximilians-University Munich, 82152 Planegg-Martinsried, Germany; schwarz@rsggmbh.de
- ⁹ Faculty of Environment, University of Waterloo, Waterloo, ON N2L 3G1, Canada; vrvantho@uwaterloo.ca
- ¹⁰ Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China; liusx@igsnrr.ac.cn (S.L.); zhouhw.18b@igsnrr.ac.cn (H.Z.)
- ¹¹ College of Resources and Environment, Sino-Danish Center, University of Chinese Academy of Sciences, Beijing 100049, China
- ¹² Department of Electrical, Computer and Biomedical Engineering, University of Pavia, 27100 Pavia, Italy; david.marzi01@universitadipavia.it
- ¹³ Program of Crop Science, Faculty of Fisheries and Food Science, Universiti Malaysia Terengganu, Kuala Nerus 21030, Terengganu, Malaysia; rudiyanto@umt.edu.my
- ¹⁴ GeoTerraImage (Pty) Ltd., Pretoria 0184, South Africa; mark.thompson@geoterraimage.com (M.T.); jens.hiestermann@geoterraimage.com (J.H.)
- ¹⁵ Radiant Earth Foundation, Washington, DC 20005, USA; hamed@radiant.earth
- ¹⁶ Group CLS, 31400 Toulouse, France; amasse@groupcls.com (A.M.); csannier@groupcls.com (C.S.)
- ¹⁷ Faculty of Science, Vrije Universiteit Amsterdam, 1081 HV Amsterdam, The Netherlands; s.wangchuk@vu.nl
- ¹⁸ RSS-Hydro SARLS, 3593 Dudelange, Luxembourg; gschumann@rss-hydro.lu (G.S.); lgiustarini@rss-hydro.lu (L.G.)
- ¹⁹ EkoSource Insight (Pty) Ltd., Johannesburg 2196, South Africa; jhallowes@ekosource.co.za
- ²⁰ European Space Agency, ESRIN, 00044 Frascati, Italy; marc.paganini@esa.int
- * Correspondence: cto@dhigroup.com

Abstract: Climate change, increasing population and changes in land use are all rapidly driving the need to be able to better understand surface water dynamics. The targets set by the United Nations under Sustainable Development Goal 6 in relation to freshwater ecosystems also make accurate surface water monitoring increasingly vital. However, the last decades have seen a steady decline in in situ hydrological monitoring and the availability of the growing volume of environmental data from free and open satellite systems is increasingly being recognized as an essential tool for largescale monitoring of water resources. The scientific literature holds many promising studies on satellite-based surface-water mapping, but a systematic evaluation has been lacking. Therefore, a round robin exercise was organized to conduct an intercomparison of 14 different satellite-based approaches for monitoring inland surface dynamics with Sentinel-1, Sentinel-2, and Landsat 8 imagery. The objective

was to achieve a better understanding of the pros and cons of different sensors and models for surface water detection and monitoring. Results indicate that, while using a single sensor approach (applying either optical or radar satellite data) can provide comprehensive results for very specific localities, a dual sensor approach (combining data from both optical and radar satellites) is the most effective way to undertake largescale national and regional surface water mapping across bioclimatic gradients.

Keywords: surface water dynamics; SAR and optical data; data fusion; water resource management; Sustainable Development Goal 6

1. Introduction

Water is key to sustainable development, being critical for socioeconomic development, energy and food production, and healthy ecosystems. Today water scarcity affects more than 40 percent of the world's population and is projected to rise further, exacerbated by climate change [1]. As the global population grows, there is an increasing need to balance the competing demands for water resources and have more efficient ways to manage water supply. The importance of ensuring the availability and sustainable management of water for all has been increasingly addressed in the global political agenda, as seen with the Sixth Sustainable Development Goal (SDG) of the United Nations 2030 Agenda for Sustainable Development [2] and the adoption of an International Decade 2018–2028 for Action on 'Water for Sustainable Development' by the UN General Assembly [3]. As the demand for freshwater increases, the importance of monitoring changes in surface waters is gaining more attention, but many countries are still lacking data to monitor the extent of their inland waters and their intra- and interannual changes.

Earth Observation (EO) is an essential source of information, which can complement national hydrometric data and services and support countries to operationally monitor changes to their surface waters. Ever since the launch of the first Earth observation satellites in the early 1970s, the mapping and monitoring of surface water has been a subject that attracts interest from researchers and practitioners in hydrology, environmental conservation, and water resource management. The field has gradually evolved and been incentivized by the steady buildup of long-term archives of global satellite data and computer resources for analyzing those data. A significant breakthrough in the adoption of EO solutions for water-related topics has been the European Commission Joint Research Center's Global Surface Water Explorer [JRC-GSWE] [4] and the Global Land Analysis and Discovery Group's Global Surface Water Dynamics [GLAD-GSWD] [5]. Despite these developments and the long track record of related successful case studies on surface water mapping, there is still a lack of clear, robust, efficient, user-oriented methods and guidelines that allow for the use of EO data at scale and on an operational basis for surface water mapping and monitoring.

The mapping of surface water with either optical or Synthetic Aperture Radar (SAR) data has been reviewed in several papers (e.g., [6,7]) and with a series of more recent papers focusing on the combined use of optical and SAR data [8–11]. This development is directly related to the Sentinel program under the European Copernicus initiative [12]. Through the Copernicus Sentinel mission, optical and SAR data in high resolution (10 m) have become globally available free of charge and with a short latency of a few days or less. The next leap in EO-based surface water detection will need to take full advantage of this enhanced observation capacity, which offers unprecedented opportunities to develop robust and cost-effective EO methods to monitor the seasonal and annual variations of surface waters. These EO methods and associated information products can be embedded in national processes for more evidence-based water policies and efficient reporting on the global water agenda. This is why the European Space Agency (ESA) has launched the WorldWater project with a principal aim of strengthening EO capacities in countries to better monitor their inland waterbodies (lakes, reservoirs, rivers, and estuaries) and, consequently, better

fulfil their commitments on water resource management and water security in the different water-related global agendas such as the 2030 Agenda on Sustainable Development [2], the 2015 Paris Agreement on climate change [13] and the Sendai Framework for Disaster Risk Reduction [14].

The overarching goal of the WorldWater project is to develop robust and scalable EO solutions for inland surface water monitoring, which can be exploited by a large community of stakeholders involved in water management from local water supplies to national water strategies, including transboundary river basin management plans and global assessment of surface water changes. As part of the project goal, a round robin exercise has been organized to conduct an intercomparison of EO algorithms for surface water detection, using the latest generation of free and open satellite data from Sentinel-1, Sentinel-2, and Landsat 8. The round robin was open to researchers, companies, and other developers of satellite-based algorithms for surface water detection. The precondition for participating in the round robin was a peer-reviewed algorithm for surface water detection based on (or adaptable to) Sentinel-1, Sentinel-2, and/or Landsat 8. Non-peer reviewed algorithms were accepted provided that adequate supplementary documentation and justification could be provided. In this paper, we present the results of the WorldWater round robin intercomparison and use them as the basis for discussing the pros and cons of different approaches to detect and monitor surface waters from Earth observation data. By using various statistical tests, we evaluate the quantitative performance of the individual algorithms and use the findings to draw some qualitative considerations about their performance. The focus is not on the algorithms themselves, as they have already proved themselves (cf. peer-reviewed or in an operational setting), but rather, on the underlying data model, that is, whether the algorithms are relying on single sensor inputs or whether they are using a dual sensor approach. Ideally, the best performing algorithms can provide spatially and temporally consistent timeseries of surface water extent dynamics that meet the user requirements, not only in terms of accuracy but also in terms of transparency, cost, and transferability. The aim is to contribute to the development of a new set of best practices for surface water monitoring, as well as identifying shortfalls and areas of further research.

2. Materials and Methods

2.1. Test Sites and Input Data

All participants in the round robin were required to produce monthly maps of inland, open surface waters at 10-m spatial resolution for 2 consecutive years over three test sites (100 × 100 km) located in 3 different countries: Colombia, Mexico, and Zambia. Optionally, participants could also submit results for an additional two test sites located in Gabon and Greenland (cf. Figure 1). Test site locations were selected to cover various eco-and climatic regions as well as to include major challenges for EO-based surface water mapping, including sites influenced by topography, clouds, canopy shading, fire scars, urban areas, and regions with permanent low backscatter (e.g., flat and impervious areas, sandy surfaces). The sites also included a diversity of waterbodies ranging from large waterbodies (wind and wave effects) to smaller waterbodies of both a permanent and seasonal nature, as well as waterbodies impacted by water constituents and shallow waters influenced by bottom reflectance. The input datasets, made available to all participants, included all Sentinel 1, Sentinel 2, and Landsat 8 images acquired over the test sites from July 2018 to June 2020. Use of ancillary datasets (such as Digital Elevation Model (DEMs) and a priori surface water maps) were allowed, but under the condition they were publicly available, e.g., the Copernicus DEM [15] and JRC-GSWE [4].



Figure 1. Location map of the test sites annotated with their dominant eco-region(s).

2.2. Surface Water Detection Models

The following sections provide a high-level summary of the fourteen contributions to the round robin intercomparison. Each contribution is referred to as a model in order to emphasize that the focus on the intercomparison was to evaluate the performance of the underlying data models, i.e., whether the surface water detection was based on optical data only (O), SAR data only (S), or integration of both optical and SAR data (O + S).

Model A [O + S] uses a histogram segmentation method to separate imagery from Sentinel-1, Sentinel-2, and Landsat 8 into water and non-water classes [16,17]. Specifically, it carries out edge detection followed by procedures to help obtain a bimodal distribution on which Otsu's method is carried out to automatically derive an optimal threshold. This model was specifically designed for fast and largescale water detection to assist in flood relief efforts. Similar methods exist that attempt to obtain local thresholds over small sections of each image [18], which can potentially yield more accurate results but at the expense of computational speed. A postprocessing step is applied on the monthly water maps derived separately from optical and SAR imagery, where water pixels are constrained to areas that are hydrologically likely to contain water, with the full timeseries of maps derived from optical imagery included as an additional constraint for the SAR-derived maps. Finally, the optical and SAR-based maps are merged to produce a single water map per month.

Model B [O + S] This surface water detection approach is based on Sentinel-2 imagery as the primary water detection dataset, with the all-weather capabilities of Sentinel-1 SAR imagery being used to "fill-in" cloud-obscured water surfaces. SAR data "in-filling" was restricted to raster cells previously detected as having recorded a surface water content from longer-term data modelling results (circa 2016 and forwards) in order to minimize SAR-generated commission errors in the target month. The water surface modelling procedure is based on a set of decision-tree-generated rules that have been derived from a comprehensive set of water and non-water feature reference points distributed across South Africa. The reference dataset consists of $\pm 60,000$ sample points that represent a wide range of seasonal and geographical variations in both water (i.e., turbidity, depth) and non-water surface conditions with potentially similar spectral characteristics, such

as burn scars, terrain shadows, and dark, non-vegetated surfaces from both natural and man-made environments. Collectively, these points ensure full representation of all spectral characteristics required in the water detection modelling process. The monthly surface water datasets represent the median surface water extent for that month, rather than the average or (absolute) maximum extent, as a result of the multirate image acquisition date compositing approach used to model water features [19,20].

Model C [O + S] uses a random forest classifier to map surface waterbodies pixel by pixel by taking advantage of the strength of both optical and SAR data in an integrated manner [21]. For optical data, the model relies on a maximum value of the NDWI composite created using both Level-1 and L-2 Sentinel-2 data. The model depends on a minimum radar backscatter intensity, from both VV and VH polarizations, of a composite for sentinel-1 SAR data. Relying composite images minimizes disturbances from clouds, turbidity, and shadows for the optical data and speckles, lake ice, and radar shadows for the SAR data. The model also uses DEM as a feature to remove false positives over a steeper terrain. All the workflows are implemented in Google Earth Engine for ease of transferability and reproducibility.

Model D [O + S] applied a combined histogram-thresholding and edge-detection approach to estimate monthly surface water extent from monthly, cloud-free Sentinel-1, Sentinel-2, and Landsat-8 scenes. Following cloud masking for optical scenes, we applied the Edge-Otsu algorithm to create binary land and water maps for each scene [17,22]. For a complete description and application of the Edge algorithm, see Markert et al., 2020. To initially segment water, histogram-thresholding was performed using the Normalized Difference Water Index (NDWI) index for optical scenes and the VV-median band for SAR scenes within already buffered surface water polygons from Pekel et al., 2016. A second segmentation was applied to full scenes to segment water and non-water, irrespective of initial water polygons. The MERIT DEM [23] was then used to derive a Height Above Nearest Drainage (HAND) model [24] and on regions less than 30 m in height relative to the nearest drainage. Final monthly surface water products combined both optical and SAR water maps by selecting the optical land–water prediction when available, and otherwise selecting the SAR-identified water pixel. Given that cloud-free optical images segment water with higher accuracy than SAR, this approach reduces error during less cloudy periods.

Model E [S] is a fully automated approach that uses dynamic thresholds to classify individual Sentinel-1 scenes. The scene-dependent thresholds to classify water are defined through the use of existing geospatial information of permanent water areas, e.g., data from the Global Surface Water Explorer (GSWE) [4]. The S-1 backscatter values of permanent water areas are derived per scene and are then statistically analyzed by using percentiles to eliminate outliers and a combination of mean and standard deviation to define the individual classification threshold. In opposite to a fixed threshold, this standardized statistical approach allows for the definition of dynamic classification thresholds per scene in order to account for variations in backscatter caused by various factors. The individually classified scenes are then combined to monthly surface water composites, in which false positives (mainly radar shadows) are removed by the use of the Multi-resolution Valley Bottom Flatness (MrVBF) index [25] derived from the Copernicus Digital Elevation Model (DEM). The automated, computationally efficient classification approach has been shown to capture seasonal changes in surface water accurately, but also shows some limitations in non-vegetated sandy areas, in which false positives occur.

Model F [O + S] used combinations of monthly percentile composite images from Sentinel-1 and Normalized Difference Vegetation Index (NDVI), Normalized Difference Water Index (NDWI), Land Surface Water Index (LSWI), Normalized Difference Snow Index (NDSI), red, NIR, and SWIR1 bands from the greenest monthly Sentinel-2 images as covariates for the mapping of monthly surface water extent in Colombia, Mexico, Zambia, and Gabon. For Greenland, covariates from Sentinel-1 were excluded and replaced by monthly minimum NDVI from Sentinel-2 [26]. Training datasets (water–non-water) were

generated using a stratified, random sample of points based on Global Surface Water data [4] and visual inspections of spectra profile based on k-means clustering results. Random forest classifier was used for classification.

Model G [S] This approach applies a novel Convolutional Neural Network (CNN) model applied to Sentinel-1 observations to detect surface water. The JRC GSWE product was used as training data, and several finetuning strategies were implemented to improve accuracy of the model in places with complex landcover types. The resulting surface water product has a 10-m spatial resolution, is not impacted by cloud coverage, and can be run in near-real time to detect any surface water changes [27].

Model H [O] uses a thresholding method based on a combination of water indexes (MNDWI > NDVI or MNDWI > EVI) to extract surface water extent from monthly composite Sentinel-2 MSI images. Different from the conventional thresholding method, this algorithm does not need to determine the threshold artificially. To obtain more accurate surface water extent maps, the clouds and cloud shadows pixels, buildup pixels, and snow/ice pixels were removed by auxiliary datasets in preprocessing, and the surface water maps with residual non-water pixels were furtherly denoised in postprocessing. For incomplete monthly surface water extent maps, the surface water frequency map was utilized to fill the gaps caused by clouds and cloud shadows. These methods had been proved effective and accurate in the construction of surface water extent continuous timeseries [28].

Model I [O] uses a multidimensional clustering analysis based on reflectance values and water indices to identify water pixels using optical scenes individually. To achieve high-performance and low memory consumption for high resolution images, this process is applied to a random subsample of the image's pixels and then coupled with a Naïve Bayes classifier responsible for generalizing the results to the complete scene. The advantage of using an unsupervised approach such as clustering is that the water pixels group is identified automatically by comparing it to other clusters (targets) in the scene. Therefore, the algorithm doesn't require ancillary data, pretraining, or any threshold calibration, and it is independent of the sensor and the coverage being analyzed. These ideas make it simple to apply the model to a great variety of conditions [29]. As the original algorithm was designed for operational use on single scenes, the monthly water surface has been derived by combining subsequent water masks through an upvote logic that considers as water those pixels that received at least two votes.

Model J [S] This model is based on an unsupervised k-means-clustering algorithm and aims to extract monthly inland waterbody extents over wide areas using multitemporal Sentinel-1 SAR data. To account for slope-induced backscatter differences caused by hills and mountains, due to the slanted acquisition geometry of SAR systems, the model included a radiometric terrain correction, as this step is not applied in the standard Sentinel-1 preprocessing chain. Moreover, the methodology added a multitemporal speckle noise filter which provides better results than a spatial filter applied independently to each SAR image. Seed points for the k-means model are then retrieved by randomly sampling the water layer of the ESA CCI GlobCover Land Cover map [30]. Each sample is represented by a set of temporal features suitable for water characterization in SAR data, such as the mean backscattering value, the maximum value, minimum value, and four "quarter composites" obtained by averaging in time all the Sentinel-1 acquisitions available within each quarter of a year. After the k-means clustering, applied with $k = 4$, the water cluster is selected by considering a majority voting procedure within the multi-polygon water boundaries of the GlobCover map. Since it is based on SAR data, the methodology can be applied in every weather and lighting condition. Being an unsupervised technique, it is quick, robust, and can be applied automatically over any region of the World [31].

Model L [O] uses the simple yet robust band ratio Normalized Difference Vegetation Index (NDVI) on Sentinel-2 images, screened with the cloud mask processor available in ESA's SNAP software. Despite the rather simplistic nature of the NDVI band ratio algorithm, results reported in other studies of this type are encouraging (e.g., [26]). Furthermore, the aim of choosing this approach was to test the application of simple and fast algorithms

for processing large amounts of images in a short time. We implemented the processing on the Web Advanced Space Developer Interface (WASDI) to process all images without the need for downloading large data quantities on a personal computer [32].

Model M [O + S] uses an efficient and opensource supervised Random Forest classifier system based on Geographic Object-Based Image Analysis (GEOBIA) [33]. It relies on two main components, which are feature extraction based on attribute profiles and a semi-supervised classification using a Random Forest Algorithm. The first step consists of computing features based on Sentinel-2 LIC without cloud detection (MNDWI) and DEM (SRTM or ArcticDEM for Greenland) and extraction of spatial features (object area). The ground truths are automatically extracted from GlobalSurfaceWater data (Pekel et al., 2016). The output from this model is maps of monthly surface water extent and a confidence index. The same automatic system is applied for all 5 test areas.

Model N [O + S] is based on a combination of different image-binarization techniques applied on monthly aggregated Sentinel-1, Sentinel-2, and Landsat-8 imagery. Dynamic, tile-based thresholding [34,35] is conducted on both SAR and optical inputs. In addition, adaptive thresholding [36] and seeded region growing [37] on each initially detected waterbody is performed on the monthly SAR imagery. Finally, fuzzy-logic classification refinement reduces water lookalikes and misclassifications (e.g., radar shadows) from the SAR-based water masks [38,39].

Model O [O + S] uses a multivariate logistic regression model to estimate monthly surface water extent from the combined usages of Sentinel-1, Sentinel-2, and Landsat-8 imagery. Models that rely upon linear distributions are often simpler and generalize well and, therefore, do not require high-quality training labels. Yet, since land-water classification has some nonlinear exceptions, such as clouds, shadows, and snow, the approach integrates logic-based masking to reduce the impact from problematic areas through specific thresholds or basic decision trees. The final approach has proved to be accurate whilst at the same time maintaining computational efficiency and simplicity that facilitates analysis and understanding at scale [8].

