Exploitation of high spatial, spectral and temporal resolution Earth observation imagery for large area land cover estimation

Isabel Rodes Arnau

THÈSE

En vue de l'obtention du

DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par : *l'Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)*

Présentée et soutenue le *10/11/2016* **par :**

Isabel RODES ARNAU

Estimation de l'occupation des sols à grande échelle pour l'exploitation d'images d'observation de la Terre à hautes résolutions spatiale, spectrale et temporelle

Exploitation of high spatial, spectral and temporal resolution Earth observation imagery for large area land cover estimation

---

JURY

Mme. Agnès BEGUE	Directrice de Recherche CIRAD	Rapporteur
M. Pierre GANÇARSKI	Professeur Université Strasbourg	Rapporteur
M. Jean-Philippe GASTELLU	Professeur Université Paul Sabatier	Président
M. Olivier HAGOLLE	Ingénieur CNES	Invité
M. Jordi INGLADA	Ingénieur CNES	Directeur
M. Grégoire MERCIER	Professeur Télécom Bretagne	Rapporteur
Mme. Silvia VALERO	Maître de Conférences Université Paul Sabatier	Examinateur

---

École doctorale et spécialité :

SDU2E : Océan, Atmosphère et Surfaces Continentales

Unité de Recherche :

Centre d'Études Spatiales de la BIOsphère (CESBIO UMR 5126)

Directeur de Thèse :

M. Jordi INGLADA

Rapporteurs :

Mme. Agnès BEGUE (Centre de Coopération Internationale en Recherche Agronomique pour le Développement, CIRAD), M. Pierre GANÇARSKI (Université Strasbourg) et *M. Grégoire MERCIER (École nationale supérieure des télécommunications de Bretagne)*

À mes parents Maribel et Juan Manuel,
à Marc et Alex,
et à Juanma et Cristina,
je dédie ce travail.

Acknowledgements

I would like to thank here all those that have contributed to make this thesis possible.

Thank you to the members of the jury, for the time you have devoted to this work, your interest and your involvement.

A heartfelt thank you to the former director of CESBIO Yann Kerr. I am sincerely grateful for your crucial support and for your advice when it was most needed.

Thank you to my colleagues in connection with the research laboratory CESBIO, who have contributed in many ways to this work. I have been lucky to share some years in Toulouse with you, and even luckier to have left France knowing I can count you among my friends. Thank you Marcela, Simone, Amanda, Malika, Mireille, Mari, Stéphanie, Jérôme, Arnaud, Ludo, Stéphane, Cédric, Ian, Clément, David, Julien, and Malik. Thank you also to the rest of colleagues of the laboratory, for such a pleasant, rich and interesting atmosphere.

Thank you to my parents Maribel and Juan Manuel. It is your support throughout the years and particularly during the duration of this work that has led me to the writing of these words. Thank you for all the efforts that you have done for me and that you continue doing, for your love, your positive approach to life and your encouragement. You are my pillars; I simply owe you everything.

Thank you to my husband Alex, my brother Juanma and my sister-in-law Cristina for your support and your love; thank you for being there.

Last but not least, thank you to my son Marc, for the happiness and strength that your gestures give me; I completed this work for you.

Contents

List of Figures

List of Tables

Abstract

The new generation Earth observation missions such as Sentinel-2 (a twin-satellite initiative prepared by the European Space Agency, ESA, in the frame of the Copernicus programme, previously known as Global Monitoring for Environment and Security or GMES) and Venμs, jointly developed by the French Space Agency (*Centre National d'Études Spatiales*, CNES) and the Israeli Space Agency (ISA), will revolutionize present-day environmental monitoring with the yielding of unseen volumes of data in terms of spectral richness, temporal revisit and spatial resolution. Venμs will deliver images in 12 spectral bands from 412 to 910 nm, a repetitivity of 2 days, and a spatial resolution of 10 m; the twin Sentinel-2 satellites will provide coverage in 13 spectral bands from 443 to 2200 nm, with a repetitivity of 5 days, and spatial resolutions of 10 to 60m.

The efficient production of land cover maps based on the exploitation of such volumes of information for large areas is challenging both in terms of processing costs and data variability. In general, conventional methods either make use of supervised approaches (too costly in terms of manual work for large areas), target specialised local models for precise problem areas (not applicable to other terrains or applications), or include complex physical models with inhibitory processing costs. These existent present-day approaches are thus inefficient for the exploitation of the new type of data that the new missions will provide, and a need arises for the implementation of accurate, fast and minimally supervised methods that allow for generalisation to large scale areas with high resolutions.

In order to allow for the exploitation of the previously described volumes of data, the objective of this thesis is the conception, design, and validation of a fully automatic approach that allows the estimation of large-area land cover with high spatial, spectral and temporal resolution Earth observation imagery, being generalisable to different landscapes, and offering operational computation times with simulated satellite data sets, in preparation of the coming missions. This approach is based on the integration of data processing algorithms, such as model learning and classification techniques, and land cover background knowledge on ecological and agricultural issues such as vegetation growth impact variables and practices. For example, the new introduction of temperature as a temporal axis for subsequent model learning incorporates an established vegetation growth factor to machine learning techniques for characterization across landscapes.

In particular, attention is paid to deal with different issues, such as automation, missing information (determined by satellite tracks, cloud reflectance effects, shadows and snow presence), limited learning and validation data, irregular time sampling (different number of images available

for each period and region, and unevenly distributed in time), data variability, and the possibility of working with different data sets and nomenclatures. To this aim, a processing methodology has been devised, automatic and not requiring any date selection or user manual work, and allowing the use of different nomenclatures and data sets. Thus, the developed methodology demonstrates the feasibility of automatic production of large scale land cover maps, dealing with data volume and variability, limited reference data and missing information, and integrating specialised knowledge and different data processing algorithms.

This work is framed by the research support programme *Terre, Océan, Surfaces Continentales, Atmosphère (TOSCA) of the Centre National d'Études Spatiales*, in relation to three research proposals: *Production de cartes d'occupation des sols à échelle régionale a partir de données SPOT HRV multitemporelles* by Jordi Inglada, and *Traitement et utilisation de séries temporelles d'images LANDSAT pour préparer le traitement et les applications de Venμs et Sentinel-2* and *Centre D'ALgorithmie et de VAlidation des Données Venμs et Sentinel-2 (CALVADOVS)* by Olivier Hagolle. It is dependent of the doctoral school Sciences de l'Univers, de l'Environnement et de l'Espace (SDU2E) of Paul Sabatier University and it has been developed in the laboratory Centre d'Études Spatiales de la BIOsphère (CESBIO) in Toulouse, France.

Keywords: remote sensing, land cover, satellite time series, multispectral data, large area mapping, automatic classification, Sentinel-2, Landsat, missing data, cloud and snow cover, temporal interpolation, stratification, spatial sampling, non-linear temporal sampling, vegetation phenology, temperature

Résumé

Les missions spatiales d'observation de la Terre de nouvelle génération telles que Sentinel-2 (préparé par l'Agence Spatiale Européenne ESA dans le cadre du programme Copernicus, auparavant appelé Global Monitoring for Environment and Security ou GMES) ou Venμs, conjointement développé par l'Agence Spatiale Française (Centre National d 'Études Spatiales CNES) et l'Agence Spatiale Israélienne (ISA), vont révolutionner la surveillance de l'environnement d' aujourd'hui avec le rendement de volumes inédits de données en termes de richesse spectrale, de revisite temporelle et de résolution spatiale. Venμs livrera des images dans 12 bandes spectrales de 412 à 910 nm, une répétitivité de 2 jours et une résolution spatiale de 10 m; les satellites jumeaux Sentinel-2 assureront une couverture dans 13 bandes spectrales de 443 à 2200 nm, avec une répétitivité de 5 jours, et des résolutions spatiales de 10 à 60m.

La production efficace de cartes d'occupation des sols basée sur l'exploitation de tels volumes d'information pour grandes surfaces est un défi à la fois en termes de coûts de traitement mais aussi de variabilité des données. En général, les méthodes classiques font soit usage des approches surveillées (trop coûteux en termes de travaux manuels pour les grandes surfaces), ou soit ciblent des modèles locaux spécialisés pour des problématiques précises (ne s'appliquent pas à autres terrains ou applications), ou comprennent des modèles physiques complexes avec coûts de traitement rédhibitoires. Ces approches existantes actuelles sont donc inefficaces pour l'exploitation du nouveau type de données que les nouvelles missions fourniront, et un besoin se fait sentir pour la mise en œuvre de méthodes précises, rapides et peu supervisées qui permettent la généralisation à l'échelle de grandes zones avec des résolutions élevées.

Afin de permettre l'exploitation des volumes de données précédemment décrits, l'objectif de ce travail est la conception et validation d'une approche entièrement automatique qui permet l'estimation de la couverture terrestre de grandes surfaces avec imagerie d'observation de la Terre de haute résolution spatiale, spectrale et temporelle, généralisable à des paysages différents, et offrant un temps de calcul opérationnel avec ensembles de données satellitaires simulés, en préparation des prochaines missions. Cette approche est basée sur l'intégration d'algorithmes de traitement de données, tels que les techniques d'apprentissage de modèles et de classification, et des connaissances liées à l'occupation des sols sur des questions écologiques et agricoles, telles que les variables avec un impact sur la croissance de la végétation ou les pratiques de production. Par exemple, la nouvelle introduction de température comme axe temporel pour un apprentissage des modèles ultérieurs intègre un facteur établi de la croissance de la végétation à des techniques d'apprentissage automatiques pour la caractérisation des paysages.

Une attention particulière est accordée au traitement de différentes questions, telles que l'automatisation, les informations manquantes (déterminées par des passages satellitaires, des effets de réflexion des nuages, des ombres ou encore la présence de neige), l'apprentissage et les données de validation limitées, les échantillonnages temporels irréguliers (différent nombre d'images disponible pour chaque période et région, données inégalement réparties dans le temps), la variabilité des données, et enfin la possibilité de travailler avec différents ensembles de données et nomenclatures. Dans ce but, une méthodologie de traitement a été mise au point, totalement automatique, où aucune sélection de date ou travail manuel de l'utilisateur est nécessaire, et permettant donc l'utilisation de nomenclatures et ensembles de données différentes. Ainsi, la méthodologie développée démontre la faisabilité de la production automatique de cartes d'occupation du sol à grande échelle, traitant le volume et la variabilité des données, les données de référence limitées et les informations manquantes, en intégrant des connaissances spécialisées et différents algorithmes de traitement de données.

Ce travail est encadré par le programme de soutien à la recherche «Terre, Océan, Surfaces Continentales, Atmosphère» (TOSCA) du Centre National d 'Études Spatiales, par rapport à trois projets de recherche: *Production de cartes d'occupation des sols à échelle régionale a partir de données SPOT HRV multitemporelles* par Jordi Inglada, et *Traitement et utilisation de séries temporelles d'images LANDSAT pour préparer le traitement et les applications de Venμs et Sentinel-2* et *Centre D'ALgorithmie et de VAlidation des Données Venμs et Sentinel-2 (CALVADOVS)* par Olivier Hagolle. Il dépend de l'école doctorale Sciences de l'Univers, de l'Environnement et de l'Espace (SDU2E) de l'Université Paul Sabatier et il a été développé au laboratoire Centre d'Études Spatiales de la BIOsphère (CESBIO) de Toulouse, France.

Mots-clés: télédétection, couverture du sol, séries temporelles satellitaires, données multispectrales, occupation des sols sur grandes surfaces, classification automatique, Sentinel-2, Landsat, données manquantes, couverture nuageuse et de neige, interpolation temporelle, stratification, échantillonnage spatial, échantillonnage temporel non-linéaire, phénologie de la végétation, température

Résumé étendu

Introduction

L'observation de la Terre depuis l'espace a considérablement transformé notre vision de la planète. Depuis le lancement du premier satellite mondial Sputnik en 1957, plusieurs satellites d'observation de la Terre (EO) ont été envoyés dans l'espace, en changeant la façon dont les humains comprennent le monde et sa dynamique. En effet, les données recueillies par les satellites d'observation de la Terre sont fondamentales pour la compréhension des caractéristiques et des changements des terres, des océans et de l'atmosphère de la planète, ainsi que pour l'analyse et la prédiction des processus naturels et des changements anthropiques sur la Terre.

Les estimations de la couverture terrestre sont nécessaires pour de nombreuses applications dans des domaines tels que l'environnement, l'agriculture, la foresterie, les transports, la recherche, la réponse aux catastrophes, la planification et de l'éducation [Kleeschulte and Büttner, 2006]. En effet, l'état de la couverture terrestre et l'analyse de ses changements sont des facteurs clés pour comprendre les processus environnementaux à l'échelle mondiale [Srivastava et al., 2012]. De nombreux objectifs prioritaires, tels que l'évaluation de l'impact de l'activité humaine sur les terres (comme la déforestation), la réponse aux catastrophes, et une multitude d'activités dans les domaines de la durabilité, de l'élaboration des politiques et de la gestion et la surveillance des terres, dépendent en grande partie de la disponibilité des informations actualisées et précises sur la couverture terrestre [Pérez-Hoyos et al., 2012] [Verbung et al., 2011] [Wiens et al., 2009]. Fig.1 montre des images d'une région du Japon par GeoEye-1 avant et après avoir été frappée par un tremblement de terre et un tsunami en 2011.

Figure 1: Images satellite GeoEye/EyeQ du Japon, au nord de la zone de Sendai, avant et après le tsunami de 2011, à une résolution spatiale de 1.65m. L'imagerie haute résolution spatiale des zones sinistrées est déterminante pour les opérations de secours et la cartographie des infrastructures.

1. Les nouvelles missions satellitaires

Avec l'arrivée des missions satellitaires de nouvelle génération et les nouvelles politiques de données gratuites qui sont en cours d'établissement, il est un moment passionnant pour la cartographie de la couverture terrestre basée sur la télédétection par satellite. Les capteurs présentent des capacités améliorées en termes de résolution et des capacités de couverture, et une nouvelle richesse des données peut désormais être acquise. La nouvelle mise à disposition gratuite des données Landsat détenues par l'Enquête United States Geological (USGS) et la politique de données libre qui sera également appliquée à certaines des futures missions sont des contributions essentielles à la recherche sur la cartographie et la surveillance de la couverture des sols, alimentant ainsi le développement des avancées techniques pour les grands ensembles de données et l'analyse multitemporelle. Ces moments sont sans précédent en termes de possibilités, mais aussi largement défiants, car le développement de nouveaux algorithmes et méthodes devient nécessaire pour être en mesure de traiter cette nouvelle richesse d'information qui n'a pas été disponible au passé.

Sentinel-2 est l'une des missions d'observation de la Terre développés par l'ESA, dans le cadre du programme Copernicus de la Commission Européenne, pour une capacité d'observation de la Terre autonome, à plusieurs niveaux, et opérationnelle. La mission comprend deux satellites identiques, Sentinel-2A et Sentinel-2B. Le premier a été lancé le 23 Juin 2015, lors de la rédaction de ce travail, et le système sera complet lorsque le satellite Sentinel-2B sera lancé fin de 2016. La configuration des satellites jumeaux offrira une couverture mondiale systématique sans précédent de terres à partir de 56 $56°S$ to $84°N$, avec une haute fréquence de révision de 5 jours à l'équateur (2-3 jours pour latitudes moyennes), des résolutions spatiales allant de 10 à 60m, et 13 bandes spectrales allant de 0.44μm to 2.28μm [Fletcher, K., 2012].

En comparaison avec leurs précurseurs, la quantité de données qui sera mise à disposition est sans précédent: la configuration Sentinel-2 fournira des résolutions spatiales plus élevées que MODIS, SPOT ou Landsat, et un nombre plus élevé de bandes spectrales, avec un temps de revisite plus court ainsi qu'une fauchée plus large que Landsat ou SPOT. Les implications de ces progrès sont claires: avec la disponibilité des ensembles de données multitemporelles de grand volume, l'élaboration de méthodes de traitement qui sont en mesure de faire face à des données volumineuses, en termes de temps de calcul et de gestion de la mémoire, devient un besoin urgent.

Le dernier des satellites Landsat est Landsat 8 ou Landsat Data Continuity Mission (LDCM), lancé le 11 Février 2013, obtient une imagerie de la Terre à une résolution de 30m (15m pour la bande panchromatique, et à 100 m pour les bandes thermiques). Il remplace Landsat 5 dans son orbite, poursuivant ainsi la configuration de deux satellites Landsat (Landsat 5 et 7) qui a duré plus de 14 ans en raison de la longévité exceptionnelle de la mission Landsat 5. En outre, il enrichit les paramètres Landsat traditionnels avec l'ajout de trois nouvelles bandes spectrales, et assure la continuité de la base de données historique Landsat, récemment mise à disposition gratuitement. Illustrations des satellites Sentinel-2 et LDCM sont présentés dans Fig.2.

En effet, une nouvelle ère de données par satellite en libre accès a été initiée par l'ouverture des archives Landsat en 2008 par l'USGS, et sera poursuivie par la politique de données du programme Sentinel et d'autres missions telles que Venμs [Hagolle et al., 2010]. L'archive USGS contient près de trois millions d'images [Loveland and Dwyer, 2012], recueillies au fil des ans et qui remontent à

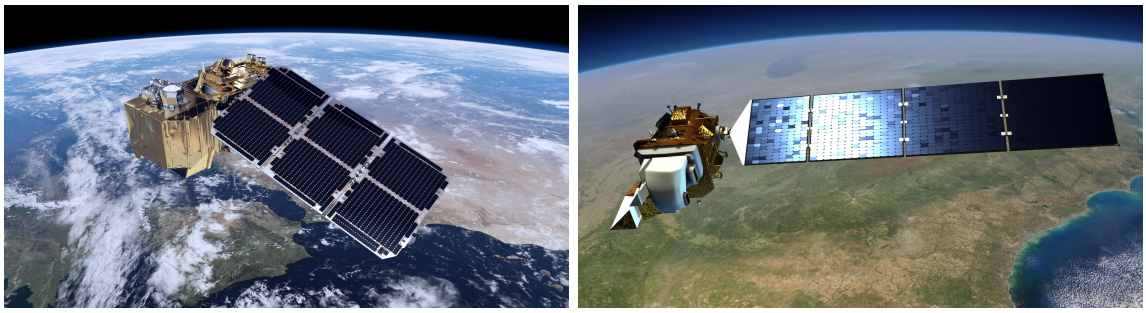


Figure 2: Illustration pour (a) un satellite Sentinel-2 (b) le récemment lancé LDCM (Landsat Data Continuity Mission ou Landsat 8) [NASA, 2014b].

1972 avec le premier des satellites Landsat. La nouvelle politique a des implications importantes: auparavant, les scientifiques travaillaient sur un nombre limité d'images quand des acquisitions Landsat étaient commercialisées, atteignant des milliers de dollars chacune, en raison de préoccupations budgétaires. Maintenant, avec l'accès à grandes séries temporelles des images multispectrales, et des volumes de données sans précédent, une révolution dans le domaine des techniques de traitement et d'analyse est nécessaire pour traiter efficacement la richesse d'information historique et actuelle qui va être mise à disposition.

2. Défis

La nouvelle richesse des informations fournies par les nouvelles missions satellitaires et les politiques de données libres expliquées dans la section précédente ouvrent la voie à une multitude de possibilités pour les recherches sur grandes surfaces et à long terme. Avec elles, un changement vers l'utilisation de séries temporelles multispectrales est rendu possible, ce qui permet à son tour l'analyse de cycle phénologique et l'identification des classes de végétation. Par rapport à la génération des cartes d'occupation des sols basée sur une ou quelques images, les séries temporelles sont des sources d'information de valeur pour la plus précise production de cartes de la couverture terrestre [Carrão et al., 2008]. Fig.3 montre quatre FORMOSAT-2 composites en fausses couleurs améliorant la végétation pour une région en Yaqui, au Mexique. La divergente évolution temporelle des différentes variétés de cultures, qui peut être exploitée pour la classification, peut être observée à partir de Décembre 2007 jusqu'à Février 2008.

Une grande variété de produits d'occupation du sol est requise dans une variété de domaines thématiques, comme décrit ci-dessus. Les informations sur la couverture terrestre peuvent varier de la surveillance de cultures et des forêts à la cartographie des zones de catastrophe pour les interventions d'urgence, comme un produit en lui-même, ou constituer un ensemble de données d'entrée pour la recherche dans autres domaines scientifiques, tels que la définition de modèles biophysiques. Dans les deux cas, les propriétés souhaitables comprennent la précision, des résultats de haute résolution, une production en court délai (de sorte que l'information puisse être rapidement mise à jour), et des capacités sur de grandes surfaces et à long terme.
Afin d'atteindre ces objectifs souhaitables, différents défis se posent:

1. D'un côté, les jeux de données multitemporelles de grand volume nécessitent le développement

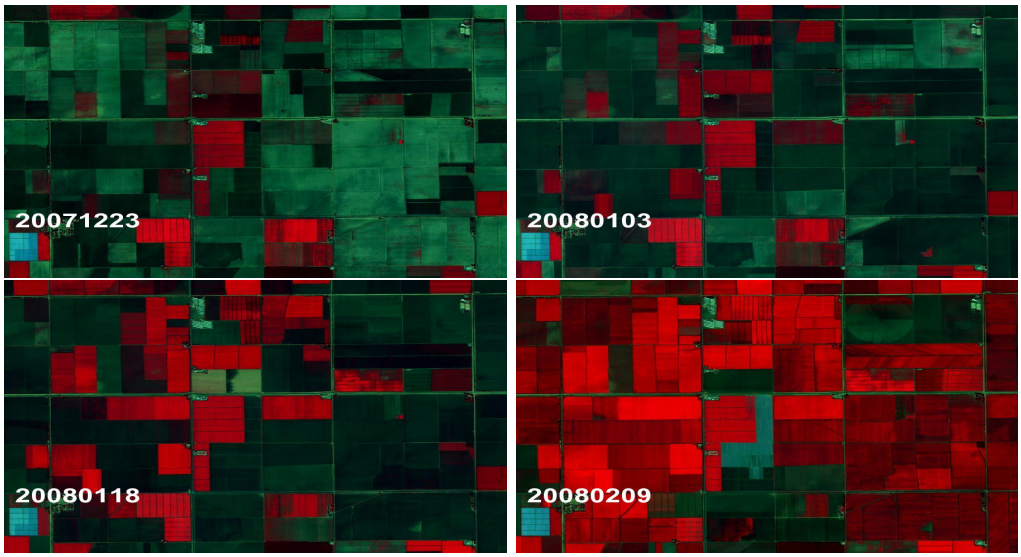


Figure 3: Séries temporelles FORMOSAT-2 de composites en fausses couleurs pour une région en Yaqui, au Mexique, de Décembre 2007 à Février 2008, avec une résolution spatiale de 8m dans le visible et le proche infrarouge.

de méthodologies de traitement qui sont en mesure de faire face à de grandes quantités de données, en termes de temps de calcul et de gestion de la mémoire. Des développements techniques sont nécessaires afin d'obtenir une méthodologie opérationnelle pour la production de cartes d'occupation du sol sur grandes surfaces.

2. D'un autre côté, l'automatisation est hautement souhaitable, car les méthodes manuelles sont coûteuses, chronophages et subjectives [Ippoliti et al., 2012]. La généralisation implique que l'automatisation requise doive être compatible avec les techniques mises en œuvre pour gérer de grands volumes de données, tant en termes de traitement opérationnel que de gestion de la mémoire.

3. En plus de cela, une grande couverture exige des mécanismes d'intégration de l'importante de variabilité spectrale de données par rapport à la diversité spatiale. En effet, les classes d'occupation du sol présentent une variabilité spectrale et temporelle due á la variation des sols et des caractéristiques à travers les territoires et, pour les classes de végétation, l'effet des différents cycles phénologiques de la végétation en raison de caractéristiques topographiques et climatiques. En conséquence, les approches traditionnelles pour la cartographie de la couverture terrestre peuvent échouer considérablement lorsqu'elles sont appliquées sûr très grandes surfaces, parce que la caractérisation des types d'occupation du sol doit tenir compte des événements déterminés par différentes conditions.

4. En outre, l'absence de données d'entrées idéales doit être adressée, avec une présence importante d'information manquante. Cela est généralement dû à l'existence de nuages, d'ombres de nuage, de neige, ou des bordures de passage du satellite dans la série temporelle, et elle peut avoir un impact significatif sur les tentatives de cartographie de l'occupation du sol si elle n'est pas gérée de façon adéquate.

5. Les irrégularités temporelles, également un facteur en l'absence de données d'entrée idéales,

résultent de l'existence de différentes dates d'acquisition pour diverses régions sur de grandes étendues, ce qui empêche l'application des techniques d'apprentissage automatique sur l'ensemble des régions cibles. Des séries temporelles complètes spatialement et temporellement sont bénéfiques pour une extraction de données fiable pour la génération de cartes d'occupation des sols.