2.3. Validation and Evaluation

These water detection models were evaluated individually and in cross-comparison using independent reference data collected from the test sites. A fundamental premise for sound scientific validation is to use reference data of higher quality than the product to be validated. There are two ways to ensure higher quality in the reference data: (i) by using a reference data source with a better resolution than the data used for production (i.e., verification by higher data) and/or (ii) by using a more accurate measurement or interpretation than being used for production (i.e., verification by higher method). A further requirement on the reference data is the ability to provide sufficient spatial and temporal representation to accurately label each unit in the sample; i.e., the ideal reference data are: (i) available for the entire region of interest, (ii) representative of the attribute at the date of interest, and (iii) available at a low cost. The balance between these criteria is often difficult to achieve and why tradeoffs and compromises may be needed when generating the final set of reference data. In the case of the round robin validation, a two-step approach was followed: (i) sample based validation (pixel based) and labelled using the production imagery (verification by higher method) and (ii) object extraction accuracy (area based) and using PlanetScope data as a reference (verification by higher data). The sample-based validation has the advantage of delivering reference data, which can be directly matched (in space and time) to the validation input, whereas the PlanetScope data offer the advantage of better capturing and, hence, better evaluation of smaller and narrower waterbodies/features. Still, the acquisition and interpretation of PlanetScope data is costly, and their representation is therefore restricted in space and time. In a final step, the temporal consistency of the optical, SAR, and dual sensor-mapping approaches were evaluated by comparing the total areal water extent mapped within each test site and

across the monthly timeseries. Each validation and evaluation step is described in more detail below.

2.3.1. Sample Based Validation

Stratified random sampling was used to generate reference points over each 100×100 km test site and within three strata across the land–water continuum: permanent water, seasonal water, and non-water. The three strata were generated from the JRC-GSWE long-term water occurrence and defined according to the following thresholds: permanent water > 90%; $0\% < \text{seasonal water} < 90\%$, and non-water = 0%. In all test sites, the target class “water” is a rare occurrence. In the case of rare occurrences, statistical equations does not allow for proper estimation of sample sizes, but stratified random sampling affords the option to increase the sample size in classes and/or regions that occupy a small proportion of area to thereby help reduce the standard errors of class/region-specific accuracy estimates [40]. It was our aim to ensure a minimum of 50 samples in each stratum, while using subsequent sample size allocations to provide a proportional allocation of samples in better accordance with the actual area of the different strata within each test site. In addition, the expected variance within each stratum was also considered; i.e., the transitional strata are expected to have the highest variance, and why it has a higher sample allocation. Thus, by taking area and expected variance into account, the following sample allocations were applied for the five test sites (cf. Table 1).

Table 1. Sample size allocations for the 5 test sites used in the round robin.

	Colombia		Gabon		Greenland		Mexico		Zambia	
	<i>per month</i>	<i>total</i>	<i>per month</i>	<i>total</i>	<i>per month</i>	<i>total</i>	<i>per month</i>	<i>total</i>	<i>per month</i>	<i>total</i>
Land	140	840	75	450	60	180	140	840	90	540
Transition zone	140	840	150	900	90	270	140	840	190	1140
Water	20	120	60	360	100	300	20	120	40	240
TOTAL	300	1800	285	1710	250	750	300	1800	320	1920

In total, 7,980 samples were allocated across the five test sites and six time periods representing every second month of the year 2019 (January, March, May, July, September, November). Each sample point was assigned to be either water or non-water by two independent and experienced interpreters using blind visual interpretations of monthly Sentinel-1 and -2 composites. To harmonize and achieve consistent reference labelling, a standard validation interface was used to ensure interpreters were looking at same area and using the same reference data and the predefined set of classes. In cases where the interpreters disagreed, a quality manager intervened to seek consensus. If consensus could not be agreed upon, the sample was rejected. For each sample we extracted, the respective water classifications and the final set of samples were used to derive standard metrics for accuracy assessments, i.e., overall accuracy (OA), producer accuracy (PA), and user accuracy (UA). For this analysis, all pixels in the individual round robin classifications not classified as water were considered to be non-water; i.e., the non-water class also included pixels being masked (e.g., due to clouds).

2.3.2. Object Extraction Accuracy

Traditionally, stratified point sampling will, in most instances, under-sample Small Waterbodies (SWB) simply because SWBs only represent a fraction of the total water area, even though they may by far exceed the larger waterbodies in numbers [41]. To deal with the issue of SWBs, we complemented the more conventional stratification, sampling, and confusion matrix-type accuracy assessments with an evaluation of object extraction

accuracy based on area-based accuracy metrics and the use of higher spatial resolution but single date (i.e., time-limited) PlanetScope data. An independent reference dataset was created from the classification and interpretation of imagery from Planet. The acquired data was PlanetScope Level 3B (Ortho Scene Products) in 3-m spatial resolution and with 4 spectral bands (RGB, NIR) (<https://www.planet.com/products/planet-imagery/>, accessed on 10 January 2022). The PlanetScope data was acquired within the coverage of each of the test sites and for two areas of approximately 25 km². The exact coverage was determined by size and type of waterbodies, i.e., covering areas with small waterbodies relative to the test site in general and representing both lakes/reservoirs and streams/rivers. For each PlanetScope coverage, we applied a supervised Gradient Boosting (lightGBM) algorithm [42] to generate water masks using the convolution layers derived from spatial filtering of Planet imagery as the explanatory variables and manually derived training samples for water and land (cf. non-water) as the response variable. The Gradient Boosting typically involved a couple of iterations to optimize results, and before finalization, all water masks were manually checked and corrected to ensure high quality. Once analyzed, the PlanetScope data was used to evaluate the object extraction accuracy of the water classifications derived using Sentinel data.

The accuracy evaluation of object extraction is based on object matching, and we focused on two elements related to this, namely: object matching and area-based accuracy measures [43]. The central idea of object matching is to estimate the maximum overlap area by computing the coincidence degree, A_{max} , between two objects.

$$A_{max} = \frac{1}{2} \left(\frac{A_{C,i} \cap A_{R,j}}{A_{C,i}} + \frac{A_{C,i} \cap A_{R,j}}{A_{R,j}} \right)$$

where $A_{C,i}$ denotes the area of the i th-evaluated object, $A_{R,j}$ is the area of the j th reference object, and $A_{C,i} \cap A_{R,j}$ represents the intersection area. For an evaluated object and candidate reference objects, each coincidence degree will be computed. Two objects will be judged as being a matching pair if the area of the coincidence degree is at a maximum, i.e., A_{max} equals 1.

The maximum overlap object matching is complemented by three area-based accuracy measures (i.e., correctness, completeness, and quality). Correctness (A_{cor}) is defined as the ratio of the correctly extracted area (A_C) and the whole extracted area (A_{DC}), whereas completeness (A_{com}) refers to the ratio of the correctly extracted area to the reference area (A_{RC}). The range of correctness and completeness is 0 to 1. If A_C fully corresponds to A_{DC} or A_{RC} , then the value is 1. If there is no overlap between A_C and A_{DC} or A_{RC} , then the value is 0; correctness and completeness interact. For instance, a large A_{DC} leads to a small correctness value, while a small A_{RC} results in a large completeness value. To amend this issue, the quality A_{qual} is designed to provide a measure of quality by balancing correctness and completeness.

$$A_{qual} = \frac{A_C}{A_{DC} + A_{RC} - A_C}$$

The range of quality is 0 to 1. If the water extraction results are the same as the reference data, then the value is 1. If none of the extracted water area overlaps with the reference area, then the value is 0. The advantage of area-based accuracy measures compared to the sample-based validation relates to the fact that the confusion matrix of the latter depends on total pixel number. In contrast, the evaluation results for two cases using area-based accuracy measures are equivalent because it relies only on the evaluation, and reference objects are independent of the total pixel number.

2.3.3. Temporal Consistency Evaluation

The purpose of temporal consistency evaluation is to identify anomalies in sequences of surface water maps. Sudden decreases in surface water can be due not only to drought and high reservoir release but also clouds and lack of valid observation. Flooding, on

the other hand, may cause an increase in surface water, but so could cloud shadows and topographic shading, as well as the impact of low-backscatter areas. More robust water detection algorithms should be able to accurately capture actual water dynamics while minimizing the influence of the other factors.

3. Results

3.1. Water Occurrence

The five test sites used for intercomparison represent very different conditions, which can also be inferred by looking at multiannual water occurrence maps for the respective test sites (cf. Figure 2). As explained in Section 2.1, site variability is, on the one hand, dictated by geographic location (i.e., from tropical to arctic, coastal to inland, and lowland to high land) and, on the other hand, by surface water characteristics. The latter is clearly illustrated in the water occurrence maps, which show the differences between test sites in terms of size and type of waterbodies, as well as the relative distribution of permanent and seasonal water (Figure 2). These different characteristics are important to bear in mind when interpreting the validation results, as they will influence the performance of the individual algorithm.

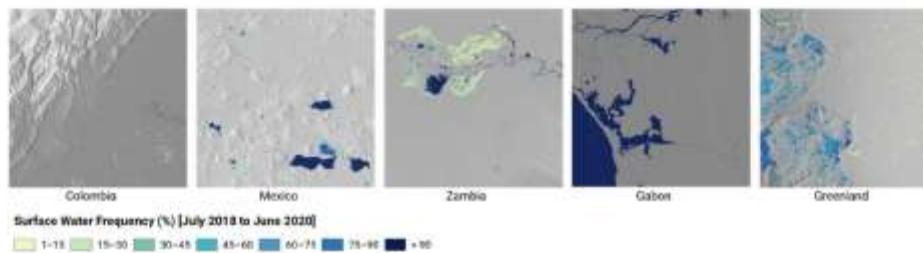


Figure 2. Examples of surface water frequency maps over the 5 test sites and as derived by Model N.

3.2. Sample Based Validation

In Table 51, we provide classification accuracies for the water extraction for all round robin submissions and for each of the three mandatory sites, as well as the optional sites, where relevant. The general performance of all models can be deemed satisfactory, with overall accuracies above or near 90% when looking across the mandatory sites. There is more ambiguity when looking at the performance in terms of user and producer accuracy and at the level of the individual sites.

In Figure 3, the classification accuracies have been grouped (median value) by input data type, i.e., algorithms using both optical and SAR vs. models based on single-sensor inputs (SAR or optical). Figure 3 shows an overall better performance of the combined sensor approach compared to single sensor approaches, although the results are not one-sided when looking at the individual sites or in terms of user and producer accuracies. In Colombia, the combined sensor approach performed best in terms of overall accuracy, but, at the level of UA and PA, the SAR and Optical models, respectively, outperform the combined approach. In Gabon, the SAR approach outperforms the other data models in terms of OA, while in Colombia and Zambia, the optical approach has much higher accuracies for, respectively, PA and UA. In Mexico, OA and UA are almost equal between the data models, but with a noteworthy (+4–5 percentage) drop in producer accuracy for the optical data models compared to the SAR and dual sensor models. The observed differences in UA and PA are closely related to site-specific characteristics. For example, the higher UA accuracies achieved in Gabon and Colombia using SAR are an indication of the benefit SAR adds in a cloud-prone region. In contrast, SAR produces a lower UA in Zambia and Mexico because of commission errors introduced by dry, sandy surfaces. In both Zambia and Mexico, it was also noted that sunglint in certain months caused erroneous cloud masking

for certain processors and hence contributed to lowering the PA for the optical data model. In Mexico, the UA for SAR is, however, only marginally lower than for the optical data model, which is impacted by bottom reflectance from shallow waters and turbidity, which both impact the optical properties (cf. spectral signal) of water more than the physical state and, therefore, the sensitivity of SAR backscatter (e.g., roughness). The Zambia site is dominated by the Kafue flats, an extensive wetland ecosystem subject to variable flooding and with a sharp contrast to the surrounding drier landscape, where fire is a major natural factor impacting the landscape. The dynamic nature and many confounding factors (e.g., fires and emergent vegetation) make Zambia a particularly challenging site, and it was also where the dual sensor approach displayed its strongest potential in balancing the individual strengths and weaknesses of optical and SAR data. In Greenland, the topography and light conditions are the main challenges. For optical data, it means higher commission errors (cf. lower UA) due to shading effects and low sun angles. The SAR model is better at dealing with these issues because it works independent of sunlight, and by using ascending and descending SAR scenes, the part of the landscape that can be monitored is increased. Still, the influence of low-backscatter areas (e.g., exposed riverbeds and in snow dominated landscapes) means the SAR data model typically suffers from commission errors and lower PA.

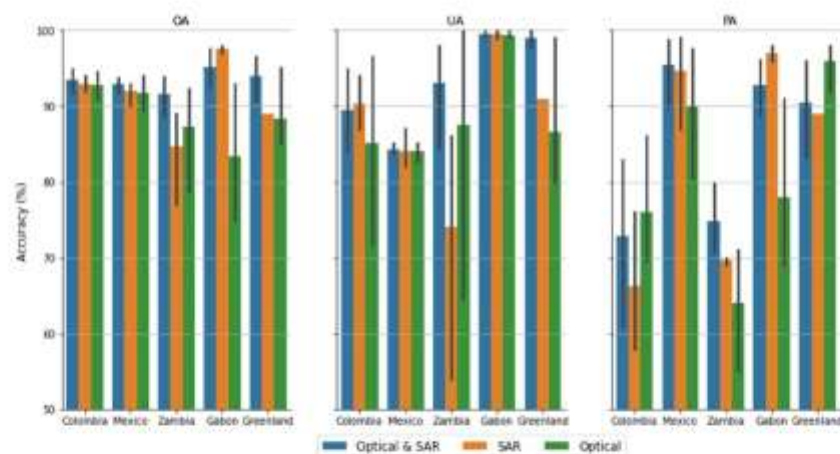


Figure 3. Accuracy statistics from the WorldWater round robin test sites, individually and overall, summarized by model input data type (OA = Overall Accuracy; UA = User Accuracy; PA = Producer Accuracy).

It is important to note that, apart from site-specific characteristics, the UA and PA are also dictated by how individual algorithms have been implemented, e.g., to what extent the individual round robin contributions have favored the importance of commission errors relative to omissions errors. The results will also depend on whether individual scenes are classified and then aggregated to a monthly water map or whether the individual scenes are merged into a monthly composite before water classification. The full accuracy statistics for the individual models is provided as supplementary material (cf. Table S1).

3.3. Object Extraction Accuracy

The 3-m PlanetScope water classification maps used to evaluate object extraction accuracy are shown in Figure 4. Like the full-size test sites, it is important to notice the variability between the sites. Individually, the PlanetScope data represent SWB regions relative to the general water characteristics within their respective test sites, yet, there is

variability between sites with, e.g., Zambia having larger waterbodies on average than Colombia.

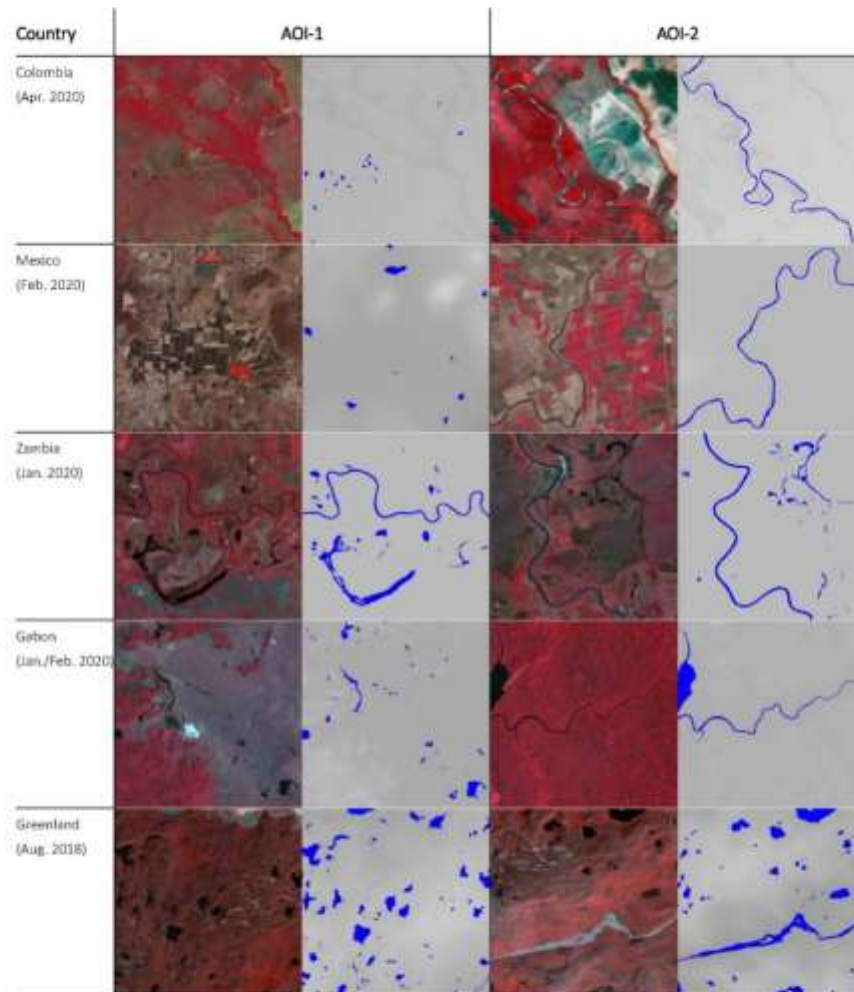


Figure 4. False colour PlanetScope QuickLooks and associated water classifications for each AOI used in the object-based validation approach (Imagery © 2022 Planet Labs Inc.).

Table S2 provides an overview of the summary statistics for object extraction accuracy for each of the three mandatory sites, as well as the optional sites, where relevant. There is a large variability between the individual contributions, and yet, with similar tendency across the sites i.e., the algorithms that integrate optical data perform better than those relying solely on SAR (Figure 5). The lowest overall accuracy is in Colombia, and this is also where the difference between the best optical approaches and the best SAR algorithm is greatest (cf. Figure 5). Figure 5 also shows the highest object extraction accuracy is in Zambia, which, together with Greenland, has the largest share of waterbodies within the test sites (cf. Figure 4). It is also noteworthy the optical data model consistently outperforms the SAR data model in all test sites except for Gabon (Figure 5).

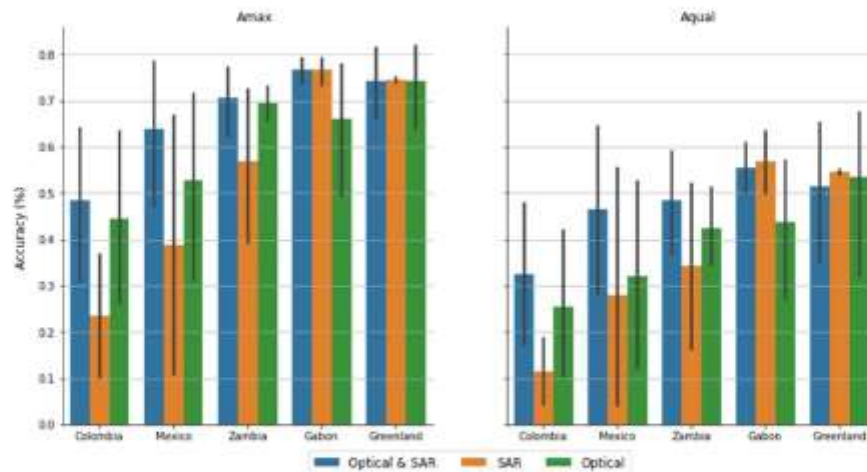


Figure 5. Object accuracy statistics from the WorldWater round robin PlanetScope sites, summarized by country and model input data type.