6. Enfin, l'existence d'inexactitudes de vérité terrain, en raison de l'existence de différentes sources de données, méthodes de collecte, résolutions de sortie, taux de mise à jour et estimations de précision, doit être prise en compte. La raison de ceci réside dans le fait qu'il a un impact sur la qualité des produits dérivés, ainsi que sur les estimations de validation.

Compte tenu du contexte de ce travail, avec l'arrivée des missions satellitaires de nouvelle génération, et les défis et les opportunités que la disponibilité de leurs nouveaux ensembles de données posera, comme décrit ci-dessus, le besoin du développement de nouveaux algorithmes et des méthodes qui sont en mesure de traiter la nouvelle type de données avec les restrictions énoncées est de fait évident.

Conclusions

Ce document présente la conception et la validation d'une approche entièrement automatique qui permet l'estimation de la couverture terrestre de grandes surfaces avec imagerie optique d'observation de la Terre de haute résolution spatiale, spectrale et temporelle, étant applicable à différentes nomenclatures et séries temporelles satellitaires et offrant des temps de calculs opérationnels (avec livraison en quelques heures), en préparation des missions à venir Sentinel-2 et Venμ. Le système développé démontre la faisabilité de la production automatique de cartes d'occupation du sol à grande échelle, en gérant le volume de données, variabilité, données de référence limitées et informations manquantes, en intégrant différentes approches de traitement de données.

1. Contributions

Le système mis en place constitue une méthodologie entièrement automatique de traitement d'image basé sur les grandes étapes suivantes: traitement de l'information manquante des séries temporelles, rééchantillonnage dans l'axe temporel, construction des vecteurs caractéristiques, échantillonnage dans l'axe spatial, apprentissage des machines à vecteurs de support, et classification. En pratique, il permet la génération rapide et automatique des cartes d'occupation du sol à grande échelle en utilisant de volumineuses séries temporelles multispectrales de haute résolution. Il offre une solution aux problèmes posés par la production d'occupation du sol à grande échelle avec la dernière génération de satellites: la couverture par différentes séries temporelles avec des dates d'acquisition non-coïncidentes (due à l'imagerie par différents passages des satellites), la présence importante d'informations manquantes à cause de diverses raisons (nuages, ombres des nuages, frontières de passage, neige), les coûts de traitement élevés pour des grands volumes d'information (séries temporelles de données multispectrales à haute résolution), une forte variabilité spectrale des classes par rapport à la diversité spatiale, et l'existence d'un ensemble des données de référence limité.

En particulier, cinq principales contributions ont été développées dans les chapitres de ce travail:

(1) En premier lieu, le cadre pour la génération automatique des cartes d'occupation du sol sur grandes surfaces avec séries temporelles de grand volume telles que celles qui doivent être fournies par la mission satellite à venir Sentinel-2. Il donne une réponse au problème de l'automatisation dans le cadre considéré. Il se compose d'une approche par pixel fondant la construction de vecteurs caractéristiques sur des dates virtuelles après un rééchantillonnage temporel sur la série temporelle d'entrée et il permet la production de résultats en temps opérationnels ainsi que la gestion de grands volumes d'informations. Sur la base de l'exploitation des informations multitemporelles et multispectrales, il utilise une combinaison de l'extraction ciblée des échantillons, le traitement de l'information manquante, l'extraction de caractéristiques caractérisantes, l'échantillonnage, l'apprentissage supervisé des modèles, et enfin la classification basée sur les modèles appris. Ce cadre permet l'application combinée des nouvelles contributions présentées dans ce travail par rapport aux méthodes de classification classiques mais aussi de l'automatisation ciblée ne nécessitant pas de supervision d'un opérateur.

(2) En deuxième lieu, la comparaison des approches du traitement de l'information manquante. Il donne une réponse au problème de l'information manquante dans le cadre considéré. Un méthode multi-temporelle pour produire des séries temporelles haute résolution temporellement et spatialement continues pour le traitement à grande échelle est présentée, basée sur l'application de techniques d'interpolations locales des séries temporelles. Différentes approches à la méthode proposée sont mises en œuvre et comparés, et leur adéquation est prouvé, particulièrement en étant possible d'obtenir des expressions analytiques successives dans les fenêtres de rayon variable pour les séries temporelles complètes.

(3) En troisième lieu, la méthode d'échantillonnage temporelle. Elle donne une réponse au problème des points temporels d'acquisition inégalés et des irrégularités temporelles avec des ensembles de données couvrant des grandes surfaces dans le cadre considéré. La méthode profite de l'approche précédente de traitement de l'information manquante afin de générer des points de données virtuels pour la construction du vecteur de caractéristiques. Il permet l'apprentissage et la classification sur grandes surfaces avec un seul modèle appris, en substituant la nécessité des différents apprenants et classificateurs, et en évitant d'autres solutions telles que la sélection manuelle de zone (trop coûteuse en termes de temps) et le rejet d'images avec une faible teneur (perte d'informations valides).

(4) En quatrième lieu, l'approche stratifiée d'échantillonnage spatial. Elle donne une réponse au problème de la grande variabilité des données en termes de réponse spectrale lorsque de grandes zones sont considérées dans le cadre donné. Elle est basée sur l'application des stratégies automatiques d'échantillonnage spatial pour la production d'ensembles d'apprentissage représentatifs, permettant ainsi un apprentissage automatique plus rapide pour une classification ultérieure. Ces stratégies complètent les stratégies d'échantillonnage temporelles décrites ci-dessus, en offrant une méthode qui fonctionne de cette façon dans l'espace pour la sélection du jeu de données, et dans le temps pour le traitement des données manquantes et la régularisation. Différentes approches pour l'échantillonnage spatial sont comparées pour la zone d'étude choisie, et en particulier la stratification qui est examinée par incorporation de concepts de sciences de l'environnement dans le but d'améliorer la performance de l'apprentissage automatique. Des informations topographiques et écoclimatiques sont appliqués pour l'identification des variables utiles de stratification.

(5) En cinquième lieu, la méthode d'échantillonnage temporelle non linéaire basée sur la température. Elle donne une réponse à la limitation de la linéarité lorsqu'il s'agit de l'extraction de paramètres de végétation. Elle se fonde sur les approches précédentes avec l'application des techniques d'interpolation (pour le traitement de l'information manquante), le rééchantillonnage temporel avec des points virtuels (pour la gestion des dates d'acquisition irrégulières et de gros volumes de données), et l'échantillonnage spatial stratifié (pour le traitement des la variabilité des données et des coûts de traitement). Un échantillonnage temporel non linéaire basée sur la température est proposé pour les classes de végétation, avec l'hypothèse que l'évolution phénologique et la réponse spectrale dans le temps de de la végétation est dépendante de la température plutôt que du temps. Un procédé pour l'analyse de la réponse spectrale dans l'espace de la température est donnée, et des approches différentes ont été développées et testées. En outre, des méthodes de correction pour le raffinement des informations sur la température d'entrée sont mises en œuvre et présentées.

Le système a été développé comme une méthodologie de traitement pour les ensembles de données Sentinel-2, mais établit un cadre général applicable à d'autres données satellitaires ou ensembles de données multicapteur dûment corregistrés et rééchantillonnés. La méthode est flexible pour travailler avec tous types de séries temporelles satellitaires raster et il est leur disponibilité, qualité et caractéristiques (spatiales, temporelles et spectrales) qui détermine les faisabilités d'extraction et les niveaux atteignables de détail. En outre, différentes classes d'informations peuvent être appliquées, selon le but de la carte d'occupation des sols; tout nombre de classes et d'informations d'occupation du sol peuvent être accueillis.

Le système est reproductible et robuste, car il peut être facilement reproduit et il peut faire face à des erreurs et artefacts dans les données d'entrée. En effet, la reproductibilité est liée à la minimisation du rôle de l'analyste dans la classification; les entrées de l'opérateur tout au long d'une procédure entraînent non-reproductibilité [Cihlar, 2000]. Il a de faibles besoins en termes de données d'entraînement, mémoire et temps de traitement. Il est totalement automatisé et ne nécessite pas d'interaction avec un utilisateur ou de travail manuel. Il restreint l'entrée de l'opérateur aux décisions initiales subordonnées aux données disponibles et à la sortie désirée, comme par exemple les catégories d'occupation du sol ciblées, le type de caractéristiques à utiliser, ou les décisions sur les paramètres des procédés décrits ci-dessus, tels que le nombre de points virtuels pour le rééchantillonnage temporel, le nombre d'échantillons d'entraînement à utiliser, ou la durée de la période temporelle à étudier. De cette façon, la reproductibilité et la cohérence de la méthode peuvent être assurés, et l'impact de la subjectivité lié à l'opérateur et les coûts associés sont réduits au minimum.

En conclusion, ce travail tente de fournir à la communauté de la télédétection, et aux utilisateurs qui ne sont pas des spécialistes du domaine, mais qui comptent sur les sorties de cartographie d'occupation du sol en tant qu'entrées pour leurs propres domaines de poursuite, avec une méthodologie pour faciliter l'exploitation de l'information des capteurs de nouvelle génération. Ce but, en permettant la production des cartes de grande superficie basées sur séries temporelles de haute résolution avec des délais de livraison opérationnels, et en particulier en étant capable de traiter des données réelles, qui ont généralement des irrégularités temporelles, des informations manquantes, et une grande variabilité des données. Les stratégies proposées essaient d'ouvrir de

nouvelles voies pour le traitement des ensembles de données de la nouvelle génération de satellites.

2. Remarques finales

L'importance d'une cartographie rapide et précise et d'une surveillance de l'occupation des sols devient évidente lorsque l'on considère la croissante souche sur les ressources de la Terre: la pression anthropique, le réchauffement climatique et le changement global ont un effet indéniable sur les processus naturels; les changements climatiques et leurs effets sur les écosystèmes sont sans précédent au cours des dernières décennies et vont s'intensifier à l'avenir; la déforestation, la désertification et l'érosion des sols menacent la sécurité alimentaire. Dans ces circonstances, la recherche sur la cartographie d'occupation du sol avec des images de télédétection de la Terre par satellite devient essentielle, permettant l'extraction des informations clés pour la cartographie des ressources passées et actuelles qui sont en constante évolution et donc la compréhension des processus de la Terre, afin de prévoir des scénarios futurs, en contribuant de cette façon à la réduction des pertes dues aux catastrophes et au développement durable.

Dans ce contexte, ce travail est conçu comme une contribution à la cartographie de l'occupation du sol rapide mais est aussi capable de gérer de grandes surfaces avec des images de télédétection. Il constitue une méthodologie pour l'estimation de l'occupation des sols de grandes surfaces avec des images satellitaires à haute résolution et des temps de calcul opérationnels (livraison en quelques heures). Il est spécifiquement conçu pour gérer grands ensembles de données avec volumes de centaines de millions de pixels (tels que ceux qui doivent être fournis par les nouvelles missions satellitaires). Il a de faibles coûts de traitement et de mémoire (il peut être mis en œuvre dans tout courant PC) et il est entièrement automatique, n'appliquant pas supervision ou interaction d'un opérateur afin d'éliminer la subjectivité. Il admet la variabilité des classes d'occupation du sol dans l'espace, et gère les informations manquantes. En outre, l'une de ses caractéristiques principales est la généralité: il peut être réglé en fonction des besoins, et généralisé à différentes nomenclatures (en fonction du but de la carte d'occupation du sol), des séries temporelles de données satellitaires (en tenant compte de leurs différentes caractéristiques spatiales, temporelles et spectrales) et des paysages de toute taille. En conclusion, ce travail a été conçu comme une contribution non seulement au domaine de la cartographie d'occupation du sol, mais aussi comme un outil pour les utilisateurs dans d'autres domaines de connaissances qui reposent sur les résultats de la cartographie d'occupation du sol en tant qu'entrée pour leurs propres activités de recherche.

Chapter 1

Introduction

Contents

1.1 General context

The world population is in accelerated augmentation. According to the 2012 Revision of population estimates and projections of the United Nations Secretariat [UNPD, 2013], the Earth population (7.2 billion in 2013) is projected to reach 8.1 billion in 2025 and 9.6 billion in 2050. Fig.1.1 illustrates the demographic growth from 10000 BCE. While it took over 2 million years to reach 5 million inhabitants, only 40 years will be needed to attain such growth at the current growth rate [Nicholas, 2004]. Such an anthropic pressure on the planet poses many risks, notably the fact that population growth needs to be paired to that of resources. Some of these risks include the extension of food and water shortages, the overexploitation of lands, the pollution of waters and seas, and the growing urbanisation in detriment of arable lands [Nicholas, 2004].