The findings from the object extraction accuracy analysis indicate that using or integrating optical data into the water detection algorithm is key to achieving accurate water object definitions. How important depends on the average size of the waterbodies and the surrounding landscape. In Colombia, where the average waterbody size/width is smaller compared to other sites, the difference between the optical algorithms and the SAR-only approaches are the largest. This is explained by the characteristics of the input data, with key spectral water detection bands from Sentinel-2 available only in 10-m spatial resolution, while the true spatial resolution of Sentinel-1 is understood to be closer to 20×20 m, although data from the widely used Sentinel-1 Ground Range Detected (GRD) product are delivered with a pixel spacing of 10×10 m. There are also some marked differences between the optical algorithms and the SAR only approaches in Mexico, which is likely caused by the dry environment and a landscape dominated by large tracts of dry, sandy surfaces, as well as the associated challenge for SAR-based water detection [44]. In contrast, the difference between optical and SAR is much less pronounced in Zambia and Gabon, which is likely related to the larger average size of the waterbodies (Zambia) and the dense tropical forest landscape causing a stark land–water contrast (Gabon).

3.4. Temporal Consistency Evaluation

The surface water area (km^2) was calculated over each test site and for each month in the 2-year observation period (cf. July 2018 to June 2020). For each test site, the surface water areas were summarized by input data model type, i.e., optical (O), SAR (S), and the fused date model (O + S). In Figure 6, the average surface water area was then plotted against time with indications of variance (i.e., minimum and maximum observed water extent within a given month) and with some key explanatory variables plotted on the secondary axis.

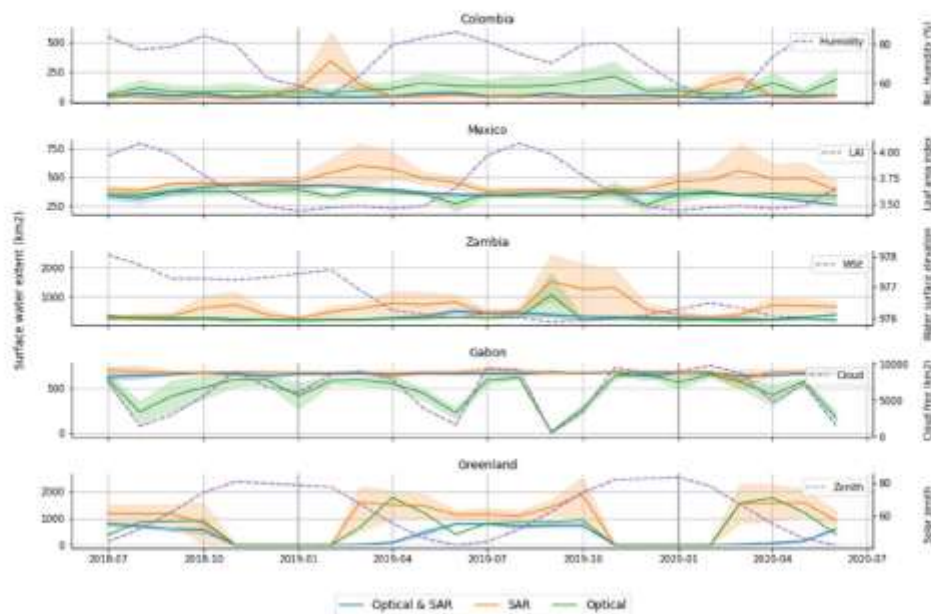


Figure 6. Monthly surface water area trajectories for the individual test sites and per sensor model. For each test site, corresponding timeseries of the key explanatory variables are equally shown, i.e., the Humidity and Leaf Area Index from the ERA-5 monthly averaged reanalysis data [45], water surface elevation from satellite altimetry [46], solar zenith angle, and cloud cover [47].

A comparison of the surface water area temporal development curves shows the variance of the fused Optical–SAR-based algorithms are much less than the single sensor solutions both within and between nearby months. If not directly, then at least indirectly, this indicates the fusing algorithms to be more reliable and have less sensitivity to temporary or seasonal phenomena that can impact water detection, including dry/moist conditions, topographic/canopy shading, and clouds.

In Colombia, the pure optical methods, in general, returns a higher surface water area across the entire timeseries. This can be attributed to false positives from topographic shading and ineffective cloud shadow masking, particularly during the humid season. In Mexico, where clouds and topography are less of a problem, there are hardly any noteworthy peaks/dips in the optical development curves. In Colombia and Mexico, the SAR peaks correspond to the dry seasons when the vegetation cover is low, resulting in an increased influence of low backscatter from dry, sandy surfaces.

In Zambia, the variance observed in both the optical and SAR data predictions is most dominant during the 2019 dry season, which was reported as having been one of the worst droughts in Western Zambia in almost 70 years. The exceptionally low water levels during this period indicates that droughts and receding water lines are likely to have an impact on water classifications. The SAR data are challenged by very dry soils, especially in the southern parts of this site, while wildfires represent another challenge for both the optical and SAR data model, as the burn scars can be difficult to separate from water. In optical imagery, burn scars have low reflectance in the near infrared and visible spectrum, and this can lead to spectral confusion with water. As fire also changes the physical and structural characteristics of the vegetated landscape, it also impacts SAR imagery. After a fire, the backscatter decreases strongly [48], and, as a result, the contrast between land and water will be lessened.

In Gabon, the cloud cover percentage over the test site is, on average, 50%, significantly impacting the optical data model, which returns estimates of water extent that strongly correlate with the cloud cover percentage. In contrast, the SAR and fused sensor approach return a much more consistent timeseries, with no apparent sensitivity to the cloud cover percentage.

Finally, in the case of Greenland, the temporal evaluation shows how limited light conditions in spring and fall (before everything freezes) hamper the optical data model. In essence, our evaluation shows the time window to collect optical imagery is short, but also that it can be extended by integration with SAR data. Using a fused data model in Greenland can also help to even out issues generated by a complex topography (e.g., cast shadows in optical imagery and foreshortening and layover effects in the SAR imagery), as indicated in Table S1.

In Figure 6, a large part of the temporal variation is explained by the performance of the individual contributions both between and within the three different sensor models. The dual sensor model has the least variation and, hence, we argue that it is the more robust in dealing with confounding factors. Figure 7 shows the average monthly surface water area statistics for the top three-performing dual sensor models. Figure 7 illustrates quite well the ability of the dual sensor model to provide consistent timeseries information that captures the seasonality of surface water dynamics in each of the test sites. The strongest seasonality is observed in Colombia and Zambia, which are the two test sites with the largest rainfall gradient. In contrast, Mexico and Gabon have less seasonal variation due to very dry (Mexico) and consistently wet (Gabon) conditions. In Greenland, the seasonality is first and foremost dictated by the temperature, i.e., thawing, and increased meltwater starting around April/May and then frost and total freezing once we enter November.

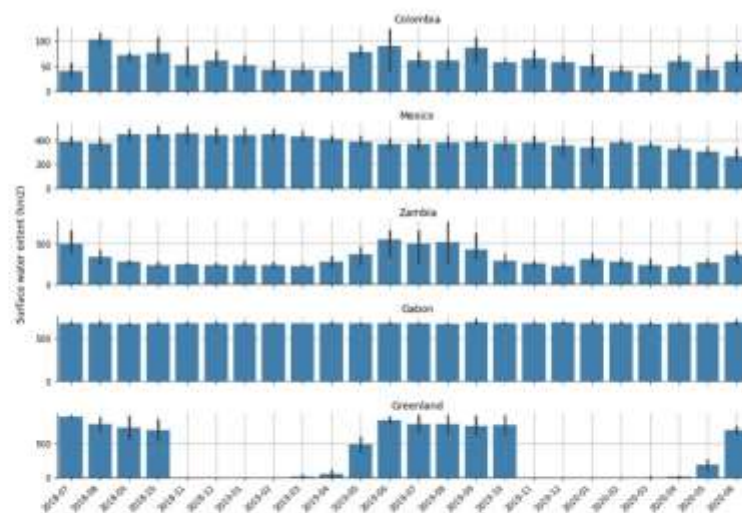


Figure 7. Interannual monthly mean surface water area dynamics and uncertainties (98% CI), as captured by the best-performing dual sensor models (i.e., models A, N, and O).

4. Discussion

The round robin evaluation was conducted over a diverse set of test sites that represented landscapes influenced by several of the known challenges for satellite-based surface water mapping, including topography, clouds, dense and inundated vegetation, fire scars, low-backscatter landcovers, low sun angles, as well as snow and ice. The intercomparison of the different round robin contributions across this diverse set of test sites supports the general hypothesis that fusing optical with SAR data produces a more robust mapping of

surface water extent dynamics across bioclimatic gradients. Yet, the findings also show that, at individual locations, the single sensor approach can outperform the fused sensor approach. By example SAR data are the better option in heavily clouded regions (cf. Gabon) while optical data are better in dry regions and in capturing smaller waterbodies. As such the round robin provides key insight to the advantage of the strengths of optical and SAR data while also identifying how a fused sensor model can help address their individual shortfalls. Moreover, the evaluation demonstrates that both supervised and unsupervised learning can provide very good results, and while steps for preprocessing and postprocessing are highly relevant to the outcome, they include many variables that are harder to quantify in terms of their individual contributions to the statistical accuracy. Still, there are several crosscutting factors that impact optical and SAR data in various ways, and which underpin why the dual sensor approach, on an overall level, outperforms single sensor approaches.

Both SAR and optical data can struggle in mountainous areas, as steep slopes can lead to shadow issues and image distortions. Orthorectification and radiometric terrain correction using a DEM are the main direct techniques to obtain the relevant geometric and/or signal correction. Yet, such correction can introduce errors, as globally available DEMs have known quality issues [49], although newer DEMs provide gradual improvements [50]. In complex terrain, shadows cast by mountains and hills will appear very dark in optical imagery, which can cause a confusion between topographic shadows and water. This means extra steps should be taken when mapping water extents to make sure the effect of terrain shadow is minimized. While there are specific methods to deal with this in optical imagery [51], SAR imagery can also be used, e.g., to remove water classified in optical imagery if it is consistently mapped as land in SAR (cf. Model A). SAR imagery is not affected by natural sunlight shadows cast by topography. However, radar sensors are side-looking, meaning they view the Earth's surface from the side of the satellite as it passes by (as opposed to looking directly from above). The side-looking nature of these radar sensors means that they can only see the side of mountains that face their sensors—they cannot see the opposite side of mountains. This is known as radar topographic shadow. Fortunately, radar sensors, such as Sentinel-1, have both ascending and descending orbits, which can collect imagery from east- and west-looking angles. Using ascending and descending imagery together helps to increase the area that can be effectively monitored using radar imagery; however, this does not solve all radar problems related to topography. Areas in deeper canyons and fjords that have a north–south orientation will likely always be in the radar signal shadow, leading to some unavoidable data gaps, and in these cases, sometimes the optical data model can help.

As both SAR and optical data can struggle in mountainous areas, using one sensor to help overcome the other is not always sufficient. Therefore, DEMs are often applied during postprocessing to mask out regions where water formation is unlikely given the topographic conditions, e.g., due to slopes or based on hydrological terrain analysis, such as the Height Above Nearest Drainage (HAND). A range of DEMs have been used for postprocessing, including the Shuttle Radar Topography Mission (SRTM) DEM (e.g., Model B, M), ALOS World 3D-30 m (Model F, J), and Copernicus DEM (Model E, N, O). Although the impact on accuracy is not quantified directly, the use of Copernicus DEM is recommended, not only because Copernicus DEM comes out favorably in statistical evaluations against other DEMs [50], but also because of the reference year (2010–2015), which is newer than SRTM (i.e., 2000) and AW3D30 (2006–2011). In essence, this means the Copernicus DEM is more likely to capture and, hence, avoids masking out newly established reservoirs, which have boomed dramatically in the past few decades [52].

Cloud cover is a major limiting factor affecting the usefulness of optical imagery. However, if clouds and their associated shadows can be effectively masked out from each image, the remaining cloud-free data in each image can be used for accurate water classification, yet the frequency of monitoring will depend on the persistency of the cloud cover. While several algorithms are available for automated cloud masking, (e.g., MAJA,

Fmask, CFMask, Tmask, IdePix, Sen2Cor, s2cloudless) none are perfect in separating clear observations from those contaminated with clouds and cloud shadows. Too aggressive cloud masking, and many waterbodies may be missed, while failure to adequately mask cloud shadows will introduce many false positives. Often, making a cloud-free optical image will require some form of image compositing and mosaicking. There are several possible ways to do this, e.g., by using the best available pixel by cloudiness (Model O), or through per-pixel band statistics such as mean/median band reflectances (Model N). Model F applies an NDVI Maximum Value Composite (MVC) procedure, which is effective for providing spatially continuous cloud-free imagery [53]. The MVC has been particularly widely adopted in vegetation studies [54], but, since the MVC emphasizes the vegetation signal, it should be used with care for monitoring water dynamics, as seasonally flooded vegetation may risk being masked. Furthermore, and as illustrated by one contribution, a synthetic timeseries can also be constructed by interpolation and gap-filling using the historical water frequency (cf. Model H). Finally, SAR data can also be used to fill in the “cloud” gaps in the optical imagery. However, even if SAR imagery is not affected by clouds, it is impacted by other issues, which can result in spurious water detection, including speckle noise and permanent low-backscatter regions. The reduction of speckle noise is important to improve the usefulness of SAR imagery. The main purpose of the noise reduction technique is to remove speckle noise while still retaining the important features in the images. Widely adopted speckle filters, such as Lee Sigma or Refined Lee, have proved effective; however, depending on the window kernel size, they may compromise the ability to map smaller water features. Therefore, attention has been drawn to other methods, such as the Gamma Map method (Model A, E) and the use of temporal filtering (e.g., mean, median, or minimum backscatter), as a means to better preserve spatial resolution (cf. Model O, N). The further advantage of using temporal filtering is the ability to also suppress the influence of high winds, which can cause wind-roughened waters that, at specific times, can vanish the contrast between open water and dry surfaces and cause Bragg scattering. With SAR data, it can also be difficult to differentiate water from other surfaces with low backscatter, such as asphalt (parking lots, airports, roads), flat rock, and, in some dry regions, sand surfaces. Long timeseries of backscatter measurements can be used to identify such areas but at the expense of computational efficiency, especially for large areas [55]. Another way is to integrate optical data to reduce potential commission errors caused by permanent low-backscatter areas (cf. Models A and O).

As additional examples, the round robin intercomparisons have also shown how the complementary use of optical and SAR data can help suppress the influence of burn scars and, to an extent, the monitoring period in light-constrained, high-latitude regions.

Aside from the challenges discussed above, there are variables and challenges which could not be fully evaluated. Unresolved issues still circulate around inundated vegetation and how to deal with the cryosphere. As the focus in this study was on open inland waters, neither of these issues was investigated. However, future improvements could be performed through the investigation of L-band SAR sensors, which penetrate vegetation better than C-band SAR data (Sentinel-1) and have potential for mapping flooded areas under vegetated canopies [56,57]. In large parts of the world, lake and river ice is an integral part of annual water dynamics, which is why we also recommend looking at scalable solutions for using optical and SAR data to monitor lake and river ice evolution [58,59] and as complementary information for open surface water dynamics.

Urban environments represent another challenge from the perspective of both optical and SAR data. For optical images, the main issue is building shadows, whereas SAR data may suffer from layover effects caused by tall buildings, as well as corner reflection (cf. double/triple bouncing). Like topography, the urban challenge is often addressed using postprocess masking, which is sensible, especially for large-area applications, as urban areas represent only a fraction of the overall landscape, and the waterbodies associated with the urban environment even less so. In addition, and as new high resolution and freely available urban footprint layers become available [60], urban masking will gradually

improve and integrating them as masking layers can help simplify the water mapping solution.

The results and above discussion point to some inherent limitations to mapping surface water when relying solely on either optical or SAR-based instruments. These limitations can be partly mitigated by using both sensors in a fused approach for surface water extent mapping. However, since the fused mapping approach will likely add to complexity, computational effort, transferability, and automation level of the mapping approach, it is important to consider exact needs and objectives before the appropriation of a specific data model.

However, if monitoring is to be conducted in a region with persistent cloud cover, or if the focus is to monitor during the wet–cloudy season, it may be worth considering if adding optical data will bring the necessary improvement to warrant the additional complexity of an operational solution. In other regions, the status of small farm dams may be the most critical information gap in supporting timely information on potential water shortages. In drier regions or during dry spells, where clouds are not an issue, monitoring should rely on optical data only to maximize the spatial resolution. However, where clouds may be an issue, the integration of SAR data will be critical to reliably monitor the status of small farm reservoirs and dams [22,61]. This reiterates that the best practices for surface water monitoring are often reliant on the study domain. In other words, a case-dependent choice of mapping approach will be needed based on certain criteria, such as ecosystem type, seasonality, climate regime, area size, and requirements for the degree of automation. Moreover, as EO technology becomes more widely adopted and mapping approaches evolve, it is further recommended that cross-comparison exercises, as presented in this paper, be repeated periodically to assess advances in surface water mapping.

5. Conclusions

The availability of satellite missions and constellations for environmental monitoring has continued to grow in the past decades, and combined with the advances in technical infrastructures for big data analysis, it is now within the realm of possibility for countries to implement satellite-based surface water monitoring systems. These systems will be vital to supporting more evidence-based planning and management of water resources and provide an ability to efficiently report and act in response to the global water agenda. By evaluating 14 different EO-based models for surface water detection, we show that single sensor approaches can produce accurate and consistent water maps under ideal conditions, and yet, across a range of challenging environments, the synergistic usage of optical and SAR data delivers more accurate and consistent outputs.

The findings in this paper therefore bear some important perspectives for formulating a new best practice where optical and SAR data are used synergistically to achieve the highest accuracy and most consistent results for monitoring surface water dynamics. While accuracy is a critical concern for selecting a surface water detection model, there are other important aspects, including computational efficiency, simplicity, and ease of implementation, which all contribute to increase understanding, maintainability, and potential scalability. In the end, specific working routines, management objectives, and individual user preferences may all contribute to how users will choose to appropriate EO technology for surface water monitoring. At larger scales across diverse ecological gradients, a synergistic approach should be preferred, but at a local scale, SAR data may be preferred for the effective and timely monitoring of water extent and potential emerging floods during cloudy periods, and similar optical data may be preferred to monitor the status of reservoirs and small waterbodies during drought periods and when clouds are not an issue.

Therefore, rather than advocating for a single “best” approach, we recommend flexibility and options to build and/or adapt surface water detection methods that meet individual user needs in terms of management goals, environmental settings, and scale of study, i.e.,

ensuring users have options for receiving data in multiple formats or from multiple sources, and with the tools necessary to process these data effectively.

The round robin evaluation presented in this paper has shown that EO datasets, methods, and tools for monitoring surface water dynamics are available and successfully applied in various contexts around the globe. The upcoming challenge will be to make the community aware of these tools and, via practical guidance, illustrate how to get started using EO data and tools to support better water resource monitoring, reporting, and management.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs14102410/s1>, Table S1. Classification accuracies (%) for each of the five round robin test sites. OA = Overall Accuracy; UA = User's Accuracy, and PA = Producers Accuracy. Table S2. Summary of object extraction accuracies. The accuracy metrics are maximum overlap area (A_{max}) and quality (A_{qual}) as a joint balanced measure of correctness (A_{cor}) and completeness (A_{com}). The overall score is the product between A_{max} and A_{qual} .

Author Contributions: Conceptualization, C.T., M.P., M.C.; methodology, C.T., T.S.; validation, K.J., R.P.M., C.T.; formal analysis and investigation, D.D., M.R., B.D., J.F., M.S., S.W., S.L., H.Z., G.S., L.G., A.M., C.S., V.V., A.H., K.M., M.C.R.C., J.-M.M., D.M., R.R., M.T., H.A., J.H. (Jens Hiestermann), J.H. (Jason Hallowes); data curation, P.R., R.P.M.; writing—original draft preparation, C.T., D.D.; writing—review and editing, J.F., M.S., S.W., S.L., H.Z., G.S., L.G., A.M., C.S., V.V., A.H., K.M., M.C.R.C., J.-M.M., D.M., R.R., M.T., H.A., J.H. (Jens Hiestermann), J.H. (Jason Hallowes), M.C., M.P.; visualization, D.D.; supervision, C.T.; funding acquisition, C.T., M.P. All authors have read and agreed to the published version of the manuscript.

Funding: This study was executed in the context of the WorldWater project, funded by European Space Agency (ESA) under the EO Science for Society programmatic element of the 5th Earth Observation Envelope Programme (EOEP-5, 2017–2021). S. Liu and H. Zhou were funded by the National Key Research and Development Program of China (Grant No. 2018YFE0106500).

Data Availability Statement: The validation data samples used in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.6539508>.

Acknowledgments: The WorldWater round robin was organized under the funding and auspices of the European Space Agency and supported by a number of international organizations and initiatives, including CNES, NASA, the European Association of Remote Sensing Companies (EARSC), and the CEOS Ad hoc team on Sustainable Development Goals (CEOS SDG AHT), as well as GEO and their Earth Observations for the Sustainable Development Goals (EO4SDG) initiative. The European Commission and ESA provided the Sentinel data. The USGS and NASA provided the Landsat imagery. The PlanetScope data was provided by Planet Labs Inc. under ESA's Third-Party Missions scheme.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UN. United Nations Sustainable Development Goals: Goal 6: Ensure Access to Water and Sanitation for All. 2020. Available online: <https://www.un.org/sustainabledevelopment/water-and-sanitation/> (accessed on 4 April 2022).
2. Long, J. The United Nations' 2030 Agenda for Sustainable Development and the Impact of the Accounting Industry. *Honor. Coll. Theses* **2019**, 260. Available online: https://digitalcommons.pace.edu/honorscollege_theses/260 (accessed on 4 April 2022).
3. General Assembly of the United Nations. *International Decade for Action: Water for Sustainable Development: 2018–2028*; UN doc A/RES/71/222 (7 February 2017); United Nations: New York, NY, USA, 2017.
4. Pekel, J.F.; Cottam, A.; Gorelick, N.; Belward, A.S. High-resolution mapping of global surface water and its long-term changes. *Nature* **2016**, *540*, 418–422. [CrossRef] [PubMed]
5. Pickens, A.H.; Hansen, M.C.; Hancher, M.; Stehman, S.V.; Tyukavina, A.; Potapov, P.; Marroquin, B.; Sherani, Z. Mapping and sampling to characterize global inland water dynamics from 1999 to 2018 with full Landsat time-series. *Remote Sens. Environ.* **2020**, *243*, 111792. [CrossRef]
6. Huang, C.; Chen, Y.; Zhang, S.; Wu, J. Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review. *Rev. Geophys.* **2018**, *56*, 333–360. [CrossRef]
7. Brisco, B. Mapping and monitoring surface water and wetlands with synthetic aperture radar. In *Remote Sensing of Wetlands: Applications and Advances*; Tiner, R.W., Lang, M.W., Klemas, V.V., Eds.; CRC Press: Boca Raton, FL, USA, 2015; pp. 119–136.

8. Druce, D.; Xiao, T.; Lei, X.; Guo, T.; Kittel, C.M.M.; Grogan, K.; Tottrup, C. An optical and SAR based fusion approach for mapping surface water dynamics over mainland China. *Remote Sens.* **2021**, *13*, 1663. [CrossRef]
9. Bioresita, E.; Puissant, A.; Stumpf, A.; Malet, J.-P.P. Fusion of Sentinel-1 and Sentinel-2 image time series for permanent and temporary surface water mapping. *Int. J. Remote Sens.* **2019**, *40*, 9026–9049. [CrossRef]
10. Markert, K.N.; Chishtie, F.; Anderson, E.R.; Saah, D.; Griffin, R.E. On the merging of optical and SAR satellite imagery for surface water mapping applications. *Results Phys.* **2018**, *9*, 275–277. [CrossRef]
11. van Leeuwen, B.; Tobak, Z.; Kovács, F. Sentinel-1 and-2 based near real time inland excess water mapping for optimized water management. *Sustainability* **2020**, *12*, 2854. [CrossRef]
12. Showstack, R. NEWS Sentinel Satellites Initiate New Era in Earth Observation. *EOS* **2014**, *95*, 239–240. [CrossRef]
13. UNFCCC. Paris Agreement. 2015. Available online: http://unfccc.int/files/essential_background/convention/application/pdf/english_paris_agreement.pdf (accessed on 4 April 2022).
14. UNDRR. Sendai Framework for Disaster Risk Reduction. 2015. Available online: <https://www.undrr.org/implementing-sendai-framework/what-sendai-framework> (accessed on 4 April 2022).
15. Airbus. Copernicus DEM Product Handbook. 2019. Available online: https://spacedata.copernicus.eu/documents/20126/0/GEO1988-CopernicusDEM-SPE-002_ProductHandbook_11.00+%281%29.pdf/40b2739a-38d3-2b9f-fe35-1184ccd17694?t=1612269439996 (accessed on 2 March 2021).
16. Donchyts, G.; Schellekens, J.; Winsemius, H.; Eisemann, E.; Van de Giesen, N. A 30 m Resolution Surface Water Mask Including Estimation of Positional and Thematic Differences Using Landsat 8, SRTM and OpenStreetMap: A Case Study in the Murray-Darling Basin, Australia. *Remote Sens.* **2016**, *8*, 386. [CrossRef]
17. Markert, K.N.; Markert, A.M.; Mayer, T.; Nauman, C.; Haag, A.; Poortinga, A.; Bhandari, B.; Thwal, N.S.; Kunlami, T.; Chishtie, F.; et al. Comparing Sentinel-1 surface water mapping algorithms and radiometric terrain correction processing in southeast Asia utilizing Google Earth Engine. *Remote Sens.* **2020**, *12*, 2469. [CrossRef]
18. Chini, M.; Hostache, R.; Giustarini, L.; Matgen, P. A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case. *IEEE Trans. Geosci. Remote Sens.* **2017**, *55*, 6975–6988. [CrossRef]
19. Thompson, M.; Hiestermann, J.; Eady, B.; Hollowes, J. Frankly My Dear I Give a Dam! Or Using Satellite Observation to Determine Water Resource Availability in Catchments. 2018. Available online: http://sbdvc.ekodata.co.za/downloads/SANCLAHS_paper.pdf (accessed on 6 May 2022).
20. Department of Science and Innovation Republic of South Africa. mzansiAmanzi—A Monthly Outlook of Water in South Africa. Available online: <https://www.water-southafrica.co.za/> (accessed on 6 May 2022).
21. Wangchuk, S.; Bolch, T. Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier: Strengths and challenges. *Sci. Remote Sens.* **2020**, *2*, 100008. [CrossRef]
22. Vanthof, V.; Kelly, R. Water storage estimation in ungauged small reservoirs with the TanDEM-X DEM and multi-source satellite observations. *Remote Sens. Environ.* **2019**, *235*, 111437. [CrossRef]
23. Yamazaki, D.; Ikeshima, D.; Neal, J.C.; O’Loughlin, F.; Sampson, C.C.; Kanae, S.; Bates, P.D. MERIT DEM: A new high-accuracy global digital elevation model and its merit to global hydrodynamic modeling. *AGU Fall Meet. Abstr.* **2017**, *2017*, H12C-04.
24. Nobre, A.D.; Cuartas, L.A.; Hodnett, M.; Rennó, C.D.; Rodrigues, G.; Silveira, A.; Saleska, S. Height Above the Nearest Drainage—a hydrologically relevant new terrain model. *J. Hydrol.* **2011**, *404*, 13–29. [CrossRef]
25. Gallant, J.C.; Dowling, T.I. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* **2003**, *39*, 12. [CrossRef]
26. Fan, X.; Liu, Y.; Wu, G.; Zhao, X. Compositing the Minimum NDVI for Daily Water Surface Mapping. *Remote Sens.* **2020**, *12*, 700. [CrossRef]
27. Guzder-Williams, B.; Alemohammad, H. Surface Water Detection from Sentinel-1. In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021.
28. Zhou, H.; Liu, S.; Hu, S.; Mo, X. Retrieving dynamics of the surface water extent in the upper reach of Yellow River. *Sci. Total Environ.* **2021**, *800*, 149348. [CrossRef]
29. Cordeiro, M.C.R.; Martinez, J.-M.; Peña-Luque, S. Automatic water detection from multidimensional hierarchical clustering for Sentinel-2 images and a comparison with Level 2A processors. *Remote Sens. Environ.* **2021**, *253*, 112209. [CrossRef]
30. Defourny, P.; Kirches, G.; Brockmann, C.; Boettcher, M.; Peters, M.; Bontemps, S.; Lamarche, C.; Schlerf, M.; Santoro, M. Land cover CCI: Product User Guide Version 2.0. 2017. Available online: http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf (accessed on 4 April 2022).
31. Marzi, D.; Gamba, P. Inland Water Body Mapping Using Multitemporal Sentinel-1 SAR Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *14*, 11789–11799. [CrossRef]
32. Schumann, G.J.P.; Campanella, P.; Tasso, A.; Giustarini, L.; Matgen, P.; Chini, M.; Hoffmann, L. An Online Platform for Fully-Automated EO Processing Workflows for Developers and End-Users Alike. In Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, Brussels, Belgium, 11–16 July 2021; pp. 8656–8659.
33. Merciol, F.; Fauqueur, L.; Damodaran, B.B.; Rémy, P.-Y.; Desclée, B.; Dazin, F.; Lefèvre, S.; Masse, A.; Sannier, C. Geobia at the terapixel scale: Toward efficient mapping of small woody features from heterogeneous vhr scenes. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 46. [CrossRef]

34. Ludwig, C.; Walli, A.; Schleicher, C.; Weichselbaum, J.; Riffler, M. A highly automated algorithm for wetland detection using multi-temporal optical satellite data. *Remote Sens. Environ.* **2019**, *224*, 333–351. [CrossRef]
35. Martinis, S.; Twele, A.; Voigt, S. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 303–314. [CrossRef]
36. Bradley, D.; Roth, G. Adaptive thresholding using the integral image. *J. Graph. Tools* **2007**, *12*, 13–21. [CrossRef]
37. Adams, R.; Bischof, L. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 641–647. [CrossRef]
38. Martinis, S.; Kersten, J.; Twele, A. A fully automated TerraSAR-X based flood service. *ISPRS J. Photogramm. Remote Sens.* **2015**, *104*, 203–212. [CrossRef]
39. Twele, A.; Cao, W.; Plank, S.; Martinis, S. Sentinel-1-based flood mapping: A fully automated processing chain. *Int. J. Remote Sens.* **2016**, *37*, 2990–3004. [CrossRef]
40. Olofsson, P.; Foody, G.M.; Herold, M.; Stehman, S.V.; Woodcock, C.E.; Wulder, M.A. Good practices for estimating area and assessing accuracy of land change. *Remote Sens. Environ.* **2014**, *148*, 42–57. [CrossRef]
41. Downing, J.A.; Prairie, Y.T.; Cole, J.J.; Duarte, C.M.; Tranvik, L.J.; Striegl, R.G.; McDowell, W.H.; Kortelainen, P.; Caraco, N.F.; Melack, J.M. The global abundance and size distribution of lakes, ponds, and impoundments. *Limnol. Oceanogr.* **2006**, *51*, 2388–2397. [CrossRef]
42. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 3146–3154.
43. Cai, L.; Shi, W.; Miao, Z.; Hao, M. Accuracy assessment measures for object extraction from remote sensing images. *Remote Sens.* **2018**, *10*, 303. [CrossRef]
44. Martinis, S.; Kuenzer, C.; Wendleder, A.; Huth, J.; Twele, A.; Roth, A.; Dech, S. Comparing four operational SAR-based water and flood detection approaches. *Int. J. Remote Sens.* **2015**, *36*, 3519–3543. [CrossRef]
45. Climate Data Store. ERA5 Climate Reanalysis. Available online: <https://cds.climate.copernicus.eu/> (accessed on 27 March 2022).
46. U.S. Department of Agriculture. Global Reservoirs and Lakes Monitor (G-REALM). Available online: https://ipad.fas.usda.gov/cropexplorer/global_reservoir/ (accessed on 22 March 2022).
47. European Union/ESA/Copernicus/SentinelHub. Sentinel-2: Cloud Probability. Available online: https://developers.google.com/earth-engine/datasets/catalog/COPERNICUS_S2_CLOUD_PROBABILITY#description (accessed on 22 March 2022).
48. Siegert, F.; Ruecker, G. Use of multitemporal ERS-2 SAR images for identification of burned scars in south-east Asian tropical rainforest. *Int. J. Remote Sens.* **2000**, *21*, 831–837. [CrossRef]
49. Uuemaa, E.; Ahi, S.; Montibeller, B.; Muru, M.; Knoch, A. Vertical Accuracy of Freely Available Global Digital Elevation Models (ASTER, AW3D30, MERIT, TanDEM-X, SRTM, and NASADEM). *Remote Sens.* **2020**, *12*, 3482. [CrossRef]
50. Guth, P.L.; Geoffroy, T.M. LIDAR point cloud and ICESat-2 evaluation of 1 second global digital elevation models: Copernicus wins. *Trans. GIS* **2021**, *25*, 2245–2261. [CrossRef]
51. Tottrup, C. Forest and land cover mapping in a tropical highland region. *Photogramm. Eng. Remote Sens.* **2007**, *73*, 1057.
52. Zarfl, C.; Lumsdon, A.E.; Berlekamp, J.; Tydecks, L.; Tockner, K. A global boom in hydropower dam construction. *Aquat. Sci.* **2015**, *77*, 161–170. [CrossRef]
53. Holben, B.N. Characteristics of maximum-value composite images from temporal AVHRR data. *Int. J. Remote Sens.* **1986**, *7*, 1417–1434. [CrossRef]
54. Pettorelli, N.; Vik, J.O.; Mysterud, A.; Gaillard, J.-M.; Tucker, C.J.; Stenseth, N.C. Using the satellite-derived NDVI to assess ecological responses to environmental change. *Trends Ecol. Evol.* **2005**, *20*, 503–510. [CrossRef]
55. Schlaffer, S.; Matgen, P.; Hollaus, M.; Wagner, W. Flood detection from multi-temporal SAR data using harmonic analysis and change detection. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 15–24. [CrossRef]
56. Tsyganskaya, V.; Martinis, S.; Marzahn, P.; Ludwig, R. SAR-based detection of flooded vegetation—a review of characteristics and approaches. *Int. J. Remote Sens.* **2018**, *39*, 2255–2293. [CrossRef]
57. Tsyganskaya, V.; Martinis, S.; Marzahn, P.; Ludwig, R. Detection of temporary flooded vegetation using Sentinel-1 time series data. *Remote Sens.* **2018**, *10*, 1286. [CrossRef]
58. Wang, X.; Feng, L.; Gibson, L.; Qi, W.; Liu, J.; Zheng, Y.; Tang, J.; Zeng, Z.; Zheng, C. High-Resolution Mapping of Ice Cover Changes in Over 33,000 Lakes Across the North Temperate Zone. *Geophys. Res. Lett.* **2021**, *48*, e2021GL095614. [CrossRef]
59. Scott, K.A.; Xu, L.; Pour, H.K. Retrieval of ice/water observations from synthetic aperture radar imagery for use in lake ice data assimilation. *J. Great Lakes Res.* **2020**, *46*, 1521–1532. [CrossRef]
60. Esch, T.; Marconcini, M.; Felber, A.; Roth, A.; Heldens, W.; Huber, M.; Schwinger, M.; Taubenböck, H.; Müller, A.; Dech, S. Urban footprint processor—Fully automated processing chain generating settlement masks from global data of the TanDEM-X mission. *IEEE Geosci. Remote Sens. Lett.* **2013**, *10*, 1617–1621. [CrossRef]
61. Perin, V.; Tulbure, M.G.; Gaines, M.D.; Reba, M.L.; Yaeger, M.A. A multi-sensor satellite imagery approach to monitor on-farm reservoirs. *Remote Sens. Environ.* **2022**, *270*, 112796. [CrossRef]

- III. Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, Planetscope and Sharpened imagery

Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, PlanetScope and Sharpened imagery

Thiago C. Tayer^{a,b*}, Michael M. Douglas^{a,b}, Mauricio C. R. Cordeiro^{c,d}, André D.N. Tayer^e, J. Nik Callow^{a,b}, Leah Beesley^{a,b}, Don McFarlane^a

^aSchool of Agriculture and Environment, The University of Western Australia, Stirling Highway, Crawley, WA 6009, Australia; ^bNorthern Australia Environmental Resources Hub, National Environmental Science Program, Brisbane, Qld, Australia; ^cAgência Nacional de Águas (ANA), Setor Policial Sul, Área 5, Quadra 3, Brasília, DF 70610-200, Brasil; ^dGéosciences Environnement Toulouse (GET), Unité Mixte de Recherche 5563, IRD/CNRS/Université, Toulouse 31400, France; ^eMaua Institute of Technology, Praça Mauá, São Caetano do Sul, SP 09580-900, Brazil.

*Correspondence: tayerthiaggo@gmail.com

Maximizing the accuracy of surface water detection in an intermittent river using the Water Detect algorithm and a sensitivity analysis to compare the performance of Sentinel-2, Planetscope and Sharpened imagery

Remote sensing classification algorithms performed on multispectral images offer much promise for mapping river systems over large areas and through time. However, the quality of the water detection results depends on the algorithm used and its input parameters, as well as the satellite's spectral quality and spatial resolution. Achieving accurate results is especially important for intermittent rivers, given the challenges of mapping their narrow linear features (e.g., thalweg/run/riffle) and isolated pools. The Water Detect algorithm is emerging as one of the best options to identify water using optical satellites without the need for ancillary data, but to date, no one has evaluated how to maximize its accuracy in an intermittent river. There is a need to compare its accuracy across a range of input parameters and satellites, especially those that can provide high spatial resolution and revisit frequency, which may not be equipped with Shortwave infrared (SWIR) sensors, useful for identifying water. The lack of SWIR bands in many high spatial resolution and high revisit frequency satellites, such as Planetscope, which only has Visible and Near-infrared (VNIR) sensors, may be overcome by image sharpening and band synthesizing, which integrates different image sources to improve the information content and quality of multispectral images. To this end, we automated an existing method for sharpening and synthesizing bands and applied it to a series of multispectral Sentinel-2 and Planetscope images. We then developed a sensitivity analysis algorithm to compare the accuracy for all possible combinations of input parameters in a given range for the water detection process – thus allowing the identification of parameters that maximize the ability to detect water. Our sensitivity analysis was performed on five sets of images, two of which were generated by the sharpening and synthesizing bands routine, i.e., Planetscope (VNIR), Sentinel-2 (VNIR and VNIR+SWIR), and Sentinel-2 sharpened with Planetscope (VNIR and VNIR+Synthetic SWIR). We used the Matthews Correlation Coefficient to evaluate the accuracy of the water detection results. We found that the ability to accurately detect surface water using spectral imagery was maximized by identifying the optimal input parameters for the Water Detect algorithm and using VNIR sensors, with relatively little gained by using image sharpening routines. Results showed that by maximizing the overall detection accuracy, we also increased the potential of detecting important features of intermittent rivers. Using automated routines to sharpen imagery and determine input parameters for the Water Detect algorithm is a significant advancement that should increase our ability to accurately detect and map water in intermittent rivers.

Keywords: remote sensing; Sentinel-2; Planetscope; water detection; sensitivity analysis; image sharpening; band synthesizing; intermittent river

Introduction

As freshwater availability and quality increasingly becomes a global concern (Vörösmarty et al. 2010), there is a growing need for more accurate water resource assessments. This is a challenge for the world's intermittent dryland rivers, which support around 40% of the global population (Koochafkan and Stewart 2008) but have little gauging infrastructure/data to guide decision-making (Callow and Boggs 2013; Jarihani et al. 2015). Accurately mapping surface water across space and through time in intermittent river systems can assist us to better understand their hydrology and ecology and improve our capacity to quantify land use and climate change impacts (Lehner et al. 2011; Mueller et al. 2016).