According to the Intergovernmental Panel on Climate Change (IPCC) report, the warming of the Earth is unequivocal, and climate changes are unprecedented over previous decades to millennia, globally reducing the amounts of snow and ice and provoking a rise of the sea levels [IPCC WG, 2013]. Fig.1.2 illustrates the significant increase of combined ocean and land surface temperature for the period 1850-2012 and relative to 1961-1990, including an annual and a decadal average.

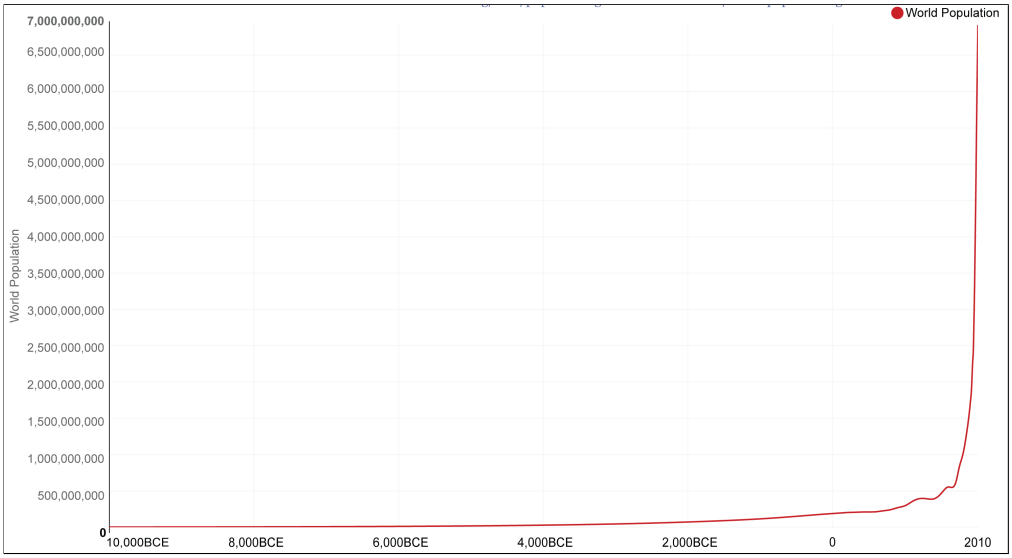


Figure 1.1: Annual world population growth from 10000 BCE and its expected development based on the United Nations Projection (2012 Revision) [Roser, 2015].

The concentration of greenhouse gases has increased to levels unprecedented in at least the 800000 years, and the planet has heated an average of 0.85°C between 1880 and 2012 [IPCC WG, 2013]. Such tendency, if continued, could have disastrous and irreversible effects. For example, since 1901, the level of the seas has augmented 19cm on average, with increasing speed. Since 1993, the rise rate is 3.2mm per year, nearly twice as fast as the rate at the beginning of the 20th century.

With the warming of the atmosphere and the ocean, severe climatic changes are expected: droughts and extreme precipitation events will become more intense and more frequent and precipitation contrasts between dry and wet regions and seasons will increase [IPCC WG, 2013]. Also, climatic warming has an impact on fauna and flora at a global level. It is expected that ecosystems will be fragilised, and that species will implant in new regions, which can have an impact on human beings and their crops. In addition to this, the rate of degradation of resources is currently intensifying, with deforestation, soil erosion and desertification threatening food security. It is estimated that millions of people will have to migrate in the future due to climate change consequences, with catastrophes such as floods, droughts, and storms, and difficulties to find nourishment due to a reduction of arable land or to arid climate extension.

Anthropic pressure, climatic warming and global change are increasingly putting the planet to a strain, as described. In these circumstances, a sustainable resources development plan is needed, based on the mapping of current resources and the forecasting of future scenarios. In this sense, the importance of satellite remote sensing of the Earth cannot be underestimated: satellite observations of land, atmosphere and ocean provide key information that can be applied for the mapping, analysis and forecasting of the Earth's processes, which in turn can reduce disaster losses and contribute to sustainable development.

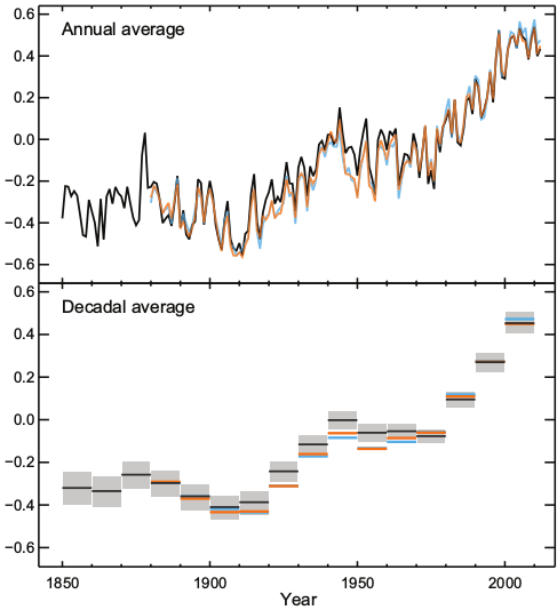


Figure 1.2: Observed surface temperature anomaly for the period 1850-2012 (land and ocean global average) relative to 1961-1990 [IPCC WG, 2013].

1.1.1 Remote sensing for Earth Observation

Observing the Earth from space has significantly transformed our view of the planet. Since the launch of the world's first satellite Sputnik in 1957, multiple Earth Observation (EO) satellites have been sent to space, changing the way humans understand the world and its dynamics. Indeed, data gathered by EO satellites is fundamental for the understanding of the features and changes of the world's lands, oceans and atmosphere, and also for the analysis and prediction of natural processes and anthropic changes on the Earth.

The well-known "Blue Marble" photograph shown in Fig.1.3 was taken by Apollo 17 astronauts en route to the moon in 1972, and it was revolutionary at that time. Since then, techniques and measurements have greatly improved in precision and sophistication, and from the early photographs, the evolution towards quantitative measurements obtained from space-born remote sensing platforms has lead to great advances in the Earth and climate sciences. For example, the mass balance of the ice sheets of Antarctica and Greenland was believed to be in equilibrium before satellite imaging, with a constant rate of ice discharge into the ocean. Later, radar satellite acquisitions showed that the flow is highly variable, and that in fact it increased with climate change [NRC, 2008].

Other examples of climate and Earth science advancements since the arrival of satellite-born remote sensors include significant improvements in weather forecasting and disaster warning that result in the saving of lives and have a great economic impact. Land cover changes monitoring and forecasting of crop productivity was demonstrated as early as the early 1970s [NRC, 2008]; since then, Landsat multispectral satellite imagery has been routinely used for crop commodity forecasting [NRC, 2008]. For example, the Famine Early Warning System Network, provid-

Figure 1.3: The Blue Marble: image of the Earth as seen by the crew of the Apollo 17 on December 7 1972 [NASA, 2014].

ing information on food insecurity and crises, applies satellite data in conjunction with other types of information, such as prices and trade, in order to predict and mitigate famines [NRC, 2008].

EO is also an example of cross-collaboration between technology and science [NRC, 2008], where the findings of scientists studying Earth satellite observations define new needs regarding technological developments and data interpretation methods, thus pushing the accomplishments in these areas in order to obtain more accurate and relevant quantitative measurements and data mining methods. Also, it offers a ground for multidisciplinary research, where Earth science, engineering, environmental studies, climatology and other sciences can successfully interact for a cross-disciplinary analysis and understanding of the processes of the Earth.

In the current context described above, with an ever-increasing demographic pressure and an intensifying resource degradation, the role of remote sensing for EO becomes even more determinant. Indeed, crucial tasks such as weather, climate, and hazard forecasting depend critically on satellite-based observations. Findings related to EO can help society to manage environmental issues and the problem of limited resources. Without any doubt, continued observation of the Earth through satellite-born remote sensing will have a key role in the addressing of the societal challenges of the future.

1.1.2 The need for land cover mapping

Land cover estimates are necessary for many applications in areas such as environment, agriculture, forestry, transport, research, disaster response, planning and education [Kleeschulte and Büttner, 2006]. Indeed, land cover status and change analysis is a key factor for understanding environmental processes at global scales [Srivastava et al., 2012]. Many priority objectives, such as impact assessment of human activity on land (such as deforestation), disaster response, and a wealth of activities in the domains of sustainability, policy making and land management and monitoring, depend largely on the availability of updated and precise land cover information [Pérez-Hoyos et al., 2012] [Verbung et al., 2011] [Wiens et al., 2009]. Fig.1.4 shows imaging of Japan by GeoEye-1 before and after the region was hit by an earthquake and a tsunami in 2011.

Figure 1.4: GeoEye/EyeQ Satellite images of Japan, north of Sendai area, before and after the tsunami of 2011, at a spatial resolution of 1.65m. High spatial resolution imagery of disaster affected areas is determinant for relief operations and infrastructure mapping.

Indeed, the Global Climate Observing System (GCOS) Secretariat in the United Nations Framework Convention on Climate Change (UNFCC) identified land cover as one of the highest priority Essential Climate Variables for the monitoring of global change [Giri et al., 2013]. In 2010, updated guidance for priority actions was issued [GCOS, 2010] for the understanding and prediction of climate change and its impacts, as well as response management for the future [Bojinski and Richter, 2010], thus recognizing a pressing need for the generation of updated large area land cover information.

The Group on Earth Observation (GEO), with its attempt to build a Global Earth Observation System of Systems (GEOSS) based on a ten-year implementation plan from 2005 to 2015, highlight the need for operational and global land cover operations [Wulder et al., 2008]. The plan is endorsed by nearly 60 governments and the European Commission, and constitutes a step towards addressing the urgent need for coordinated observation of the state of the Earth. In particular, GEO has defined nine areas of societal benefits derived from Earth Observation, as listed below [GEO, 2005]:

- Biodiversity: understanding, monitoring and conserving biodiversity
- Agriculture: supporting sustainable agriculture and combating desertification
- Ecosystems: improving the management and protection of terrestrial, coastal and marine ecosystems
- Weather: improving weather information, forecasting and warming
- Water: improving water resource management through better understanding of the water cycle
- Climate: understanding, assessing, predicting, mitigating and adapting to climate variability and change
- Energy: improving management of energy resources
- Health: understanding environmental factors affecting human health and well-being
- Disasters: reducing loss of life and property from natural and human-induced disasters

In all of these areas, land cover mapping plays an important role, with a variety of desired products [Wulder et al., 2008], contributing benefits to society. In the biodiversity area, for example, land cover observation is of great value for the analysis of changes affecting biodiversity and for ecosystem characterization; in agriculture, for the monitoring of the distribution of crop production types and of forest changes and deforestation; in ecosystems, for the study of land cover and vegetation changes. Fig.1.5 shows vegetation maps derived from the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) for a region in Rondonia, Brazil, where severe deforestation occurred in the period covered by the images, from 2000 to 2006.

Figure 1.5: ASTER-derived vegetation index images of Brazil, Rondonia region, showing the intensification of deforestation between 2000 and 2006, with a spatial resolution of 15m in the visible and the near-infrared. Brighter areas correspond to deforested regions. [NASA, 2014b]

Similarly, in the disasters area some useful products include the production of fire extent maps and the location of land cover changes due to disasters, as well as population and infrastructure map generation. In the energy and climate areas, some important Earth Observation products include the computation of biomass yield estimates from agriculture and forestry, and of greenhouse gas emissions caused by land cover change, for example.

1.2 Specific context

With the advent of the new generation satellite missions and the new free data policies that are recently being established, it is an exciting time for the mapping of land cover using satellite remote sensing. Sensors exhibit enhanced capabilities in terms of resolution and coverage capabilities, and a new richness of data can now be acquired. The new free availability of Landsat data held by the United States Geological Survey (USGS) and the free data policy that will also be applied to some of the future missions are key contributions to research on land cover mapping and monitoring, fuelling the development of technical advancements for large datasets and multitemporal analysis. These times are unprecedented in terms of the opportunities laying ahead, but also greatly challenging, as the development of new algorithms and methodologies is needed to be able to process a new richness of information that was not available in the past.