Mapping surface water using remote sensing is a particular challenge in intermittent rivers because water contracts down to narrow linear features (e.g., thalweg/run/riffle) and isolated pools which requires accurate water detection methods as well as reliable image datasets. The increasing spatial and spectral resolution of remote sensing platforms may overcome these barriers and allow better mapping of hydrologic features and connections. Remotely-sensed technologies provide a cost-effective alternative to traditional ground-based monitoring methods and offer advantages such as higher frequency, temporal and spatial continuity, as well as lower costs (Callow and Boggs 2013). Identifying intermittent rivers' water features indicative of no flow periods and mapping persistent pools using remote satellite data can help understand the impacts of hydrological alterations on riverine habitats in intermittent systems, which can improve the management of surface- and groundwater resources and assist the preservation of biodiversity. However, there is a critical gap in our current ability to map these features as accurately as possible due to the compromise inherent with water detection algorithms (e.g., accuracy, processing time, user bias) and imagery datasets (e.g., spectral bands, ground pixel resolution, and revisit frequency – most of the satellites with high- resolution and revisit frequency lack of SWIR bands). Increasing the overall accuracy of the water detection process can potentially improve the identification of those features. Efficient water detection methods and high-resolution multispectral imagery with high revisit frequency, such as Planescope, have the potential to make the detection of intermittent river features more accurate.

Water detection methods based on multispectral imagery offer advantages due to their spatial resolution, coverage extent, and revisit frequency (Marcus and Fonstad 2008; Jiang et al. 2014; Liu, Yao, and Wang 2021; Gao 1996; Qiao et al. 2012; Yang et al. 2017; Callow and Boggs 2013), but their efficacy depends directly on the method used, cloud cover and the spectral quality and resolution of the chosen satellite. From the many methods that use optical sensors to identify water, the Water Detect algorithm (Cordeiro, Martinez, and Peña-Luque 2021) stands out as potentially one of the best options due to its accuracy Tottrup et al. (2022) and open-source code. In the original study, this algorithm showed higher accuracy in mapping water bodies smaller than 0.5 ha, and at a country scale, than other commonly used unsupervised methods (e.g., Canny-edge MNDWI, Multi-Otsu MNDWI, FMask, Sen2Cor, and MAJA) (Cordeiro, Martinez, and Peña-Luque 2021). Similarly, Peña-Luque et al. (2021) confirmed the advantage of using multidimensional clustering over other traditional water detection methods (Canny-edge and Hue, Saturation, Value - HSV) when classifying water at both local (i.e., reservoir) and regional levels. The Water Detect algorithm was also compared with other optical water detection methods and tested in several countries by Tottrup et al. (2022); results indicated that Water Detect had higher average accuracy than other tested methods that used only optical input. However, care needs to be taken when selecting the best input parameters for the Water Detect algorithm for a specific region or sensor. If sub-optimal parameters are chosen, the resulting classification accuracy can be compromised. Sensitivity analysis algorithms designed to test and compare the results of multiple input parameters can be used to identify optimal parameter values, i.e., those that most accurately detect water, for each use case. Currently, no automated script exists that performs a sensitivity analysis for the Water Detect algorithm meaning there is little guidance for practitioners wishing to use this method for water mapping and a lack of understanding of how the dynamic change of input parameters influences the accuracy of this method.

Optical satellites with high spatial resolution and high revisit frequency have the potential to provide quality input data to water detection methods and improve the mapping of intermittent river features and the overall result accuracy. If a river's surface water extent changes markedly through time, which is common in intermittent systems, the spatial

resolution and the frequency of cloud-free imagery constrain our ability to quantify these changes. In this sense, the PlanetScope constellation (Planet Team 2021) is a game-changer, producing daily global coverage at high spatial resolutions (3 m). The high frequency- and spatial resolution of imagery obtained from PlanetScope now means that riverine surface water can be mapped with much higher precision than previously achieved with coarser and less frequent imagery from platforms such as Landsat, Sentinel, or MODIS. However, while PlanetScope accessibility, frequency, and pixel resolution represent a significant advance, their sensors lack longer spectral wavelengths (SWIR bands). This can be problematic for water mapping as these wavelengths are known to be sensitive to water absorption and are widely used in water mapping. For instance, the Water Detect algorithm was initially proposed to be applied to Sentinel-2 images with SWIR bands. PlanetScope sensors also have no onboard calibration devices, making it hard to systematically implement corrections to the multiple generations of this satellite constellation, resulting in spectrally-variable output data (Huang and Roy 2021). While initially problematic, this difference in spectral quality and number of bands among PlanetScope images may be overcome by spectral data sharpening and band synthesizing.

Image sharpening has been developed to get the best of both spectral and spatial resolution by integrating different image sources to improve the information content and quality of multispectral images (Kaplan and Avdan 2018). Generally, high-resolution multispectral imagery is merged with lower-resolution images to create a hybrid product with the highest resolution possible while preserving spectral quality (Fonseca et al. 2011). Some satellites, such as Landsat, Worldview, and Quickbird, provide a panchromatic (Pan) band, which can be used in sharpening. Satellites with no Pan band can use other satellites' high-resolution bands to emulate a Pan band, provided the pixel resolution and the sharpening algorithm are suited to the specific process. For instance, to take advantage of PlanetScope's four spectral resolution bands (3m), Li et al. (2020) tested methods to combine spectrally-corrected Sentinel-2 imagery with high-resolution but uncorrected, PlanetScope-0 imagery for Earth Observation studies. Li et al. (2020) concluded that it is feasible to sharpen 3 m VNIR (Visible – Blue/Green/Red, and Near-Infrared – NIR) and synthesize Red-edge and Shortwave Infrared (SWIR) reflectance on days that Sentinel-2 and PlanetScope-0 are spatially overlapping. Sharpening and synthesizing bands can also be highly beneficial for surface water mapping. For instance, sharpening methods, such as the Landsat-MODIS fusion (Jarihani et al. 2015), could be adapted to PlanetScope-Sentinel. Given that water has a strong absorption on the SWIR band (Xu 2006), synthesizing bands allows the application of many methods for water detection that uses the SWIR band (Feyisa et al. 2014; Xu 2006; Fisher, Flood, and Danaher 2016; Wang et al. 2018; Jarihani et al. 2015), even if the chosen satellite is not equipped with the sensor to capture this wavelength band. Although using high resolution and high-frequency imagery combined with sharpening and band synthesizing techniques have the potential to improve the accuracy of surface water mapping, their specific benefit to the Water Detect algorithm remains unknown. Moreover, as the sharpening and band synthesizing process can require considerable processing time, there is a need to ascertain the benefit of this step for surface water detection.

This study aims to improve our ability to detect surface water in an intermittent river using multispectral imagery. Specifically, we compared the performance of a higher spatial resolution and narrower spectral range (PlanetScope-0) with a lower spatial resolution and wider spectral range image (Sentinel-2) and assessed the benefits of combining both images through sharpening to identify surface water features in an intermittent river over a time series analysis. To achieve this, we (i) automated the sharpening routine proposed by Li et al. (2020); (ii) developed a sensitivity analysis algorithm to test all possible combinations of input

parameters (spectral indices, maximum number of clusters, and regularization) and determine optimum inputs for the Water Detect algorithm that generates the most accurate detection of surface water given a specific image type, (iii) determined an input combination for Water Detect that performed consistently over multiple images and time, representing the best global fitting input parameters for each sensor, and (iv) performed a visual inspection of the results to assess special cases. The authors are unaware of any previous research using Planetscope or sharpened images as input data or any sensitivity analysis algorithm that can automatically test and assess input parameters and results for the Water Detect algorithm. Our sensitivity analysis algorithm and the learnings from this study will assist the development of best practices in mapping surface water in intermittent rivers.

Materials and Methods

Study Area

The Fitzroy River, located in the semi-arid Kimberley region of north-western Australia (Figure 1), experiences extreme flood flows during the wet season (November to April) but dries to poorly interconnected or disconnected pools in the dry season (May to October) (Beesley et al. 2021). Using remote sensing techniques to map these pools and maintenance of their interconnection accurately can provide crucial information on pool persistence and improve our understanding of the role of surface water and groundwater in the river's hydrological functioning.

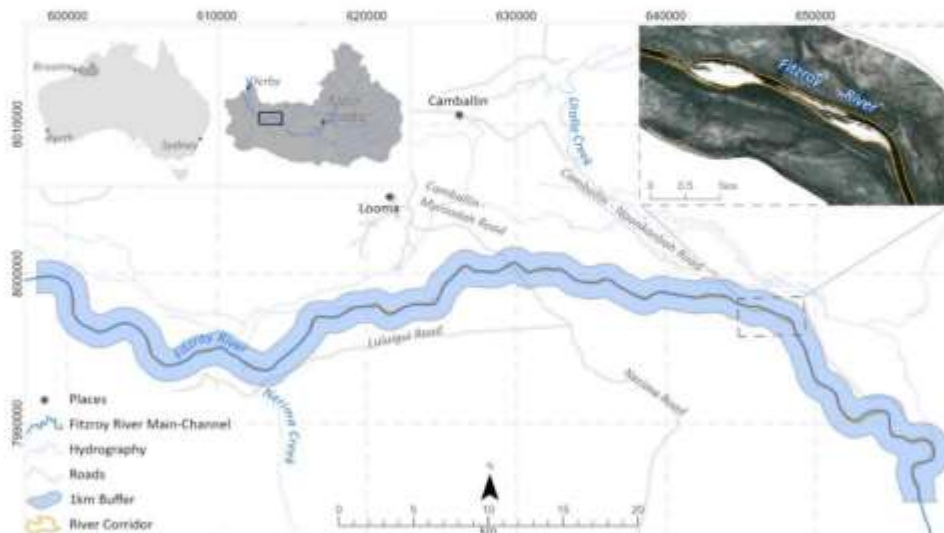


Figure 1. Location of the study area in the Fitzroy River, Kimberley, Western Australia. The inset dashed box on the right shows an example of the estimated Fitzroy River corridor.

This study focused on the lower Fitzroy River (Figure 1), where there is an extreme variation in water extent within and between years, as this region has little or no influence from groundwater discharge (Taylor et al. 2018). We deliberately selected an area of the river where longitudinal surface water connectivity breaks down during the dry season so that we could

detect narrow linear features (e.g., thalweg/run/riffle) and isolated pools. The marked variation of water extent through time benefits the testing of image classification accuracy as it helps cover different scenarios. Specifically, we chose time points to maximize variation in surface water within a hydrological year, such as mid-wet (February), early-dry (June), and late-dry season (October), and among years (2018-2020), given that in 2019 the catchment experienced the lowest (348mm) total rainfall of the measured record (2003–2022). The other years (2018 and 2020) presented total rainfall closer to the historical average (Department of Water and Environmental Regulations Water Information Reporting system). The river corridor extent was estimated using slope values from a Digital Elevation Model (DEM) derived from LiDAR (2 m) measurements provided by the Western Australian Department of Water and Environment Regulation (DWER). Values >20 degrees were considered likely riverbanks given the lower Fitzroy River has deeply scoured channels. After filtering slope values, the river corridor was visually defined and manually polygonized using ArcGIS Pro and Esri base maps with a total area of 10.08 km². Moreover, the river corridor was buffered on each side by 1 km to cover a wider area of interest (169.98 km²), or buffered zone, given the hydrological and ecological importance of the immediate floodplain (Figure. 1). The resulting polygon was used to clip PlanetScope, Sentinel-2, and sharpened images for further analysis.

Image Datasets and Data Processing Tools

Sentinel-2

We used nine cloud-free Sentinel-2 surface reflectance NBART images (Nadir BRDF Adjusted Reflectance + Terrain Illumination Correction) captured in February, June, and October of 2018, 2019, and 2020 downloaded from the Geoscience Australia - Digital Earth Australia (DEA) database (<http://www.ga.gov.au/dea> (accessed on February 8, 2021)) using the National Computational Infrastructure (NCI), reprojected to UTM/WGS84. Geoscience Australia used several pre-processing techniques for Sentinel-2 (F. Li et al. 2012, 2010; Vincenty 1975), with a detailed description of the steps at: <https://cmi.ga.gov.au/data-products/dea/190/dea-surface-reflectance-nbart-sentinel-2-msi#basics> (accessed on April 1, 2021).

PlanetScope

As a source of high spatial resolution imagery, we used cloudless Level-3B PlanetScope-0 analytic surface reflectance orthoscenes atmospherically corrected by Planet Labs to surface reflectance using the 6S radiative transfer model (Kotchenova et al. 2006) with ancillary data from MODIS. The PlanetScope-0 satellites have four spectral bands (Visible + Near-infrared - VNIR): Blue (B: 455–515 nm), Green (G: 500–590 nm), Red (R: 590–670 nm), and Near-infrared (NIR, 780–860 nm) with a spatial resolution of 3 x 3 m (Planet Team 2021). We used 82 images captured in February, June, and October of 2018, 2019, and 2020 (~9 tiles per month). The acquisition of images on dates close to Sentinel-2 image capture helps produce better sharpening quality results (Fonseca et al. 2011). PlanetScope images were acquired using an academic license provided by Planet Labs Inc (<https://planet.com> (accessed on February 12, 2021)).

Image Sharpening and Band Synthesizing

Li's et al. (2020) method for sharpening and synthesizing Sentinel-2 bands with PlanetScope was automated. We followed the same workflow found in Li et al. (2020) but used different packages intending to improve performance (see Data Availability Statement for script).

The first step of the automated sharpening process was to check if both images were in the same projection and if the reference image (Planetscope) was within the extent of the target image (Sentinel-2). Each band of the Sentinel-2 and the Planetscope images were co-registered using AROSICS (Scheffler et al. 2017) to ensure that all pixels were in the same geographic position. In the AROSICS co-registry process, the target image (Sentinel-2) is resampled to match the reference (Planetscope) pixel resolution (3 m). The co-registered Planetscope visible + near infra-red or VNIR (Red, Green, Blue and Near-infrared) bands were spatially degraded to Sentinel-2 spatial resolutions (i.e., 10 or 20 m) using an OpenCV 2D convolution filter (Bradski 2000) with a 41 x 41 matrix (-20 ... 20 – with center $i=0, j=0$). Each value of the convolutional matrix is defined by the following equations extracted or analytically derived from Li et al. (2020):

$$\varpi_{\lambda}(i, j) = \frac{1}{2\pi(\sigma_{\lambda}/pr)^2} e^{-\frac{i^2+j^2}{2(\sigma_{\lambda}/pr)^2}} \quad (1)$$

$$\sigma_{\lambda} = \frac{1}{\sigma_{f,\lambda} 2\pi} \quad (2)$$

$$\sigma_{f,\lambda} = \sqrt{-\frac{f^2}{2\ln(\text{MTF})}} \quad (3)$$

Where, $\varpi_{\lambda}(i, j)$ is the value of each cell in the convolution matrix, $\sigma_{f,\lambda}$ is the standard deviation (m^{-1}) of the modulation transfer function (MTF - values for each band published in (ESA 2021)), pr is the reference image (Planetscope) pixel spatial resolution value (m^{-1}), and f is the spatial frequency (m^{-1}), which depends on pixel resolution or the target image (i.e., $f=1/20 m^{-1}$ for 10 m bands and $f=1/40 m^{-1}$ for 20m bands). Before applying the convolution matrix to the respective Sentinel-2 band, λ , $\varpi_{\lambda}(i, j)$ values were normalized, so the 41 x 41 matrix sums to one ($1/\sum_{i=-20}^{20} \sum_{j=-20}^{20} \varpi_{\lambda}(i, j)$).

Because Planetscope is equipped with sensors to capture only four bands (VNIR), and water has a high absorption spectrum on the Short-wave Infrared (SWIR) band (Xu 2006), a 3 m equivalent SWIR band was synthesized from Sentinel-2 SWIR 1 (Band 11) and SWIR 2 (Band 12). For that, each of the Sentinel-2 SWIR bands was resampled and co-registered with Planetscope bands. While Jarihani et al. (2015) suggested calculating indices and then sharpening data for water footprint mapping, Planetscope lacks the same spectral bands as Sentinel-2 and necessitates the band synthesizing first, then index approach. All Planetscope bands were degraded to 20 m resolution using Equation (1) (with $f=1/40 m^{-1}$) and the respective SWIR MTF values to calculate the convolutional matrix. The degraded 20 m bands were used as input variables in multiple linear regression equations (Equation (5)), executed with Scikit-learn (Pedregosa et al. 2011), using each band as explanatory (or independent) variables and the 20 m Sentinel SWIR bands as the response (or dependent) variable. The regression coefficients were used in Equation (4) with the Planetscope bands to synthesize the SWIR bands. Equation (4) and Equation (5) (as per Li et al. (2020)) are shown below.

$$\hat{\rho}_{Planet_{\lambda}}^3(i, j) = \omega_{blue,\lambda} \rho_{Planet_{blue}}^3(i, j) + \omega_{green,\lambda} \rho_{Planet_{green}}^3(i, j) + \omega_{red,\lambda} \rho_{Planet_{red}}^3(i, j) + \omega_{NIR,\lambda} \rho_{Planet_{NIR}}^3(i, j) \quad (4)$$

Where, $\hat{\rho}_{Planet_{\lambda}}^3(i, j)$ is the synthetic Planetscope values for the considered band (λ ; e.g., SWIR 1, SWIR 2), $\omega_{blue,\lambda}$, $\omega_{green,\lambda}$, $\omega_{red,\lambda}$ and $\omega_{NIR,\lambda}$ are weights calculated as per Equation (5), and $\rho_{Planet_{blue}}^3$, $\rho_{Planet_{green}}^3$, $\rho_{Planet_{red}}^3$ and $\rho_{Planet_{NIR}}^3$ are the 3m Planetscope bands.

$$\rho_{S2_{\lambda}}^{20} = \omega_{blue,\lambda} \rho_{Planet_{blue}}^{20} + \omega_{green,\lambda} \rho_{Planet_{green}}^{20} + \omega_{red,\lambda} \rho_{Planet_{red}}^{20} + \omega_{NIR,\lambda} \rho_{Planet_{NIR}}^{20} \quad (5)$$

Where, $\rho_{S2\lambda}^{20}$ is the Sentinel-2 values for the considered band λ (e.g., SWIR 1, SWIR 2), ω is the multiple linear regression coefficient values calculated for each planet band (e.g., Green, Blue) and target synthesized band λ (e.g., SWIR 1, SWIR 2), and $\rho_{Planet_{blue}}^{20}$, $\rho_{Planet_{green}}^{20}$, $\rho_{Planet_{red}}^{20}$ and $\rho_{Planet_{NIR}}^{20}$ are the degraded to 20 m Planetscope bands.

Lastly, a high-pass modulation equation (HPM) (Schowengerdt 2007; Z. Li et al. 2020; Vivone et al. 2019) was applied for each VNIR band (Equation (6)) and to synthetic SWIR bands (Equation (7)) to finish the sharpening process.

$$\hat{\rho}_{S2\lambda}^3(i, j) = \frac{\rho_{S2\lambda}^{10 \rightarrow 3}(i, j)}{\rho_{Planet_{\lambda}}^{(3 \rightarrow 10)^{-3}}(i, j)} \times \rho_{Planet_{\lambda}}^3(i, j) \quad (6)$$

Where, $\hat{\rho}_{S2\lambda}^3(i, j)$ is the sharpened Sentinel-2 λ band (VNIR) at 3 m pixel resolution, $\rho_{S2\lambda}^{10 \rightarrow 3}(i, j)$ is the resampled Sentinel-2 λ band (10 m) to 3m, $\rho_{Planet_{\lambda}}^{(3 \rightarrow 10)^{-3}}(i, j)$ is the 3 m to 10 m spatially-degraded Planetscope values before being resampled back to 3 m, and $\rho_{Planet_{\lambda}}^3(i, j)$ is the Planetscope λ band at 3 m pixel resolution.

$$\hat{\rho}_{S2\lambda}^3(i, j) = \frac{\rho_{S2\lambda}^{20 \rightarrow 3}(i, j)}{\rho_{Planet_{\lambda}}^{(3 \rightarrow 20)^{-3}}(i, j)} \times \hat{\rho}_{Planet_{\lambda}}^3(i, j) \quad (7)$$

Where, $\hat{\rho}_{S2\lambda}^3(i, j)$ is the sharpened Sentinel-2 λ band (SWIR1, SWIR2) at 3 m pixel resolution, $\rho_{S2\lambda}^{20 \rightarrow 3}(i, j)$ is the resampled Sentinel-2 λ band (SWIR1, SWIR2) at 3 m, $\rho_{Planet_{\lambda}}^{(3 \rightarrow 20)^{-3}}(i, j)$ is the synthetic Planetscope λ band degraded to 20 m and resampled back to 3m and $\hat{\rho}_{Planet_{\lambda}}^3(i, j)$ is the synthetic Planetscope λ band.