1.2.1 The new satellite missions: Sentinel-2

Sentinel-2 is one of the Earth Observation missions developed by ESA, in the frame of the Copernicus Programme of the European Commission, for an autonomous, multilevel and operational Earth Observation capacity. The mission comprises two identical satellites, Sentinel-2A and Sentinel-2B. The former was launched on 23 June 2015, during the writing of this thesis, and the system will be complete when the Sentinel-2B satellite is launched at the end of 2016. The configuration of twin satellites will offer an unprecedented systematic global coverage of land from $56°S$ to $84°N$, with a high revision frequency of 5 days at the equator (2-3 days at mid-latitudes)), spatial resolutions ranging from 10 to 60m, and 13 spectral bands ranging from 0.44μm to 2.28μm [Fletcher, K., 2012].

Venμs, jointly developed by the French Space Agency (*Centre National d'Études Spatiales*, CNES) and the Israeli Space Agency (ISA), will deliver images in 12 spectral bands from 412 to 910 nm, a repetitivity of 2 days, and a spatial resolution of 10 m. In comparison with their precursors, the amount of data that will be made available by the new satellite missions is unprecedented: the Sentinel-2 configuration will provide higher spatial resolutions than MODIS, SPOT or Landsat, and a higher number of spectral bands, with a shorter revisit time and a wider swath than Landsat or SPOT. The implications of such advances are clear: with the availability of large volume multitemporal datasets, the development of processing methodologies that are able to cope with voluminous data, in terms of computation times and memory management, becomes an urgent need.

The latest of the Landsat satellites is Landsat 8 or Landsat Data Continuity Mission (LDCM), already launched on 11 February 2013, and imaging the Earth at a resolution of 30m (15m for a panchromatic band, and 100m for the thermal bands). It replaces Landsat 5 in its orbit, thus continuing the configuration of two Landsat satellites (Landsat 5 and 7) that stretched over 14 years due to the exceptional longevity of the Landsat 5 mission. Also, it enriches the traditional Landsat settings with the addition of three new spectral bands, and ensures the continuity of the historical Landsat database, recently made freely available. Two illustrations of the satellites Sentinel-2 and LDCM are given in Fig.1.6

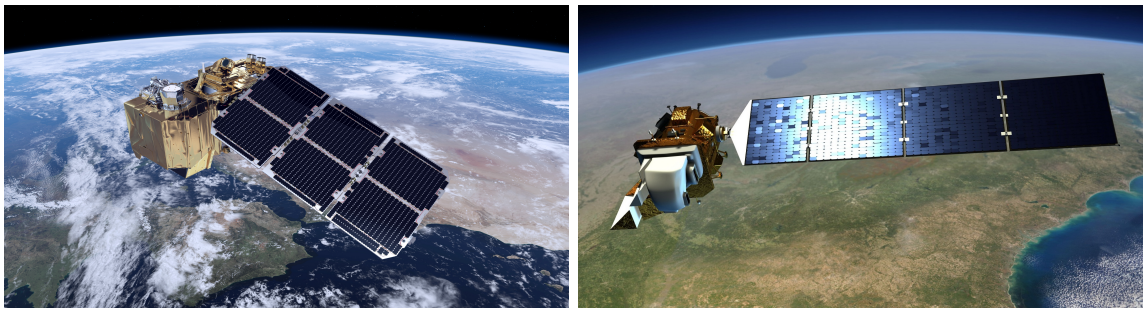


Figure 1.6: Illustration for (a) a Sentinel-2 satellite (b) the recently launched LDCM (Landsat Data Continuity Mission or Landsat 8) [NASA, 2014b]

Indeed, a new era of open-access satellite data has been initiated by the opening of the Landsat archive in 2008 by the USGS, and is to be continued by the data policy of the Sentinel programme

and other missions such as Venμs [Hagolle et al., 2010]. The USGS archive contains nearly three million images [Loveland and Dwyer, 2012], collected over the years and dating back to 1972 with the first of the Landsat satellites. The new policy has important implications: before, scientists worked on limited numbers of images when Landsat acquisitions were commercialised, reaching thousands of dollars each, due to budget concerns. Now, with access to large multispectral time series, and unprecedented mass volumes of data, a revolution in the field of processing and analysis techniques is required to effectively process the historical and current-day wealth of information that is made available.

1.2.2 Challenges

The new richness of information provided by the new satellite missions and the free data policies explained in the previous section paves the way for a wealth of opportunities for both large-area and long-term investigations. With them, a shift towards the use of multispectral time series is made possible, which in turn allows for phenological cycle analysis and identification of vegetation classes. In comparison to land cover map generation based on one or few images, time series are valuable sources of information for the production of more accurate land cover maps [Carrão et al., 2008]. Fig.1.7 shows four FORMOSAT-2 false colour composites enhancing vegetation for a region in Yaqui, Mexico. The diverging temporal evolution of different crop varieties, which can be exploited for classification, can be observed from December 2007 to February 2008.

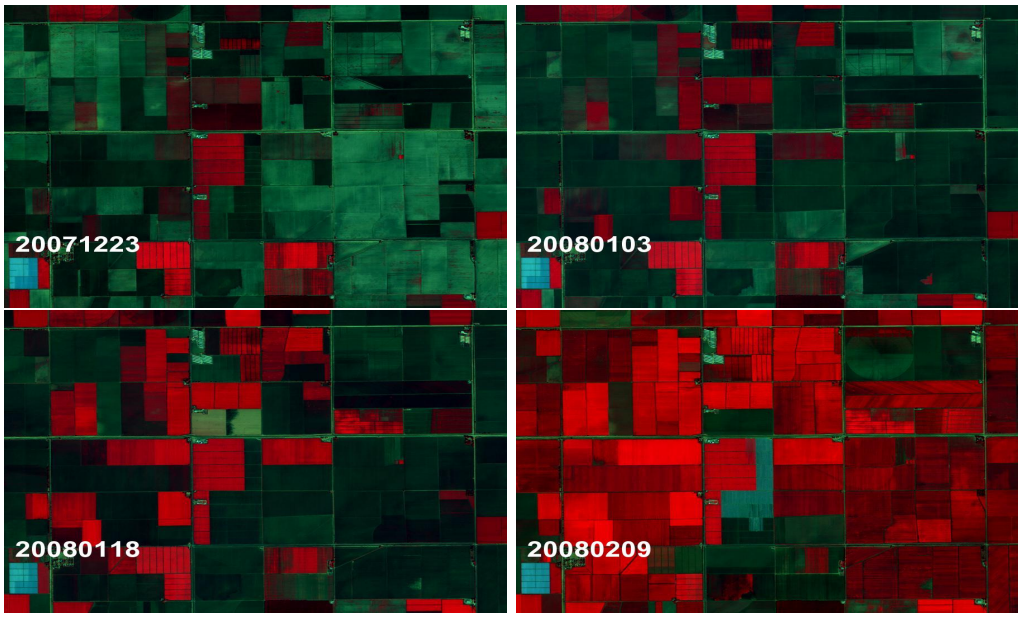


Figure 1.7: FORMOSAT-2 time series of false colour composites for a region in Yaqui, Mexico, from December 2007 to February 2008, with a spatial resolution of 8m in the visible and the near-infrared.

A great variety of land cover products is needed in a variety of thematic areas, as described in Sec.1.1.2. Land cover information can range from crop and forest monitoring to disaster area mapping for emergency response, as a product in itself, or constitute a input dataset for research on other scientific fields, such as the definition of biophysical models. In either case, desirable

properties include accuracy, high resolution results, a timely production (so that information can be updated rapidly), and large-area and long-term capabilities.

In order to achieve this desirable goals, different challenges are posed:

1. On the one hand, large volume multitemporal datasets require the development of processing methodologies that are able to cope with great volumes of data, in terms of computation times and memory management. Technical developments are needed in order to obtain an operational methodology for the generation of large-area land cover maps.

2. On the other hand, automation is highly desirable, as manual methods are costly, time-consuming and subjective [Ippoliti et al., 2012]. The generalisation that automation requires needs to be compatible with the techniques implemented to manage great volumes of data, both in terms of operational processing and of memory management.

3. In addition to this, large area coverage requires mechanisms for integrating the important spectral variability of data in relation to spatial diversity. Indeed, land cover classes present spectral and temporal variability due to soil and characteristics variation across territories and, for vegetation classes, the effect of differing vegetation phenological cycles due to topographic and climatic characteristics. As a consequence, traditional approaches for land cover mapping can significantly fail when applied to very large areas, as the characterisation of land cover types needs to take into account occurrences determined by different conditions.

4. Also, an absence of ideal input data needs to be addressed, with an important presence of missing information. This is generally due to the existence of clouds, cloud shadows, snow, or satellite passage borders in the time series, and it can have a significant impact on land cover mapping attempts if not adequately managed.

5. Temporal irregularities, also a factor in the absence of ideal input data, result from the existence of different acquisition dates for different regions over large areas, preventing the application of machine learning techniques on whole target regions. Spatially and temporally complete time series are beneficial for a reliable data extraction for land cover map generation.

6. Finally, the existence of ground truth inaccuracies, due to the existence of different data sources, collection methods, output resolutions, update rates and accuracy estimates, needs to be taken into account. The reason for this is the fact that it has an impact on the quality of derived products, and on validation estimations. This reference data problem is further discussed in Sec.2.4. In spite of being problematic, this last challenge remains outside the scope of this work.

1.3 Objectives of this thesis

Given the context of this work, with the arrival of the new generation satellite missions, and the challenges and opportunities that the availability of their new datasets will arise, as described in sections 1.1 and 1.2, the development of new algorithms and methodologies that are able to process the new type of data with the stated requirements is made obvious.

In this framework, the objective of this thesis is the conception, design, implementation and validation of a methodology that allows for the estimation of large-area land cover with high spatial, spectral and temporal resolution Earth observation imagery, being generalisable to different landscapes, and offering operational computation times (with delivery in a matter of hours) with satellite data sets in preparation of the coming Sentinel-2 mission. It is intended as a contribution not only to research in the domain of land cover mapping, but also as an orientation for users who are not specialists of the field but that rely on land cover mapping outputs as inputs for their own pursuit areas.

Furthermore, this work tries to contribute with the analysis of strategies in both the temporal and the spatial dimensions for the development of operational large-area land cover mapping algorithms to be applied on high resolution time series. It aims at the management of large volume datasets of hundreds of millions of pixels, the assurance of the consistency of data, and the admittance of variability of land cover classes over space, which are desirable properties [Aitkenhead and Aalders, 2011], [Wulder et al., 2003], as well as the management of missing information.

More particularly, this thesis provides solutions to the challenges described in Sec.1.2.2. It presents an automatic classification methodology for land cover map generation over large areas, based on the application of supervised learning with support vector machines and not requiring operator supervision or interaction. It makes use of multitemporal multispectral time series and ancillary topographical and climatic information and demonstrates the feasibility of operator-independent large-scale land cover mapping with operational processing times. To make this methodology possible, three novel contributions in relation to classical classification approaches have been developed:

1. A multitemporal approach to produce temporally and spatially continuous high resolution time series for large-scale processing. It is based on the joint application of local interpolation techniques in the time series and of temporal resampling strategies with virtual timepoints. It overcomes the problem of missing information in the time series due to cloud, cloud shadows, snow and satellite passage borders, as well as that of temporal irregularities due to different acquisition dates for different regions over large areas. It allows for the production of continuous time series that can be subsequently processed by standard processing algorithms and for the application of single machine learning algorithms on whole target areas.

2. An automatic training data extraction approach for the targeted selection of pertinent training samples aiming at the enhancement of classifier performance and the reduction of data volume on which to base classifier learning. It consists of the application of stratification techniques for training sample selection based on strata obtained from ancillary information. It overcomes the problems of large data volumes and great variability inherent to large areas by using additional information to guide the selection of samples. It allows for a faster classification model learning, as well as for an enhanced characterisation of land cover classes for classification.

3. A non-linear temperature-driven sampling approach aiming at the extraction of robust features for an improved characterisation of land cover classes and an improved performance of classification algorithms. It is based on the reduction of the variability of spectral responses

by applying an axis transposition along cumulative temperature curves, thus adding a virtual dimension determined by temperature. It allows for a more accurate characterisation of classes that can be exploited for land cover map generation in extended regions.

In conclusion, this thesis tries to provide the remote sensing community with a methodology to ease the exploitation of information from the new generation sensors, allowing for the production of large area land cover maps based on multitemporal high resolution time series with operational delivery times, and particularly dealing with real datasets, which generally present temporal irregularities and missing information. It intends to be robust, reproducible, and with low requirements in terms of training data, memory needs, user supervision and processing times; the suggested strategies try to open new paths for the processing of the new generation satellite datasets.

1.4 Organisation

This thesis is organised into seven chapters, each of which are described in this section. The general objectives of each chapter are outlined, and the logical connection between them is given.

Chapter 1 introduces the general context of the thesis and the need for land cover mapping, as well as the specific context and the challenges and opportunities posed by the new coming generation of satellite missions. In this framework, the objectives of the thesis and its organisation are described.

Chapter 2 presents the chosen study area and the datasets used to develop the proposed methodology, comprising satellite data time series, learning and validation data, and exogenous information on topography, pedology and meteorological data. Dataset sources and preprocessing tasks are described; also, data-related inaccuracies are given, and the difficulties that these error sources pose for land cover map generation and its validation are analysed.

Chapter 3 describes the general methodological approach that has been developed for the automatic generation of land cover maps at large scales. It pays special attention to the automation necessity that arises when large area map generation is coupled with the great volumes of information that will be provided by the coming Earth Observation satellite missions. The problem of automatic classification is described, an account of the state of the art is given, and the different stages of the proposed methodology are described and discussed. Tests yielding classification accuracies in the subsequent chapters of this work follow the methodological approach here described.

Chapters 4 to 6 present the same structure in five sections for clarity purposes: description of the problem to be solved, state of the art, description of the proposed method, presentation and analysis of undertaken experiments and obtained results, and final discussion and conclusions.

Chapter 4 introduces the problem of missing information in time series for large area map generation. A state of the art for missing information detection and processing is given, and a multitemporal method to produce temporally and spatially continuous HR time series for large-scale processing is presented. A temporal local interpolation and resampling method is

proposed, and different interpolation approaches are compared. Undertaken experiments and results are described and analysed in order to yield final conclusions.

Chapter 5 explains the problems of temporal irregularity, data variability and great volumes of data that are commonly encountered when large area land cover analysis with satellite data time series is considered. It provides an overview of the state of the art for the processing of such issues, and a method for the generation of large-area land cover maps with operational processing times is described, based on automatic spatial sampling techniques. Different spatial sampling methodologies are compared for the purpose, and in particular stratification for strata derived from topographic and ecoclimatic information. Experiments and results are given and analysed, and conclusions are drawn.

Chapter 6 presents a non-linear temperature-based temporal sampling method for large area processing. The problem of using temporal sampling based on linear time when vegetation parameters are considered is explained, and the state of the art is given. The proposed method is based on the application of integrals of temperature information for feature extraction over vegetation areas, with the hypothesis that phenological evolution and temporal spectral response of vegetation land cover is dependent on temperature rather than time. This information is obtained from SAFRAN meteorological data, and two correction methods for the refinement of input temperature information based on height and radiation variables are implemented and presented. A new non-linear class characterization in the temperature space is thus obtained, different from the linear temporal characterization given in Chapter 4, and having an impact on classification results. Then, experiments, results and conclusions are given.

Chapter 7 provides final conclusions for the undertaken work and describes the contributions of this thesis. Also, perspectives for future research are given.

Chapter 2

Study area and datasets

Contents

This chapter introduces the study area and datasets used in this work, including satellite imagery time series and exogenous information such as topography and meteorological data, of which the proposed methodology makes use. Similitudes and differences between the targeted Sentinel-2 data and the Landsat time series that have been applied for simulation are explained. Training and validation data sources and general nomenclature are also described. Dataset collection and preprocessing methods are given and discussed, and inherent data-related difficulties are explained. By providing information on the spatial, temporal and spectral characteristics of the input data, the quality of the reference data, type of nomenclature, and preprocessing methods, this chapter provides the frame for the interpretation of results presented in the following chapters.

2.1 Study area

The chosen site comprises an area of $84000 km^2$ in the Pyrenees region in southwest Europe. It includes the Pyrenees mountain chain, a natural border between France and Spain extending from the Gulf of Gascogne (Cap Higuer) to the Mediterranean Sea (Cap de Creus), and also a considerable extent of its bordering regions, mainly in the French territory, as shown in Fig.2.1.

The Pyrenees mountain chain culminates at 3404m with the Aneto peak, and constitutes a real geographical barrier between the Iberian peninsula and the rest of Europe to the north. Physiographically, three main regions can be distinguished from west to east: the Atlantic Pyrenees

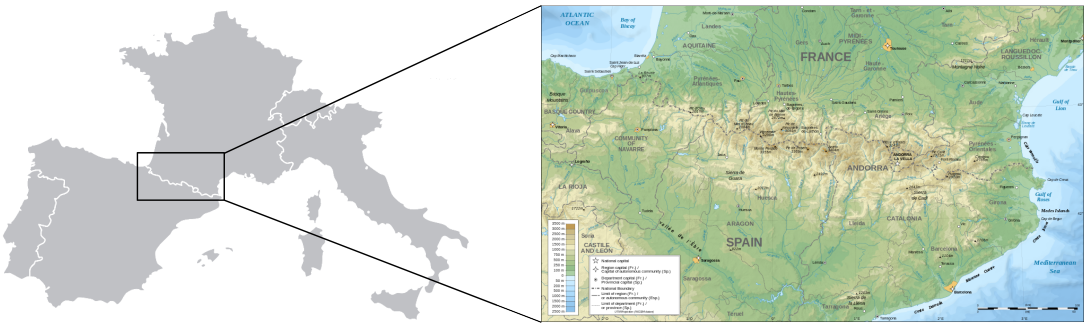


Figure 2.1: Location map of the study area, covering a region of $410x210km^2$ in both French and Spanish territories, for the automatic generation of land cover maps at 30m resolution.

or Low Pyrenees, with elevation gradually increasing eastwise (not surpassing the 2000m of altitude); the Central Pyrenees, with the highest summits (surpassing the 3000m of altitude, such as the Aneto peak or Monte Perdido at 3335m); and the Eastern Pyrenees or Oriental Pyrenees, characterised by a uniform elevation (with high peaks such as Canigó at 2784m) followed by a sudden decline towards the east [Ibarra Benlloch, P., 2002]. The landscape of the chain is characterised by an absence of great lakes, and the presence of a great number of small lakes and ponds, some of the bigger ones being artificial dams. Some glaciers are present in the Atlantic and Central Pyrenees, towards the northern slopes, near the crest of the highest peaks. From north to south, the wideness of the chain varies from 150km in the Central Pyrenees to 25-30km in the Atlantic Pyrenees and 10km in the Oriental Pyrenees [Ibarra Benlloch, P., 2002].

A climatic dissymmetry characterizes the Pyrenees area, putting into contact two wide types of climate, oceanic and Mediterranean. The snow-line varies between 2800 to 3100 metres above sea level, and both snow and rain precipitation quantities are greater in the western half of the area. To the north, rainfall is well spread throughout the year; to the south, especially in lower areas, a dry season appears in the summer and the number of rain days in the summer is lower [Gruber, 1980a]. In the north slopes the oceanic character is dominant: in low areas, rainfall is considerable (900 to 1500mm) and favours Atlantic vegetation; in higher areas, rain and snow fall are more important (1500 to 3000mm). In the south slopes, three climatic areas can be distinguished: Mediterranean climate up to 600 or 700m (with precipitation under 800m and a dry summer period); subMediterranean climate between 600 and 1300m (precipitation between 800 to 1200mm and disappearing dry summer period); and finally high mountain climate over 1300m with well-distributed annual precipitation from 1200 to 2000mm [Gruber, 1980b].

Regarding temperature ranges, typically the area has been studied in stages, with some differences regarding their definitions and altitude limits [Ibarra Benlloch, P., 2002], [Gruber, 1980a], [Gruber, 1980b]. Following [Ibarra Benlloch, P., 2002], the alpine stage, over 2000-2300m of altitude, has an annual average temperature of under 3°C, the coldest month having an average minimum of -8°C and an average maximum of 0°C. The subalpine stage, between 1600 and 2200m (from 1400m for northern slopes and from 1800-2000m for southern slopes) has an annual average of 3°C to 6°C, with minimum averages for the coldest month between -4°C and -8°C and maximum averages between 0 and 3°C. It is widely represented in the Pyrenees chain, with a freezing period of 5 to 6 months. The mountain stage, between 500 and 1700 m, shows less extreme temperatures,

with an annual average between 6°C and 12°C. For the coldest month, the average minimum ranges between -2°C and -4°C, and the average maximum between 3°C and 10°C. The freezing period is reduced to 3 to 5 months and not coincident with the summer months. Finally, the colinean stage, under 400m, corresponds to smoother temperatures more favourable for crops, with an annual average over 12°C, a minimum average for the coldest month over 2°C, and a maximum average over 10°C.

Climatic conditions, including temperature ranges and precipitation characteristics are determinant for the distribution of vegetation. Gruber gives a comprehensive stage study of the Pyrenees in [Gruber, 1980b], including a supraMediterranean and a mesoMediterranean stage for the southern slopes corresponding to the colinean stage of the northern slopes, as well as a description of the main vegetation types to be found in each of the stages. [Ibarra Benlloch, P., 2002] provides also a general characterisation of the Pyrenees including vegetation patterns.

The Pyrenees range embeds the Andorran Principate, and partially includes both Spanish and French territories. On the French side, three regions are concerned: from west to east, Aquitaine (Pyrénées Atlantiques department), Midi-Pyrénées (Haute-Pyrénées, Haute-Garonne and Ariège) and Languedoc-Roussillon (Aude and Pyrénées Orientales). On the Spanish side, four autonomic communities: País Vasco (Álava and Guipuzcoa provinces), Navarra, Aragón (Zaragora and Huesca), and Catalunya (Lleida, Girona and Barcelona). All these administrative regions are included in the study area, as well as further regions in the proximity of the mountain chain, in particular the vast plain areas north of the range, as far as Biarritz, Bayonne and Labenne in the Atlantic coast, and comprising Toulouse in the Haute-Garonne. The site extends towards the east to Narbonne in the Mediterranean sea and north to the Parc naturel régional des Grands Causses. Overall, the study region presents marked differences in terms of altitude, with a significant gradient between the north plane regions and coastal areas, the Pyrenean foothills, and the high peaks of the Pyrenees mountainous range.

2.2 Data description

Aiming at the automatic production of large scale land cover maps with a view to prepare the Sentinel-2 mission, with the launch of the second satellite of the twin-satellite configuration expected soon (Sentinel-2B has been announced as ready for environmental testing during the time of writing in June 2016, [EC, 2014]), the methodology here presented is based on the use of two different Landsat missions satellite data time series simulating the target mission datasets. The application of exogenous information, when available, in combination with the satellite time series, is implemented and tested using topographic and meteorological data in order to improve classification results. This section presents the collection, preprocessing, and characteristics of the different datasets that have been applied, as well as those used for training and validation.

2.2.1 Landsat data time series for simulation

A combination of Landsat-7 and Landsat-5 time series covering the chosen Pyrenees site have been used in order to simulate the soon-available Sentinel-2 twin-satellite data. The next subsections under this heading summarize the main characteristics of the Landsat and Sentinel programmes

and give an overview of the Landsat-5, Landsat-7 and Sentinel-2 missions. Also, the specifics of Sentinel-2 time series simulation with a combination of Landsat-5 and Landsat-7 datasets are given.

2.2.1.1 Landsat programme

NASA's Landsat is the longest-running Earth Observation programme with civil objectives with 8 satellites launched between 1972 and 2013. It has provided over 40 years of continuous observation of the Earth; Fig.2.2 provides a graphical illustration of the life spans of each of the missions of the programme and their expected continuity.

Landsat features a unique spatio-temporal combination, with resolution fine enough to monitor land cover changes, and footprint large enough for wide-area applications [Wulder et al., 2012], together with a good revisit cycle. The programme has an unmatched value, with an archive offering a history of the evolution of the planet surface during the last 42 years. Furthermore, with a change to an open data policy in 2008, making available the nearly three million images of the United States Geological Survey (USGS) archive [Loveland and Dwyer, 2012], Landsat has also made possible the explosion of research for large-area applications with time series, of which this work is an example. The new data policy has other implications, as explained in Chap.1, for example in the field of international treaties, as it allows independent monitoring and verification of the compliance to agreements related for instance to deforestation [Wulder et al., 2012].