Ground Truthing

The Fitzroy River is located in a very remote area with limited access to the river, making the collection of ground truth information on water extent logistically difficult, costly, and impossible across all images. Opportunistic hand-held GPS water extent points were collected during two fieldwork campaigns in October 2018 and September 2019. Although the ground truth points were insufficient to validate the entire area of interest, they aided in interpreting satellite imagery and creating hand-classified validation maps. Validation was also assisted by using high-resolution images (10 cm) from the Western Australian Land Information Authority (Landgate).

To overcome the scarcity of ground truth data, we manually polygonized the water extent for all images by interpreting areas identified as water. We used supporting data, including high-resolution aerial imagery and LiDAR DEM products (2 m), to understand preferential runoff flow paths and a combination of Sharpened Sentinel-2 with SWIR bands (3 m) and natural and false colors of Planetscope (3 m) to enhance contrast. The resulting hand-digitized polygons were smoothed using ArcGIS's 'smooth polygon' tool to give features a more natural and rounded appearance, as manual polygonizing can leave sharp and unnatural edges. The interpreted data was used as a ground truth reference in subsequent quantitative accuracy analysis.

Automated Sensitivity Analysis for the Water Detect Algorithm

The Water Detect algorithm (Cordeiro, Martinez, and Peña-Luque 2021) identifies surface water features by combining spectral water indices (e.g., Normalized Difference Water Index-NDWI, Modified Normalized Difference Water Index-MNDWI, and the Multiband Water

Index-MBWI) with VNIR-SWIR bands to highlight features of interest that are clustered based on multidimensional agglomerative clustering and a naive Bayesian classifier. In addition to choosing the main water spectral indices and the combination of bands, two other parameters can significantly influence the results of Water Detect: the maximum clustering and the regularization of the normalized spectral indices. These are particularly important for this study, given that the area of interest is a buffered corridor around the main river channel. In this case, the number of possible targets to be identified as clusters is lower than an entire Sentinel-2 scene, such as that used in the original paper by Cordeiro, Martinez, and Peña-Luque (2021). Moreover, the regularization of spectral water indices promotes a shrinkage of the water indices variance and can avoid a water cluster being split in two in the presence of different water constituents (i.e., organic and inorganic matter). The main input parameters for Water Detect must be configured in an initialization file (.ini), which holds all necessary inputs.

We developed a sensitivity analysis algorithm (see Data Availability Statement for script) to test all possible combinations of inputs parameters (i.e., spectral indices, maximum clustering, and regularization) for Water Detect within a specified range and assessed accuracy, determining the most accurate inputs for each specific case, and producing a most-to-least accuracy ranking. The developed sensitivity algorithm has four main inputs: 1) the Water Detect default initialization file as per Cordeiro et al. (2021); 2) the range of maximum clustering and regularization given by lowest, highest, and step values; 3) the images to be tested, and 4) the ground truth raster to be used in the accuracy assessment.

The first step to finding the most accurate Water Detect input parameters for each tested image was to calculate all possible combinations between the range of maximum clustering and regularization values and the images to be tested. We automatically changed the initialization file for each unique combination based on the number of bands, maximum clustering, and regularization and used the modified initialization file to execute Water Detect. The reference points (ground truth) were then compared with the classification results (water mask) using the Scikit-learn Metrics module and Matthews correlation coefficient (MCC) function (Pedregosa et al. 2011). The MCC is a performance measure of the quality of binary and multiclass classifications proposed by Matthews (1975) and revised by Baldi et al. (2000), which accommodated the Pearson correlation coefficient to assess the correlation in confusion matrices. The MCC is less influenced by imbalanced datasets, as it considers both accuracy and error rates and uses all confusion matrix values (Bekkar, Kheliouane Djemaa, and Akrouf Alitouche 2013). In intermittent rivers, the ratio between non-water and water classes will invariably become smaller as the dry season progresses and the relative number of water pixels decreases. Therefore, using metrics that account for skewed or biased datasets is necessary. It ranges from -1 (total disagreement between predicted scores and true labels' values) to 0 (prediction no better than random) to 1 perfect prediction (Fernández et al. 2018). A high MCC value means high accuracy and low misclassification of positive and negative classes (Chicco, Tötsch, and Jurman 2021).

To assess the Water Detect algorithm's input parameter values and resulting accuracy, we used 3, 9, 1 as lowest, highest and step values for maximum clustering (or 3, 4, 5, 6, 7, 8, 9), and 0.01, 0.1, 0.01 for regularization (or 0.01, 0.02 ... 0.09, 0.1). We tested five sensor configurations as input for the sensitivity analysis algorithm: (i) Planetscope (VNIR – NDWI/NIR); (ii) Sentinel-2 (VNIR – NDWI/NIR); (iii) Sharpened Sentinel-2 (VNIR – NDWI/NIR); (iv) Sentinel-2 (VNIR-SWIR – MNDWI/NDWI/SWIR-2); and (v) Sharpened Sentinel-2 (VNIR-Synthetic SWIR – MNDWI/NDWI/SWIR-2). The accuracy of the classified images was evaluated using all valid points retrieved from each reference ground truth raster (-19×106 points) and the respective values of those points in each classified raster. We, therefore, applied our sensitivity analysis algorithm across 3,150 configurations (i.e., 70

possible input combinations for each sensor and nine time periods – February, June, October of 2018, 2019, and 2020).

Comparing sensor performance

The water detection performance of different sensors (Planetscope – VNIR, Sentinel-2 – VNIR, Sharpened Sentinel-2 – VNIR, Sentinel-2 – VNIR-SWIR, and Sharpened Sentinel-2 – VNIR-Synthetic SWIR) were compared over various periods (February, June, October of 2018, 2019, and 2020). Firstly, we compared each sensor performance using all the combinations considered for the sensitivity analysis and then used box- and 3D- plots to aid in interpreting accuracy dispersion, patterns, and regions of high performance. Secondly, we compared the performance of the water detection results using the best fitting input parameters, identified by the developed sensitivity analysis, with global fitting parameters (inputs that presented high accuracy over all considered images) to confirm that the global fitting parameters had good adherence with optimum input parameters. We then compared the performance of each sensor using the global fitting parameters.

Although the sensitivity test algorithm can extract top-performing input combinations for each image, for reproducibility purposes, we also aimed to find a combination that would perform consistently well over multiple images and time periods for each sensor, i.e., to identify global fitting parameters. We identified the global fitting parameters for each sensor by averaging the MCC results over time for each combination of maximum clustering and regularization. In addition, we analyzed the top five parameter combinations for each time step and sensor.

The complete workflow for this study is presented in Figure 2.

Results

Water Detect Input Parameter Analytics and Accuracy Assessment

We found a higher level of agreement in the ability to detect water for VNIR compared to SWIR sensors. The Sharpened Sentinel-2 (VNIR) outperformed the other sensors (mean MCC: 0.746); followed by Sentinel-2 (VNIR) (0.721), Planetscope (VNIR) (0.704), Sharpened (SWIR) (0.55) and the Sentinel-2 (SWIR) (0.48). We observed a considerable range between maximum and minimum MCC values, with Sharpened (SWIR) presenting the highest range (1.135), followed by Planetscope (0.89), Sentinel-2 (SWIR) (0.842), Sharpened (VNIR) (0.794), and Sentinel-2 (VNIR) (0.741). Not surprisingly, the standard deviation was also high, with Sentinel-2 (SWIR) (std: 0.351) as the topmost, followed by Sharpened (SWIR) (0.336), Planetscope (0.185), Sharpened (VNIR) (0.179), and Sentinel-2 (VNIR) (0.165).

The correct use of input parameters for the 'Water Detect' algorithm had a greater influence on results than the choice of sensor. Even expecting variability, the wide range of accuracy values showed that finding the optimal input parameters is essential, as some combinations performed poorly and others outstandingly. This highlights the importance of using the sensitivity analysis algorithm developed here. As shown in Figure 3, the difference between the interquartile range (IQR) MCC values of images with SWIR bands was much more significant than images with only VNIR bands, which indicated that VNIR sensors were more robust and accurate for water identification using the Water Detect in the tested environment.

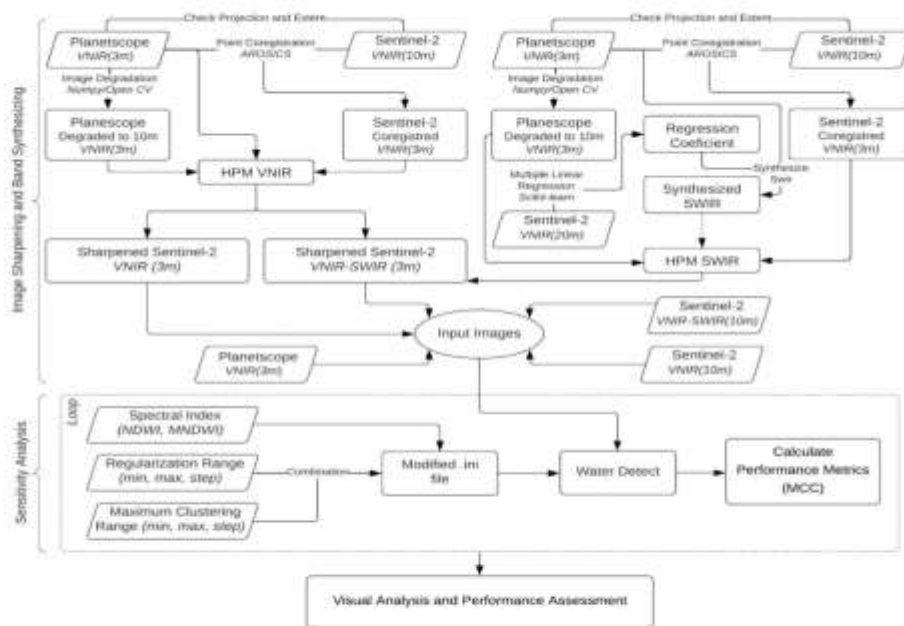


Figure 2. Processing workflow for Sharpening, Synthesizing bands, and Sensitivity analysis, from the raw input data to performance assessment.

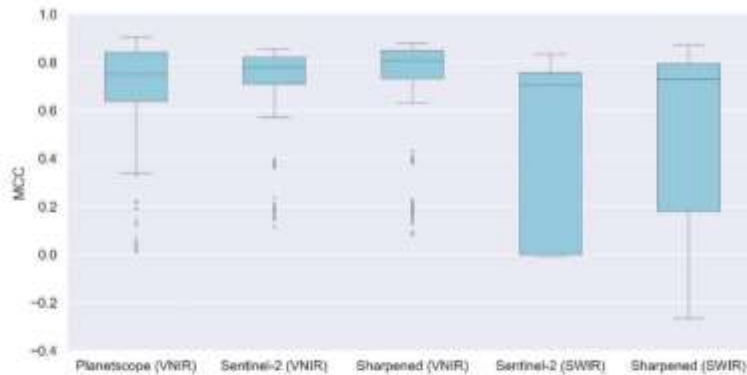


Figure 3. Boxplots represent all sensors' MCC median, interquartile range, extreme values, and outliers. High values indicate better performance, while narrow ranges indicate greater robustness over time.

Global Fitting Parameters

The overall best-fitting results for all sensors tended to gravitate to low maximum clustering and high regularization, although regularization seemed to have a lesser effect on the accuracy of water detection results (Figure 4). This finding was also true when evaluating only the top five combinations for each sensor. For instance, four out of five had a mean maximum number of clusters between 3 and 5, with Sentinel-2 (VNIR) presenting a value of 6. Similarly, for mean optimal regularization, four out of five sensors ranged from 0.07 and 0.1, with Sharpened (VNIR) presenting a value of 0.04. Furthermore, our data suggest that the best overall fitting (i.e. global conditions) for maximum clustering and regularization is: Planetscope (VNIR) (max. cluster: 3, regularization: 0.09), Sentinel-2 (VNIR) (6, 0.07), Sharpened (VNIR) (5, 0.04), Sentinel-2 (SWIR) (3, 0.08) and Sharpened (SWIR) (4, 0.1). We highlight that there was a minimal difference between the top five most accurate combinations of each sensor, with the MCC difference between the first and last (1 and 5) of 0.007 (Planetscope VNIR), 0.012 (Sentinel-2 VNIR), 0.017 (Sharpened VNIR), 0.006 (Sentinel-2 SWIR), and 0.023 (Sharpened SWIR). In this case, we could choose as global fitting parameters any of the top five most accurate input combinations for each sensor with almost no change in accuracy (< 2.3%).

The proposed global fitting parameters showed excellent adherence with individual fitting, suggesting that the Water Detect can perform close to maximum accuracy (Figure 5). Planetscope (VNIR), Sentinel-2 (VNIR), and Sharpened (VNIR) exhibit exceptional adherence with mean differences in MCC values of 0.014, 0.013, and 0.016, respectively. Conversely, Sentinel-2 (SWIR) and Sentinel-2 Sharpened (SWIR) showed the lowest adherence with a mean difference in MCC values of 0.037 and 0.053, respectively (Figure 5). Figure 6 shows the water detection results for the most accurate sensor and each time step, see how isolated pools and potential runs were successfully mapped.

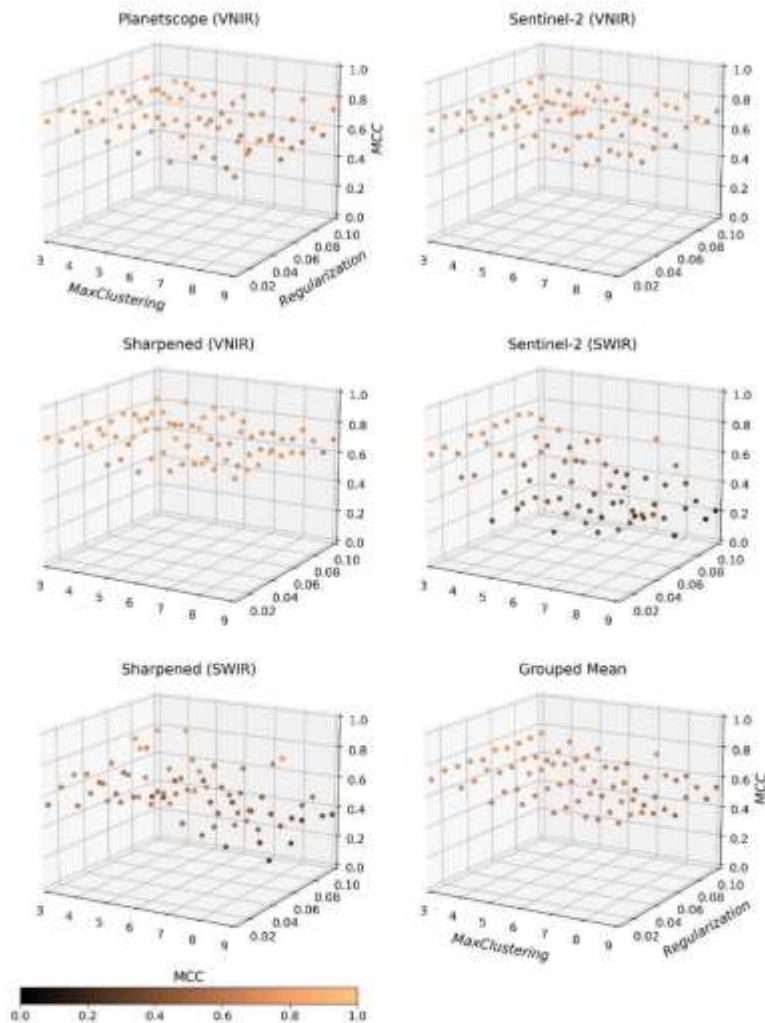


Figure 4. 3D plots comparing different maximum cluster and regularization parameters for each sensor with their MCC values. The grouped mean is shown on the lower left.

Comparison of Sensors using Global Fitting Parameters

To compare the primary sensors and the sharpened versions, we used the Global fitting parameters described above that represented the most overall accurate water extraction for each sensor. The Sharpened (VNIR) showed the highest MCC performance mean values (0.822), followed by Planetscope (VNIR) (0.809), Sentinel-2 (VNIR) (0.791), Sharpened (SWIR) (0.783), and finally Sentinel-2 (SWIR) (0.753) (Figure7). While the Sharpened (VNIR) presented the most consistent performance over time, Planetscope (VNIR) showed the highest MCC values in six of the nine time steps. However, it also had the highest standard deviation

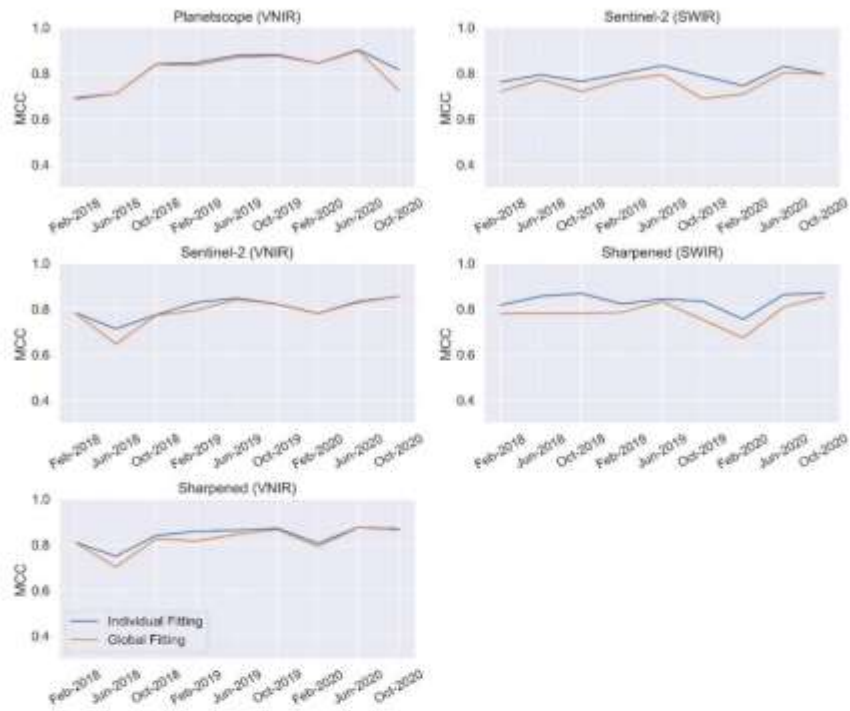


Figure 5. Comparison of performance between the most accurate images using individually-fitted parameter combinations and the global-fitting parameter combination. Horizontal lines show the adherence between the two fittings.

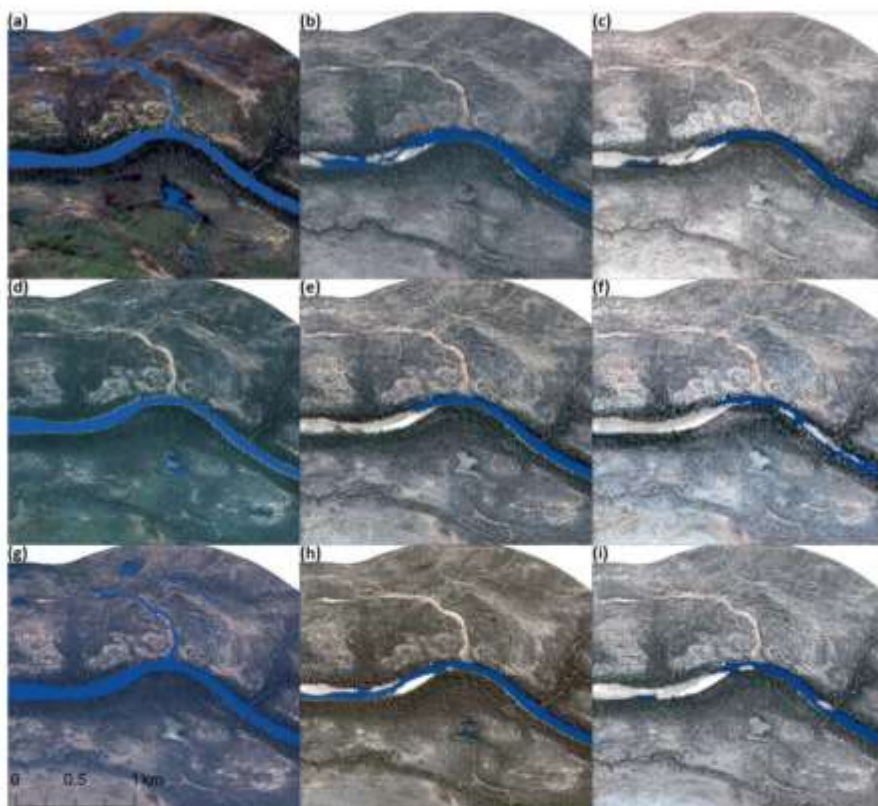


Figure 6. Example of best performance achieved across classified images for each timestep where blue indicates identified water with Planetscope as a base map in the background. Sharpened (VNIR) showed the best performance over the other considered sensors in February 2018 and October 2020 (a, i); Sharpened (SWIR) in June 2018 (b); and Planetscope in October 2018, February 2019, June 2019, October 2019, February 2020 and June 2020 (c, d, e, f, g, and h).

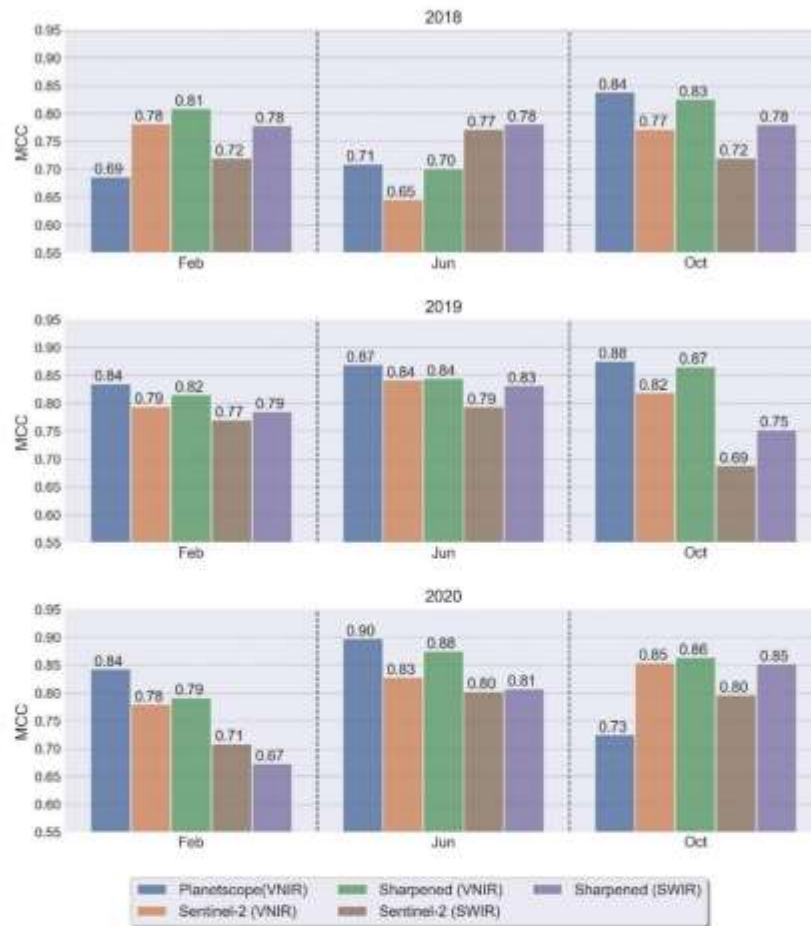


Figure 7. Comparison of sensor performance over time using the chosen Global Fitting Parameters. Bars show MCC values for each time and sensor, as per the legend on the bottom.

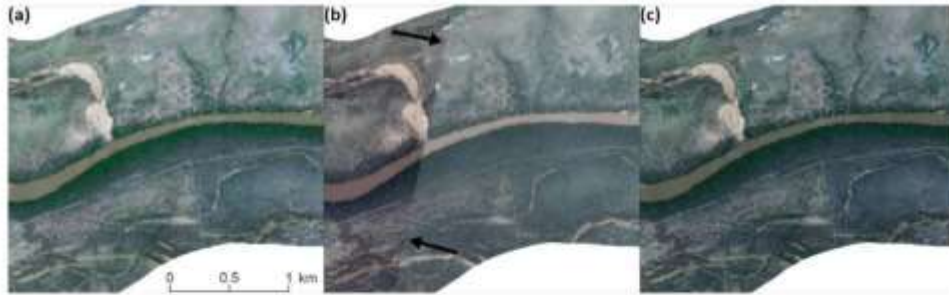


Figure 8. Visual comparison between Sentinel-2 (a) on the left, Planetscope (b) in the center, and Sharpened Sentinel-2 (c) on the right. The black arrows in the center indicate the difference between scenes (edge effect) when merging raw Planetscope imagery without any treatment.

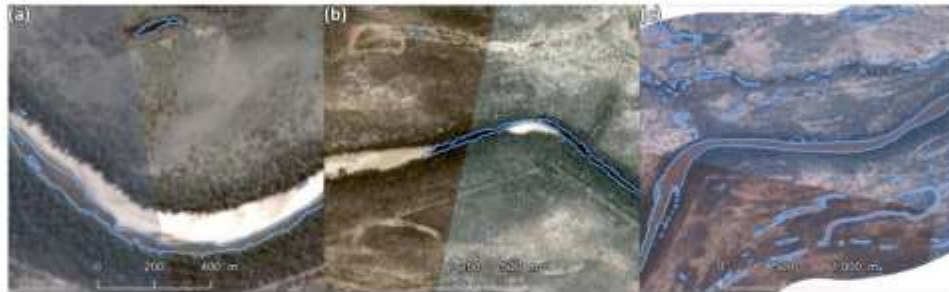


Figure 9. Water identification in raw merged Planetscope imagery at regions affected by edge effect (a, b, c). The difference between scenes did not seem to affect water identification outcomes in merged Planetscope images.

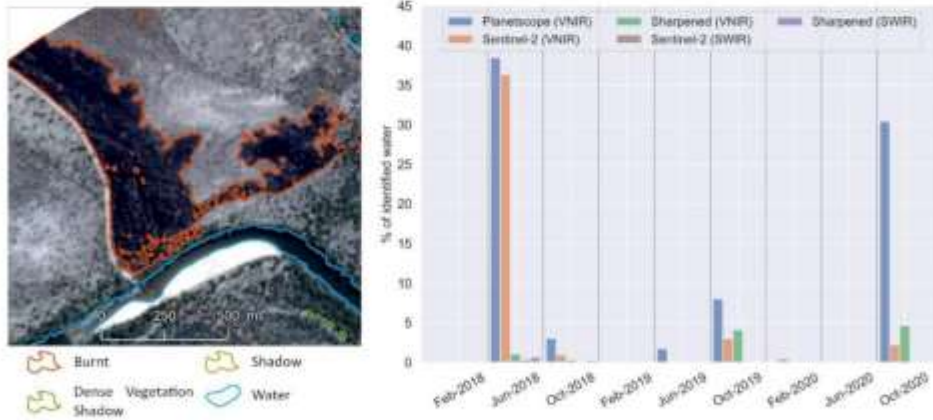


Figure 10. Most typical cases of misclassification – Burnt areas. On the left is an example showing burnt areas in orange, and on the right, a bar plot showing the percentage of burnt areas erroneously classified as water.

because of poor performances in Feb/2018 and Oct/2020, which can be due to sensor calibration issues and appears to be corrected in the sharpened image. Comparably, Sentinel-2 (VNIR) and Sharpened (SWIR) images also had excellent performance over time, with mean MCC values very close to Planetscope (VNIR). In fact, the Water Detect algorithm consistently identified water in all images with good performance, i.e., the difference between the highest mean (Sharpened VNIR) and the lowest (Sentinel-2 SWIR) was only 0.069 or 8%. Furthermore, the Sentinel-2 Sharpened (VNIR) MCC values appeared to follow the highest accuracy between their source images (i.e., Sentinel-2 (VNIR) and Planetscope (VNIR)), which suggests that the sharpening process can benefit water detection consistency over time and helps to explain its overall good performance. Even though in our case, the performance metrics were almost negligible between Sharpened (VNIR), Planetscope (VNIR), and Sentinel-2 (VNIR).

Visual Analysis

Visual assessment is an important step in image data validation as human interpretation can detect nuances that are not readily highlighted by examining image statistics. Our visual assessment indicated that the sharpening method seamlessly unified all Planetscope scenes (Figure 8). The resulting image looked more natural with obvious contrast when compared with the merged raw images. For example, in the area pointed by the black arrows in Figure 8B, the fusion results are clear, with smooth edges, obvious contrast, and critical information in the image can be clearly assessed. However, the edge effects did not seem to affect water identification outcomes in Planetscope images (Figure 9 a, b, c).

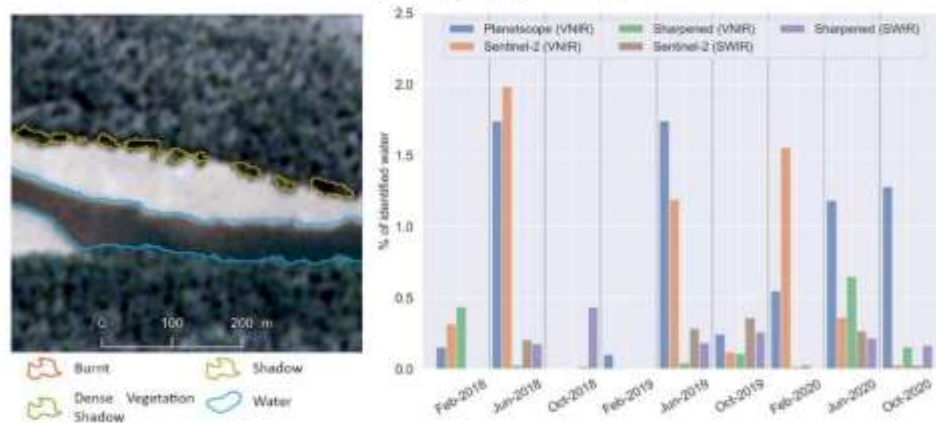


Figure 11. Most typical cases of misclassification – Heavily shaded areas. On the left is an example showing heavily shaded areas in yellow, and on the right, a bar plot showing the percentage of heavily shaded areas erroneously classified as water.

The Water Detect algorithm was able to identify surface water features in all images with excellent classification performance. This was also the case when narrow linear features between pools (i.e., runs) were analyzed, a critical feature in river fragmentation/connectivity studies. However, we observed a few misclassification cases, even when each sensor's best global fitting parameters were used. For instance, in all VNIR images (Planetscope, Sentinel-2, and Sharpened), burnt areas were erroneously classified as water. This can dramatically affect the classification accuracy, as seen in Figure 10. Also, heavily shaded areas were

frequently mistaken for water in VNIR images (Figure 11). In some SWIR (Sentinel-2 and Sharpened) images, parts of the riparian vegetation and a few spots with dense vegetation throughout the floodplain were misclassified as water, as were some sparse vegetation areas in drier months (June and October) (Figure 12).

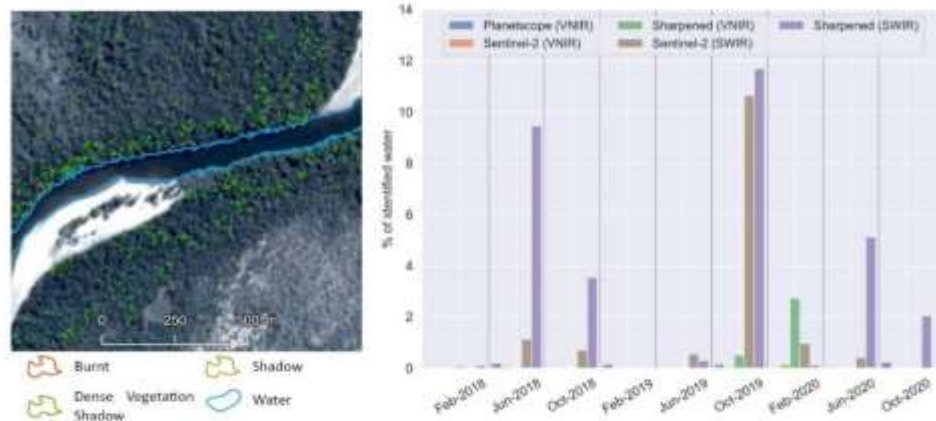


Figure 12. Most typical cases of misclassification – Dense vegetation. On the left is an example showing dense vegetation in green, and on the right, a bar plot showing the percentage of shadows on dense vegetation areas erroneously classified as water.

Discussion

This study of an intermittent river found that the ability to accurately detect surface water using multispectral imagery was maximized by identifying the optimal input parameters for the Water Detect algorithm and using VNIR sensors, with relatively little gained by using image sharpening routines. Our automated sensitivity test for the Water Detect algorithm was highly effective at identifying parameter inputs that maximized the algorithm's accuracy. The use of sub-optimal parameter inputs resulted in considerable declines in accuracy. We found a higher level of agreement in the ability to detect water for VNIR compared to SWIR sensors highlighting the possibility of accurate water detection using simple sensors. The benefits of sharpening Sentinel-2 images with Planetscope images were marginal, especially if the target area did not include heavily shadowed or extensive burnt areas; thus, sharpening should be considered only after weighing the benefits against the additional processing time. Using automated routines to sharpen imagery and determine input parameters for the Water Detect algorithm is a significant advancement that should increase our ability to accurately detect and map water in intermittent rivers.

The correct choice of input parameters for Water Detect considerably influenced the ability of the algorithm to accurately classify surface water in the study river. For instance, optimal parameters for clustering and regularisation identified by the developed sensitivity test algorithm increase accuracy by 130% compared to suboptimal values (MCC = 0.87 vs. -0.266, respectively) if considering the sensor with the largest accuracy range (Sharpened-SWIR). Specifically, we found the global optimum maximum clustering value to be 6 for Sentinel-2. This finding is in accordance with the results of Cordeiro, Martinez, and Peña-Luque (2021), where accuracy results for Sentinel-2 peaked and stabilized with maximum clustering values between 5 and 10. Our study is the first systematic test of regularization for the Water Detect algorithm, and our results suggest that in conjunction with maximum clustering, classification

accuracy is higher when regularization is also higher. However, this parameter varied greatly among individual fittings, suggesting that it would impact the classification performance less than the maximum clustering value. Ground-truthed data are required to confirm our sensitivity test and performance assessment. However, in the absence of such data, we recommend using our proposed global fitting parameters for each sensor and combinations of bands as a starting point. We are particularly confident in these parameter values for semi-arid locations, given the excellent performance achieved in our study. However, as our study was only undertaken in one location, we recognize that the global fitting parameters presented here might not be ideal for all uses and should be used mindfully.

Our automation of Li et al.'s (2020) sharpening routine is a significant advancement that makes it possible to easily and quickly sharpen many images or time series. This automation is advantageous due to its speed and also because it reduces the possibility of human error, which can arise when many images are sharpened manually. The resulting images were clear and seamlessly unified compared to raw merged Planetscope images and are visually comparable to the results achieved by Li et al. (2020). The sharpened images also outperformed, albeit slightly, raw merged images when identifying water, which is another indicator that the sharpening process successfully fused the best qualities of the source images.

While our automated image sharpening was successful, our study revealed that sharpening had a negligible effect on the accurate identification of water in most situations. For instance, the maximum increase in accuracy associated with sharpening in our study system was only 1.52%. Nonetheless, when one of the source images performed poorly, the sharpened image tended to perform as well as the most accurate source image (Sentinel-2 or Planetscope), making its accuracy more consistent over time. This can be attributed to a given sensor's spectral quality or spatial resolution since we used the best combination of both in the sharpened image. Depending on a project's objective, this could justify the additional processing time despite marginal improvement. Furthermore, VNIR images outperformed SWIR images in all scenarios but one, opening the possibility of using a plethora of images as input for Water Detect. This is an encouraging finding, given that four-band images are much more common and less expensive to capture.

Sentinel-2 (VNIR) images performed surprisingly well relative to Sentinel-2 sharpened (VNIR) and Planetscope (VNIR), even though its spatial resolution is more than three times as coarse. This is especially important as the marginal improvement in performance from Planetscope (VNIR) and Sentinel-2 sharpened (VNIR) incurs the expense of greater processing time and the financial cost of purchasing Planetscope versus the freely available Sentinel-2 imagery. When comparing Sentinel-2 (VNIR) to Sentinel-2 (SWIR), our results indicate that VNIR bands were more robust and performed better over time, which is a very relevant outcome since Cordeiro, Martinez, and Peña-Luque (2021) initially proposed the Water Detect algorithm to be used only with SWIR images. Although this allows the use of VNIR bands as an option or default in the subsequent upgrades of Water Detect, our study area is located in a remote semi-arid region with no build-up areas, which can impact the transferability for regional scales. Studies such as Xu (2006) and Zhang et al. (2011) suggest that MNDWI (SWIR bands) has the broadest applicability and is a better alternative in distinguishing water from shadow than NDWI, especially in areas of interest that include cities. Thus, future research should apply our sensitivity algorithm and workflow in areas with those characteristics to enable a comparison.

It is possible that the relatively large size of both the target river (Fitzroy River) and the study area minimized the difference in performance between high resolution (Planetscope and Sentinel-2 Sharpened) and coarser images (Sentinel-2), somewhat diminishing the overall

influence of spatial resolution needed for very narrow features. The Fitzroy River has a width of 100 m in wet months, and the catchment area is large. Locations where the classification would benefit from higher spatial resolution, such as in very small pools, and runs/riffles, were relatively minor, which does not significantly affect overall accuracy. However, even though we did not assess the detection accuracy of these important intermittent river features, results showed that by maximizing the overall detection accuracy, we also increased the potential of detecting them. Nonetheless, we believe Sentinel-2 is a solid option for many studies, especially on catchment or sub-basin scales. Of course, the extent of the area of interest and the size of target features will directly influence the choice of a particular spatial resolution to better represent each study's needs.