The consistency of measurements has been ensured throughout the missions, with incremental improvements. Landsats 1 to 5 carried the Multispectral Scanner System (MSS) with four spectral bands in the visible (Landsat 3 added shortly a thermal band) and a spatial resolution of 80m. Landsats 4 and 5 introduced the Thematic Mapper (TM) with seven spectral bands including middle- and thermal-infrared and worked at 30m resolution (120m for the thermal band). Landsat 7, carrying the Enhanced Thematic Mapper Plus (ETM+), continued the TM configuration with a 60m thermal band and adding a panchromatic band at 15m.

The recently launched Landsat 8 (LDCM) features the Operational Land Imager (OLI) and the Thermal Infrared Sensor (TIRS), which add to the previous setting an ultra-blue band, a short-wave infrared band, and a further thermal-infrared band. Thus, eleven bands are available with the last of the programme missions, with resolutions of 30m except for the panchromatic band (15m) and the thermal-infrared bands (100m) [Loveland and Dwyer, 2012].

Landsats 1 to 3 featured an orbit at 917km and a revisit time of 18 days, while the rest of missions have been configured for an orbit of 705km and a revisit time of 16 days. Tab.2.1 gives an overview of the Landsat missions characteristics, specifying altitude, revisit times, sensors and lengths of operation.

2.2.1.2 Landsat-5 and Landsat-7 data

Only Landsat-5 and Landsat-7 data have been used in this work. Launched in 1984 and 1999, respectively, the satellites have a sun-synchronous orbit at an altitude of 705 km. Their period of revolution is 99 minutes, thus obtaining an orbiting of the Earth of approximately 14.5 times per

System	Launch (Service end)	Sensors	Resolution (m)	Altitude (km)	Revisit (days)
Landsat-1	23/07/1972 06/01/78	RBV MSS	80 80	917	18
Landsat-2	22/01/1975 25/02/82	RBV MSS	80 80	917	18
Landsat-3	05/03/1978 31/03/83	RBV MSS	40 80	917	18
Landsat-4	16/07/1982 TM failure 08/1993	MSS TM	80 30	705	16
Landsat-5	01/03/1984 05/06/2013	MSS TM	80 30	705	16
Landsat-6	05/10/1993 05/10/93	ETM	15 (pan) 30 (ms)	705	16
Landsat-7	15/04/1999	ETM+	15 (pan) 30 (ms)	705	16
Landsat-8 (LDCM)	11/02/2013	OLI TIRS	15 (pan) 30 (ms) 100	705	16

Table 2.1: Landsat missions characteristics ([NASA, 2014], [USGS, 2014]).

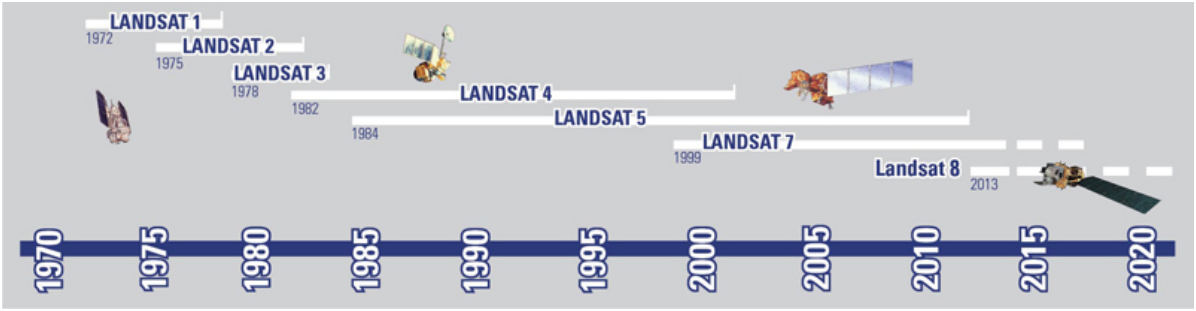


Figure 2.2: Landsat missions chronology ([USGS, 2014].

day. With an inclination of 98.2° and a swath of 185 km, the satellites have a repeat coverage of 16 days (8 days with two satellites). Landsat-7 is still operational in spite of the Scan Line Corrector (SLC) failure that it experienced in 2003. Landsat-5 was decommissioned on 5 June 2013, and its orbit is now used by Landsat-8 (LDCM). These general characteristics of the two satellites are summarised in Table.2.2.

The TM/MSS and ETM+ sensors of Landsat-5 and Landsat-7 range from 0.45μm to 2.35μm in the spectrum, yielding imagery in seven reflective bands (Blue, Green, Red, Near-infrared, Middle-infrared, Thermal, Middle-infrared) at a resolution of 30m except for the Thermal band, at 120m for Landsat-5 and at 60m for Landsat-7. A panchromatic band at 15m resolution is also available with Landsat-7 and Landsat-8. Details on the available bandwidths of the satellites are summarised in Table.2.3.

	Landsat-5 / 7
Launch Date	March 1, 1984 / April 15, 1999
Status	decommissioned / operational despite Scan Line Corrector (SLC) failure (May 31, 2003)
Sensors	TM, MSS / ETM+
Altitude	705 km
Inclination	98.2°
Orbit	polar, sun-synchronous
Equatorial Crossing Time	nominally 9:45 AM / 10AM (+− 15 min.) local time (descending node)
Period of Revolution	99 minutes; ≃14.5 orbits/day
Swath	185 km
Repeat Coverage	16 days (8 days with 2 satellites)

Table 2.2: Features of the Landsat 5/7 satellites [USGS, 2014]

Landsat-5	TM width (μm)	Resol. (m)	Landsat-7	ETM+ width	Resol.
Band 1	0.45 - 0.52	30	Band 1	0.45 – 0.52	30
Band 2	0.52 - 0.60	30	Band 2	0.53 – 0.61	30
Band 3	0.63 - 0.69	30	Band 3	0.63 – 0.69	30
Band 4	0.76 - 0.90	30	Band 4	0.78 – 0.90	30
Band 5	1.55 - 1.75	30	Band 5	1.55 – 1.75	30
Band 6	10.40 - 12.50	120	Band 6	10.40 – 12.50	60
Band 7	2.08 - 2.35	30	Band 7	2.09 – 2.35	30
			Pan Band	0.52 - 0.90	15

Table 2.3: Bandwidths and resolutions of the Landsat 5/7 sensors [NASA, 2014]

2.2.1.3 Preprocessing

Landsat scenes for the study site (scenes 200-030, 199-030, 198-030 and 197-030) were provided by the United States Geological Survey (USGS) and the European Spage Agency (ESA) for the period 2009 to 2011, orthorectified and non-orthorectified, respectively. Preprocessing of these scenes was performed at CESBIO by O. Hagolle and M. Huc and consisted of the following stages, as described in [Hagolle and Huc, 2011]: orthorectification of the ESA data with CNES SIGMA [Bignalet-Cazalet et al., 2010], reprojection to Lambert-93, calibration based on [Chander et al., 2009], aerosol detection [Hagolle et al., 2008], and atmospheric, environment and slope correction.

Data were tiled in tiles of 110x110km with an overlap of 10km, as illustrated in Fig.2.6. The resulting time series consist of different number of images for each period and region, unevenly distributed in time, and with significant percentages of missing data. Fig.2.3 illustrates, for each of the 8 Landsat tiles here considered and corresponding to the chosen study area, acquisition dates for which any quantity of information is available, independently of its size in terms of pixel count. Valid information for each date is often scarce and can even approach 0% depending on the presence of clouds, cloud shadows and passage tracks. Missing information is studied in Chap.4.

The tile notation used in this work follows the convention L93xxyy defined in [Hagolle and Huc, 2011] allowing the calculation of tile origins (0,0). L93 stands for the projection Lambert 93, and computation uses step=100020m and margin=9990m:

$UL_x = xx \times step$ (Upper Left Corner, pixel centre)

$UL_y = yy \times step$
$LR_x = (xx + 1) \times step + margin$
$LR_y = (yy + 1) \times step - margin$

As exemplified by the authors, L930563 corresponds to Lambert 93, $xx = 05$, $yy = 63$, and thus the coordinates of the top left pixels in Lambert 93 are $(UL_x = 800160m, UL_y = 6301260m)$. The described naming convention is retained here.

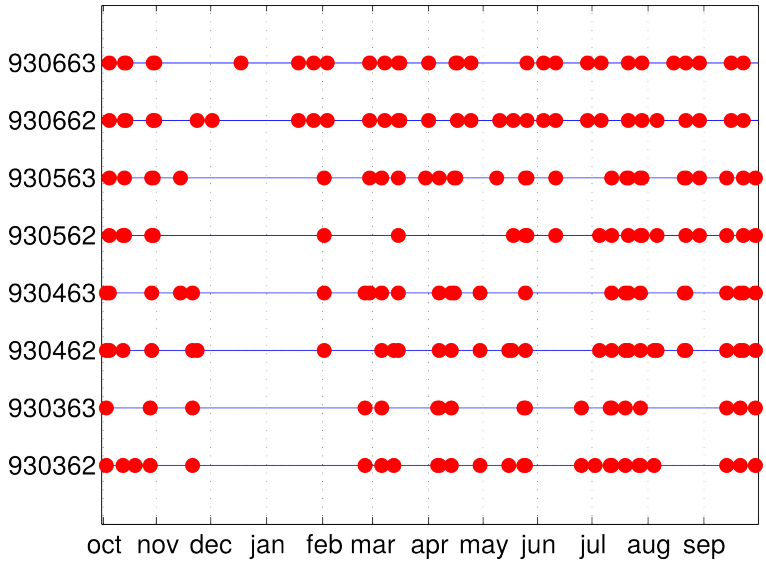


Figure 2.3: Data date correspondences for the studied crop year 2009-2010, irregularly spaced in time and different for each numbered Landsat tile belonging to the study area. Available valid information for each date is often scarce and can even be close to 0%, depending on passage tracks and presence of clouds and shadows. See Tab.4.1 in in Chap.4 for valid information percentages.

The MultiTemporal method for Cloud Detection (MTCD) by [Hagolle et al., 2010] was applied by O. Hagolle and M. Huc at CESBIO in order to obtain cloud and cloud shadow masks for each tile, as illustrated in Figs.2.4 and 2.5, the approach obtaining better discrimination of cloud-affected and cloud-free pixels than traditional reflectance-based methods. The original RGB composite, selected areas for clouds and clouds shadows, and final masks with labeled areas are shown, for both a Landsat tile and a detail of the same tile. Finally, passage border and snow masks were also derived, snow detection being based on spectral behaviour by application of [Riggs et al., 1994] and [Irish, 2000].

2.2.1.4 Sentinel programme

The Copernicus programme, previously known as Global Monitoring for Environment and Security (GMES) is a programme of the European Commission aiming at an autonomous and operational European Earth Observation capacity with six thematic areas: land, marine, atmosphere, climate change, emergency management and security.

The Sentinel programme is fundamental part of Copernicus with the current development of a fleet of new satellites including radar and multi-spectral imaging for Earth Observation, grouped

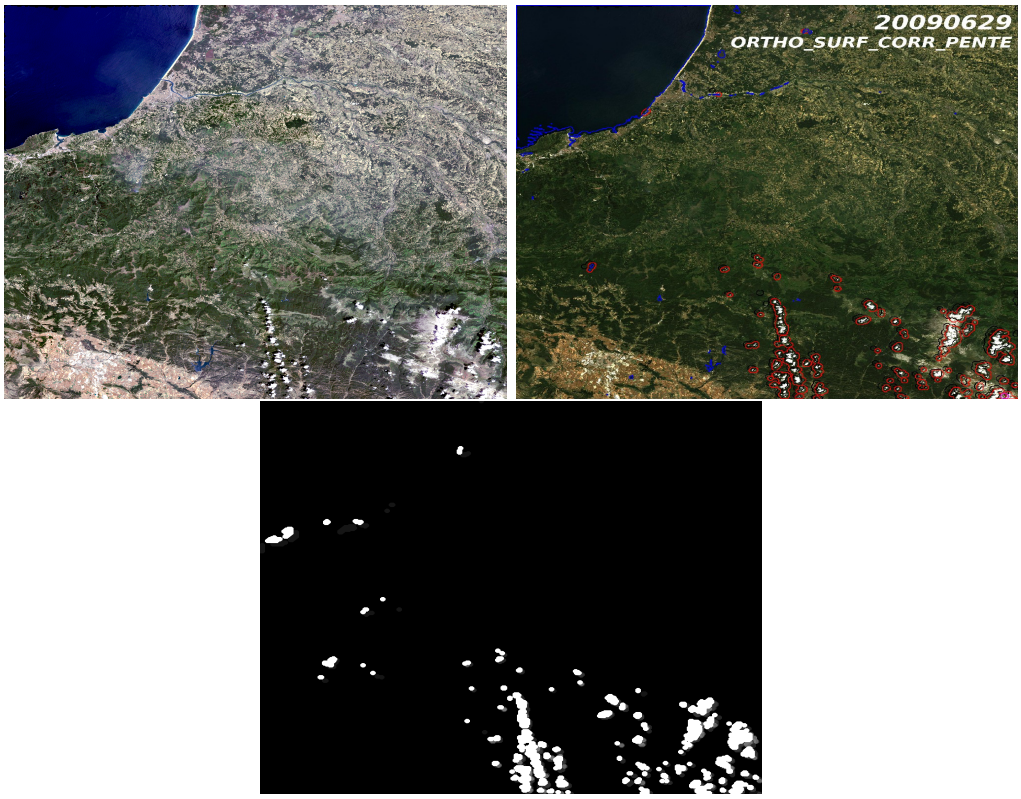


Figure 2.4: RGB composite with observable cloud cover (mostly in the bottom right area), detected cloud regions (in red), and derived masks. From [Hagolle and Huc, 2011].