Even though all tested images demonstrated excellent overall performance in water detection, we observed three significant misclassification cases linked to dark surfaces: burnt areas, heavily shaded areas, and shaded areas in dense vegetation. The first two cases were more pronounced on VNIR images and the latter on SWIR images. Previous studies frequently report that shadows and dark surfaces, such as burnt areas, are often misclassified as water. For example, Xu (2006) noted that the NDWI was not appropriate for distinguishing water from shadows in built-up areas and proposed the MNDWI, which improved water discrimination over the NDWI in the presence of high turbidity waters and shadows in built-up areas. Zhang et al. (2011) found that the multi-band spectral relationships are more effective in dealing with shadow misclassification in mountainous areas, whereas MNDWI performed better in city areas. According to Zhai et al. (2015), this happens because NDWI values of water and shadows overlap, while MNDWI minimizes such problems with minimal intersections between shadow and other features. However, MNDWI still misclassifies vegetation in shaded areas as water, mainly because in such areas the SWIR band spectral value is usually smaller than the green band causing MNDWI values to increase (Guo et al. 2017). Burnt areas are also problematic, like shadows, due to their similar spectral reflectance to surface water, especially when dealing with 4-band images and NDWI. Nevertheless, unlike shadows, burnt regions are usually large and contiguous, severely influencing classification accuracy when present over extensive areas. Usually, in semi-arid regions, fires occur naturally or by controlled burning. In northern Australia, controlled fire techniques are commonly applied by Aboriginal communities to reduce the intensity of bushfires (Jackson, Finn, and Featherston 2012; Bird, Bliege Bird, and Parker 2005; McGregor et al. 2010). SWIR bands should be considered if attempting to detect water in these heavily burnt areas, especially when examining the floodplain since this is a lesser problem if only the river's main channel is of interest. A possible solution to minimize misclassification would be introducing other methods to the classification routine specifically designed to mask heavily shaded and burnt areas. This is possible (Mostafa 2017; Yamazaki, Liu, and Takasaki 2009; Pereira 2003; Roy et al. 2005; Frantz et al. 2016) but beyond the scope of the present study.

Conclusions

Our results revealed that the ability to accurately detect surface water in an intermittent river using multispectral imagery was maximized by identifying the optimal input parameters for the Water Detect algorithm. By maximizing the overall detection accuracy, we also increase the potential of detecting important intermittent river features, such as runs, riffles, and small isolated pools, which could be the focus of assessments in future studies. Furthermore, VNIR images outperformed SWIR images and should be considered an option in future versions of the 'Water Detect' algorithm. We also found a minor overall improvement in the performance of water detection after sharpening, but the benefits were marginal, especially if the target area did not include heavily shadowed or extensive burnt areas. Thus, sharpening and synthesizing

bands should be considered only after weighing the benefits against the additional processing time. Sentinel-2 performed well and proved a robust option for studies on catchment or sub-basin scales. Further research is needed to test the workflow presented in this study in other locations to establish if the tested images' performance is similar to our results. Furthermore, the developed sensitivity analysis algorithm should also be applied in other locations to refine the global fitting parameters.

Declaration of Interest Statement

The authors declare no conflict of interest.

Author Contributions

Conceptualization, TCT, MMD; methodology, TCT; software, TCT, ADNT, MCRC; writing—original draft preparation, TCT; writing—review and editing, TCT, MMD, LB, JNC, DM, MCRC, ADNT; visualization, TCT, ADNT. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

Code for Sharpening Sentinel-2 with Planetscope and Water Detect Sensitivity Test associated with the current submission is available at https://github.com/tayerthiaggio/Sharpenig-S2_Planetscope and <https://github.com/tayerthiaggio/WaterDetect-SensitivityTest>, respectively. Updates will also be published on GitHub, and the final DOI will be cited.

Acknowledgments

This project was carried out as part of the lead author's PhD research, with financial support from the Australian Government's Research Training Program (RTP) and the National Environmental Science Program (NESP) through its Northern Australia Environmental Resources Hub and the Australian Institute of Nuclear Science and Engineering through the Postgraduate Research Awards (PGRA).

References

- Baldi, Pierre, Soren Brunak, Yves Chauvin, Claus A.F. Andersen, and Henrik Nielsen. 2000. "Assessing the Accuracy of Prediction Algorithms for Classification: An Overview." *Bioinformatics (Oxford, England)* 16 (5). Bioinformatics: 412–24. doi:10.1093/BIOINFORMATICS/16.5.412.
- Beesley, Leah S., Bradley J. Pusey, Michael M. Douglas, Chris S. Keogh, Mark J. Kennard, Caroline A. Canham, Paul G. Close, Rebecca J. Dobbs, and Samantha A. Setterfield. 2021. "When and Where Are Catfish Fat Fish? Hydro-Ecological Determinants of Energy Reserves in the Fork-Tailed Catfish, *Neoarius Graeffei*, in an Intermittent Tropical River." *Freshwater Biology* 66 (6). John Wiley & Sons, Ltd: 1211–24. doi:10.1111/FWB.13711.
- Bekkar, Mohamed, DrHassiba Kheliouane Djemaa, and DrTaklit Akrouf Alitouche. 2013. "Evaluation Measures for Models Assessment over Imbalanced Data Sets." *Journal of Information Engineering and Applications* 3 (10). www.iiste.org.
- Bird, Douglas W, Rebecca Bliege Bird, and Christopher H Parker. 2005. "Aboriginal Burning Regimes and Hunting Strategies in Australia's Western Desert." *Human Ecology* 33 (4). doi:10.1007/s10745-005-5155-0.

- Bradski, G. 2000. "The OpenCV Library." Dr. Dobb's Journal of Software Tools. [https://www.scirp.org/\(S\(351jmbntvnsjt1aadkposzje\)\)/reference/ReferencesPapers.aspx?ReferenceID=1692176](https://www.scirp.org/(S(351jmbntvnsjt1aadkposzje))/reference/ReferencesPapers.aspx?ReferenceID=1692176).
- Callow, J. N., and G. S. Boggs. 2013. "Studying Reach-Scale Spatial Hydrology in Ungauged Catchments." *Journal of Hydrology* 496 (July). Elsevier: 31–46. doi:10.1016/j.jhydrol.2013.05.030.
- Chicco, Davide, Niklas Tötsch, and Giuseppe Jurman. 2021. "The Matthews Correlation Coefficient (Mcc) Is More Reliable than Balanced Accuracy, Bookmaker Informedness, and Markedness in Two-Class Confusion Matrix Evaluation." *BioData Mining* 14 (1). BioMed Central Ltd: 1–22. doi:10.1186/S13040-021-00244-Z/TABLES/5.
- Cordeiro, Maurício C.R., Jean Michel Martinez, and Santiago Peña-Luque. 2021. "Automatic Water Detection from Multidimensional Hierarchical Clustering for Sentinel-2 Images and a Comparison with Level 2A Processors." *Remote Sensing of Environment* 253 (February). Elsevier: 112209. doi:10.1016/j.rse.2020.112209.
- ESA. 2021. "Sentinel-2 L1C Data Quality Report," no. 65: 1–52.
- Fernández, Alberto, Salvador Garcia, Mikel Galar, Ronaldo C. Prati, Bartosz Krawczyk, and Francisco Herrera. 2018. "Performance Measures." *Learning from Imbalanced Data Sets*. Cham: Springer, Cham, 47–61. doi:10.1007/978-3-319-98074-4_3.
- Feyisa, Gudina L., Henrik Meilby, Rasmus Fensholt, and Simon R. Proud. 2014. "Automated Water Extraction Index: A New Technique for Surface Water Mapping Using Landsat Imagery." *Remote Sensing of Environment* 140 (January). Elsevier: 23–35. doi:10.1016/j.rse.2013.08.029.
- Fisher, Adrian, Neil Flood, and Tim Danaher. 2016. "Comparing Landsat Water Index Methods for Automated Water Classification in Eastern Australia." *Remote Sensing of Environment* 175 (March). Elsevier: 167–82. doi:10.1016/j.rse.2015.12.055.
- Fonseca, Leila, Laercio Namikawa, Emiliano Castejon, Lino Carvalho, Carolina Pinho, and Aylton Pagamisse. 2011. "Image Fusion for Remote Sensing Applications." In *Image Fusion and Its Applications*. InTech. doi:10.5772/22899.
- Frantz, David, Marion Stellmes, Achim Röder, Joachim Hill, David Frantz, Marion Stellmes, Achim Röder, and Joachim Hill. 2016. "Fire Spread from MODIS Burned Area Data: Obtaining Fire Dynamics Information for Every Single Fire." *International Journal of Wildland Fire* 25 (12). CSIRO PUBLISHING: 1228–37. doi:10.1071/WF16003.
- Gao, Bo Cai. 1996. "NDWI - A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space." *Remote Sensing of Environment* 58 (3). Elsevier: 257–66. doi:10.1016/S0034-4257(96)00067-3.
- Guo, Qiandong, Ruiliang Pu, Jialin Li, and Jun Cheng. 2017. "A Weighted Normalized Difference Water Index for Water Extraction Using Landsat Imagery." *International Journal of Remote Sensing* 38 (19). Taylor & Francis: 5430–45. doi:10.1080/01431161.2017.1341667.
- Huang, H., and D.P. Roy. 2021. "Characterization of Planetscope-0 Planetscope-1 Surface Reflectance and Normalized Difference Vegetation Index Continuity." *Science of Remote Sensing* 3 (June). Elsevier BV: 100014. doi:10.1016/j.srs.2021.100014.

- Jackson, Sue, Marcus Finn, and Pippa Featherston. 2012. "Aquatic Resource Use by Indigenous Australians in Two Tropical River Catchments: The Fitzroy River and Daly River." *Human Ecology* 40 (6): 893–908. doi:10.1007/s10745-012-9518-z.
- Jarihani, Abdollah A., Joshua R. Larsen, John N. Callow, Tim R. McVicar, and Kasper Johansen. 2015. "Where Does All the Water Go? Partitioning Water Transmission Losses in a Data-Sparse, Multi-Channel and Low-Gradient Dryland River System Using Modelling and Remote Sensing." *Journal of Hydrology* 529 (October). Elsevier: 1511–29. doi:10.1016/J.JHYDROL.2015.08.030.
- Jiang, Hao, Min Feng, Yunqiang Zhu, Ning Lu, Jianxi Huang, and Tong Xiao. 2014. "An Automated Method for Extracting Rivers and Lakes from Landsat Imagery." *Remote Sensing* 2014, Vol. 6, Pages 5067-5089 6 (6). Multidisciplinary Digital Publishing Institute: 5067–89. doi:10.3390/RS6065067.
- Kaplan, Gordana, and Ugur Avdan. 2018. "Sentinel-2 Pan Sharpening - Comparative Analysis." In *Proceedings*, 2:5158. MDPI AG. doi:10.3390/eecs-2-05158.
- Koohafkan, P, and B A Stewart. 2008. "Drylands, People and Land Use." In *Water and Cereals in Drylands*, edited by P Koohafkan and B A Stewart, 5–15. London: The Food and Agriculture Organization of the United Nations and Earthscan. www.earthscan.co.uk.
- Kotchenova, Svetlana Y., Eric F. Vermote, Jr. Frank J. Klemm, and Raffaella Matarrese. 2006. "Validation of a Vector Version of the 6S Radiative Transfer Code for Atmospheric Correction of Satellite Data. Part I: Path Radiance." *Applied Optics* 45 (26). Optical Society of America: 6762–74. doi:10.1364/AO.45.006762.
- Lehner, Bernhard, Catherine Reidy Liermann, Carmen Revenga, Charles Vörösmarty, Balazs Fekete, Philippe Crouzet, Petra Döll, et al. 2011. "High-Resolution Mapping of the World's Reservoirs and Dams for Sustainable River-Flow Management." *Frontiers in Ecology and the Environment* 9 (9). John Wiley & Sons, Ltd: 494–502. doi:10.1890/100125.
- Li, Fuqin, David L.B. Jupp, Medhavy Thankappan, Leo Lymburner, Norman Mueller, Adam Lewis, and Alex Held. 2012. "A Physics-Based Atmospheric and BRDF Correction for Landsat Data over Mountainous Terrain." *Remote Sensing of Environment* 124 (September). Elsevier: 756–70. doi:10.1016/J.RSE.2012.06.018.
- Li, Fuqin, Leo Lymburner, Norman Mueller, Peter Tan, Anisul Islam, David L.B. Jupp, and Shanti Reddy. 2010. "An Evaluation of the Use of Atmospheric and BRDF Correction to Standardize Landsat Data." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 3 (3): 257–70. doi:10.1109/JSTARS.2010.2042281.
- Li, Zhongbin, Hankui K. Zhang, David P. Roy, Lin Yan, and Haiyan Huang. 2020. "Sharpening the Sentinel-2 10 and 20 m Bands to PlanetScope-0 3 m Resolution." *Remote Sensing* 12 (15). MDPI AG: 2406. doi:10.3390/RS12152406.
- Liu, Zhaofei, Zhijun Yao, and Rui Wang. 2021. "Assessing Methods of Identifying Open Water Bodies Using Landsat 8 OLI Imagery." *Environmental Earth Sciences* 75. Accessed December 23. doi:10.1007/s12665-016-5686-2.
- Marcus, W. Andrew, and Mark A. Fonstad. 2008. "Optical Remote Mapping of Rivers at Sub-Meter Resolutions and Watershed Extents." *Earth Surface Processes and Landforms* 33 (1). John Wiley & Sons, Ltd: 4–24. doi:10.1002/ESP.1637.

- Matthews, B. W. 1975. "Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme." *Biochimica et Biophysica Acta (BBA) - Protein Structure* 405 (2). Elsevier: 442–51. doi:10.1016/0005-2795(75)90109-9.
- McGregor, Sandra, Violet Lawson, Peter Christophersen, Rod Kennett, James Boyden, Peter Bayliss, Adam Liedloff, Barbie McKaige, and Alan N. Andersen. 2010. "Indigenous Wetland Burning: Conserving Natural and Cultural Resources in Australia's World Heritage-Listed Kakadu National Park." *Human Ecology* 38 (6). Springer New York LLC: 721–29. doi:10.1007/S10745-010-9362-Y/FIGURES/9.
- Mostafa, Yasser. 2017. "A Review on Various Shadow Detection and Compensation Techniques in Remote Sensing Images." <https://doi.org/10.1080/07038992.2017.1384310> 43 (6). Taylor & Francis: 545–62. doi:10.1080/07038992.2017.1384310.
- Mueller, N., A. Lewis, D. Roberts, S. Ring, R. Melrose, J. Sixsmith, L. Lymburner, et al. 2016. "Water Observations from Space: Mapping Surface Water from 25 Years of Landsat Imagery across Australia." *Remote Sensing of Environment* 174 (March). Elsevier: 341–52. doi:10.1016/J.RSE.2015.11.003.
- Pedregosa, Fabian, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12 (85): 2825–30.
- Peña-Luque, Santiago, Sylvain Ferrant, Mauricio C.R. Cordeiro, Thomas Ledauphin, Jerome Maxant, and Jean Michel Martinez. 2021. "Sentinel-1&2 Multitemporal Water Surface Detection Accuracies, Evaluated at Regional and Reservoirs Level." *Remote Sensing* 13 (16). Multidisciplinary Digital Publishing Institute: 3279. doi:10.3390/RS13163279.
- Pereira, José M.C. 2003. "Remote Sensing of Burned Areas in Tropical Savannas." *International Journal of Wildland Fire* 12 (4). CSIRO PUBLISHING: 259–70. doi:10.1071/WF03028.
- Planet Team. 2021. "Planet Application Program Interface: In Space for Life on Earth." San Francisco, CA. <https://api.planet.com>.
- Qiao, Cheng, Jiancheng Luo, Yongwei Sheng, Zhanfeng Shen, Zhiwen Zhu, and Dongping Ming. 2012. "An Adaptive Water Extraction Method from Remote Sensing Image Based on NDWI." *Journal of the Indian Society of Remote Sensing* 40 (3). Springer: 421–33. doi:10.1007/s12524-011-0162-7.
- Roy, D. P., Y. Jin, P. E. Lewis, and C. O. Justice. 2005. "Prototyping a Global Algorithm for Systematic Fire-Affected Area Mapping Using MODIS Time Series Data." *Remote Sensing of Environment* 97 (2). Elsevier: 137–62. doi:10.1016/J.RSE.2005.04.007.
- Scheffler, Daniel, André Hollstein, Hannes Diedrich, Karl Segl, and Patrick Hostert. 2017. "AROSICS: An Automated and Robust Open-Source Image Co-Registration Software for Multi-Sensor Satellite Data." *Remote Sensing* 9 (7). MDPI AG: 676. doi:10.3390/rs9070676.
- Schowengerdt, Robert A. 2007. "Remote Sensing: Models and Methods for Image Processing." Academic Press, 515.
- Taylor, Andrew R, Glenn A Harrington, Steven Clohessy, Warrick R Dawes, Russell S Crosbie, Rebecca C Doble, Daniel L Wohling, et al. 2018. "Hydrogeological Assessment of the Grant Group and Poole Sandstone-Fitzroy Catchment, Western Australia." A technical report to the Australian Government from the CSIRO Northern Australia Water

Resource Assessment, part of the National Water Infrastructure Development Fund: Water Resource Assessments, CSIRO, Australia. <https://publications.csiro.au/rpr/download?pid=csiro:EP183648&dsid=DS5>.

- Tottrup, Christian, Daniel Druce, Rasmus Probst Meyer, Mads Christensen, Michael Riffler, Bjoern Dulleck, Philipp Rastner, et al. 2022. "Surface Water Dynamics from Space: A Round Robin Intercomparison of Using Optical and SAR High-Resolution Satellite Observations for Regional Surface Water Detection." *Remote Sensing 2022, Vol. 14, Page 2410* 14 (10). Multidisciplinary Digital Publishing Institute: 2410. doi:10.3390/RS14102410.
- Vincenty, T. 1975. "Direct and Inverse Solutions of Geodesics on the Ellipsoid with Application of Nested Equations." *Survey Review* 23 (176). Taylor & Francis: 88–93. doi:10.1179/sre.1975.23.176.88.
- Vivone, Gemine, Luciano Alparone, Andrea Garzelli, and Simone Loli. 2019. "Fast Reproducible Pansharpening Based on Instrument and Acquisition Modeling: AWLP Revisited." *Remote Sensing 2019, Vol. 11, Page 2315* 11 (19). Multidisciplinary Digital Publishing Institute: 2315. doi:10.3390/RS11192315.
- Vörösmarty, C. J., P. B. McIntyre, M. O. Gessner, D. Dudgeon, A. Prusevich, P. Green, S. Glidden, et al. 2010. "Global Threats to Human Water Security and River Biodiversity." *Nature* 2010 467:7315 467 (7315). Nature Publishing Group: 555–61. doi:10.1038/nature09440.
- Wang, Xiaobiao, Shunping Xie, Xueliang Zhang, Cheng Chen, Hao Guo, Jinkang Du, and Zheng Duan. 2018. "A Robust Multi-Band Water Index (MBWI) for Automated Extraction of Surface Water from Landsat 8 OLI Imagery." *International Journal of Applied Earth Observation and Geoinformation* 68 (June). Elsevier: 73–91. doi:10.1016/J.JAG.2018.01.018.
- Xu, Hanqiu. 2006. "Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery." *International Journal of Remote Sensing* 27 (14). Taylor and Francis Ltd.: 3025–33. doi:10.1080/01431160600589179.
- Yamazaki, Fumio, Wen Liu, and Makiko Takasaki. 2009. "Characteristics of Shadow and Removal of Its Effects for Remote Sensing Imagery." *International Geoscience and Remote Sensing Symposium (IGARSS)* 4. doi:10.1109/IGARSS.2009.5417404.
- Yang, Xiucheng, Shanshan Zhao, Xuebin Qin, Na Zhao, Ligang Liang, Claudia Kuenzer, Deepak R Mishra, Weimin Huang, and Prasad S Thenkabail. 2017. "Mapping of Urban Surface Water Bodies from Sentinel-2 MSI Imagery at 10 m Resolution via NDWI-Based Image Sharpening." *Remote Sensing* 9 (596): 19. doi:10.3390/rs9060596.
- Zhai, Ke, Xiaoqing Wu, Yuanwei Qin, and Peipei Du. 2015. "Comparison of Surface Water Extraction Performances of Different Classic Water Indices Using OLI and TM Imageries in Different Situations." *Geo-Spatial Information Science* 18 (1). Taylor & Francis: 32–42. doi:10.1080/10095020.2015.1017911.
- Zhang, Fang Fang, Bing Zhang, Jun Sheng Li, Qian Shen, Yuanfeng Wu, and Yang Song. 2011. "Comparative Analysis of Automatic Water Identification Method Based on Multispectral Remote Sensing." *Procedia Environmental Sciences* 11 (PART C). Elsevier: 1482–87. doi:10.1016/J.PROENV.2011.12.223.