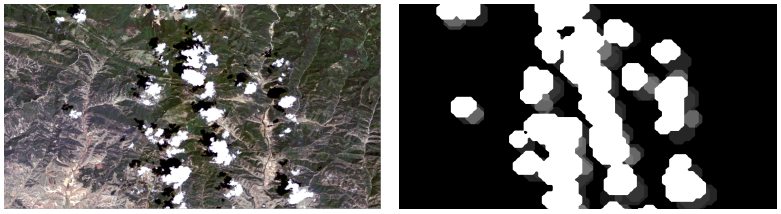


Figure 2.5: Detail of RGB composite with visible cloud and cloud shadow regions and corresponding cloud and cloud shadow masks (grey levels indicate missing information type and detection characteristics).

in five different families. It is undertaken by the European Space Agency (ESA), with around 30 existing or planned Contributing Missions making data available to Copernicus, such as ERS, Envisat, MetOp, the French SPOT and PLEIADES, the German TerraSAR-X, and the Earth Explorers. Notably, two of the Earth Explorers have been proposed by CESBIO: the Soil Moisture and Ocean Salinity mission (SMOS), launched on 2 November 2009, and the BIOMASS mission, scheduled for launch in 2020.

Each of the Sentinel missions is constituted by a constellation of two satellites, in order to obtain the revisit and coverage requirements of the Copernicus programme. Briefly, the Sentinel missions are described below [EC, 2014]:

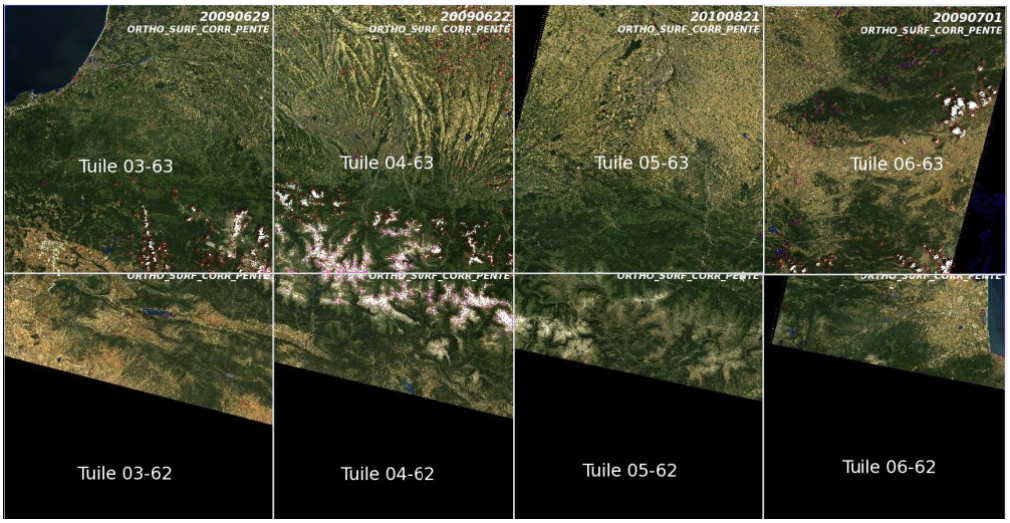


Figure 2.6: Landsat data preprocessed at CESBIO ($110 \times 110 km^2$, 3667x3667p, 8 tiles, 30m) [Hagolle and Huc, 2011]

- Sentinel-1: polar-orbiting radar imaging mission for land and ocean monitoring.
- Sentinel-2: multi-spectral high-resolution imaging mission for land monitoring and emergency services.
- Sentinel-3: multi-instrument mission (optical, radar and altimetry data) for surface and temperature measuring for forecasting, environmental and climate monitoring.
- Sentinel-4: payload on geostationary Meteosat Third Generation-Sounder (MTG-S) for atmospheric monitoring .
- Sentinel-5: payload on polar-orbiting MetOp Second Generation satellite for atmospheric monitoring.
- Sentinel-5 Precursor: mission for data gap reduction between Envisat and Sentinel-5 for atmospheric monitoring.

The first satellite of the Sentinel-1 mission (Sentinel-1A) was launched on 3 April 2014, and it is currently operational, with the launch of the second one (Sentinel-1B) being planned for 2016. The first of the Sentinel-2 satellites (Sentinel-2A) was launched on 23 June 2015, during the time of writing, and the launch of the second one (Sentinel-2B) to complete the Sentinel-2 twin-satellite configuration is programmed for the end of 2016. Finally, the Sentinel-4 and Sentinel-5 missions are scheduled for 2020 [EC, 2014].

2.2.1.5 Sentinel-2 data

The Sentinel-2 mission comprises two twin polar-orbiting satellites, phased at 180° in the same orbit, with a high revisit time (under cloud-free conditions, 10 days at the equator with one satellite, and 5 days with 2 satellites, resulting in 2-3 days at some latitudes) [Fletcher, K., 2012].

With an inclination of 98.62° and a swath of 290 km, they orbit at an altitude of 786 km and their period of revolution is 100.7 minutes, orbiting the Earth 14+3/10 times per day. General characteristics of the two satellites are summarised in Table.2.4.

	Sentinel-2
Status	being developed
Sensors	MSI (MultiSpectral Instrument)
Altitude	786 km
Inclination	98.62°
Orbit	polar, sun-synchronous
Equatorial Crossing Time	nominally 10:30 AM local time (descending node)
Period of Revolution	100.7 minutes; 14+3/10 orbits/day
Swath	290 km
Repeat Coverage	10 days (5 days with 2 satellites)

Table 2.4: Features of the Sentinel-2 satellites [Fletcher, K., 2012] [ESA, 2010]

The MSI sensors of the Sentinel-2 satellites range in the spectrum from the visible to the short-wave infrared (0.43 to 2.28μm), yielding imagery in 13 reflective bands including four bands at 10m spatial resolution (Blue, Green, Red and Near-infrared), six bands at 20m (four bands in the vegetation red-edge domain and two large middle-infrared bands) and three bands at 60m (for aerosol and water vapour retrieval and cirrus cloud detection) [Fletcher, K., 2012]. Details on the available bandwidths of the satellites are given in Table.2.5. Fig.2.7 illustrates the spectral bands of the Sentinel-2 satellites and their corresponding spatial resolutions.

Sentinel-2	Center (nm)	Width (nm)	Resol. (m)
Band 1	443	20	60
Band 2	490	65	10
Band 3	560	35	10
Band 4	665	30	10
Band 5	705	15	20
Band 6	740	15	20
Band 7	783	20	20
Band 8	842	115	10
Band 8a	865	20	20
Band 9	945	20	60
Band 10	1375	30	60
Band 11	1610	90	20
Band 12	2190	180	20

Table 2.5: Bandwidths and resolutions of the Sentinel-2 satellites [ESA, 2010]

2.2.1.6 Sentinel-2 time series simulation

Sentinel-2 aims at ensuring the continuity of Landsat and SPOT data, and indeed the spectral band configuration of Sentinel-2 was developed around the use of Landsat and SPOT wavelengths, with some modifications and additions. For example, Landsat's near-infrared band was narrowed to reduce water vapour contamination and to enhance sensitivity to soil iron oxide content; a new

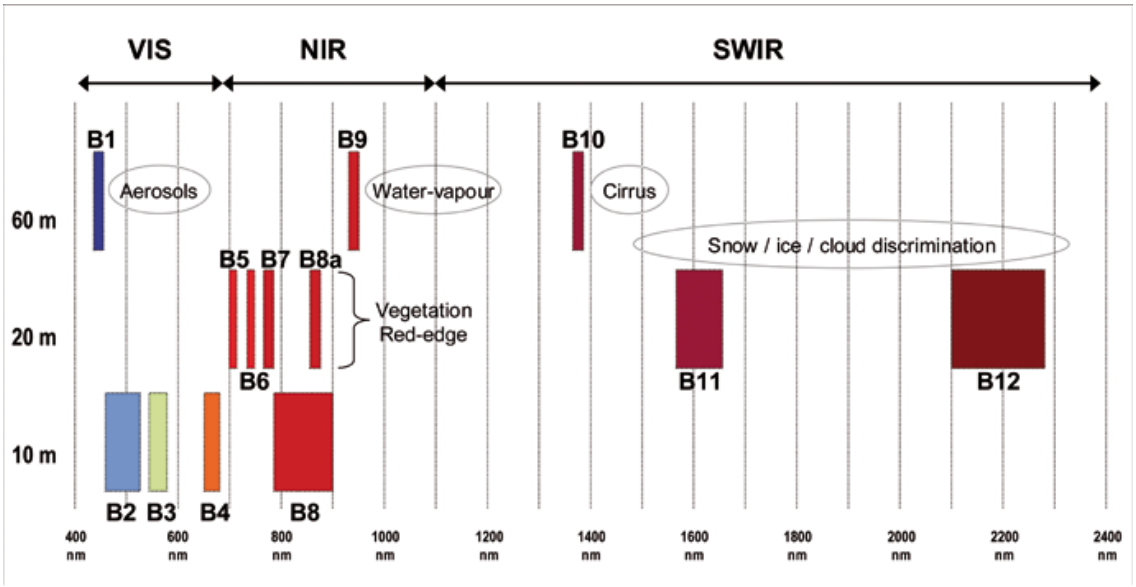


Figure 2.7: Sentinel-2 13 spectral bands (nm) and their corresponding spatial resolutions (m) [Fletcher, K., 2012]

.

band in the blue domain was included for precise aerosol correction, and another at 1375 nm for cirrus cloud detection, due to their impact on reflectance values.

Fig.2.8 shows graphically the similitudes between the spectral bands of Sentinel-2 and Landsat 8 (LDCM). The Landsat 8 OLI sensor adds two new bands to those of Landsat-7 and Landsat-5 (which have been used here), a coastal band and a cirrus band, with some refinement for six of the heritage bands. No thermal band is present in the Sentinel-2 satellites.

Thus, due to their similarities, as shown in Tab.2.6, Landsat data (30 m resolution, 185 km field of view, 7 bands including SWIR, and a repetition cycle of 8 days achieved through the combination of data from the satellites Landsat-5 and 7) is used here as a simulator for Sentinel-2 (10, 20 and 60 m resolution, 290 km field of view, 13 bands including SWIR, and a repetition cycle of 5 days with two satellites), although spatial, temporal and spectral resolutions will be superior with the latter. Also, the Landsat repetition cycle is only partially attainable with the two combined 16-day revisit frequencies, since only the central area of Landsat-7 is usable since the failure of the ETM+ scan line corrector in 2003, resulting in a 22% data loss per capture [Loveland and Dwyer, 2012].

The spectral bands of Landsat-5 and Landsat-7 range from 0.45 to 12.50μm, and they include blue (B1), green (B2), red (B3) and near-infrared channels (B4), two middle-infrared channels (B5 and B7), and a thermal band (B6, which is not used here, as Sentinel-2 lacks such a band). The 13 Sentinel-2 bands will range from 0.43 to 2.28μm, six of its spectral channels being considered as heritage of the Landsat system [ESA, 2010]. Indeed, the Sentinel-2 B2, B3, B4, B8, B11 and B12 bands correspond to the Landsat B1, B2, B3, B4, B5 and B7 spectral channels, what explains our choice of data. As only the four VNIR bands of Sentinel-2 will attain a 10m spatial resolution, with the three bands dedicated to atmospheric correction having a resolution of 60m, land cover

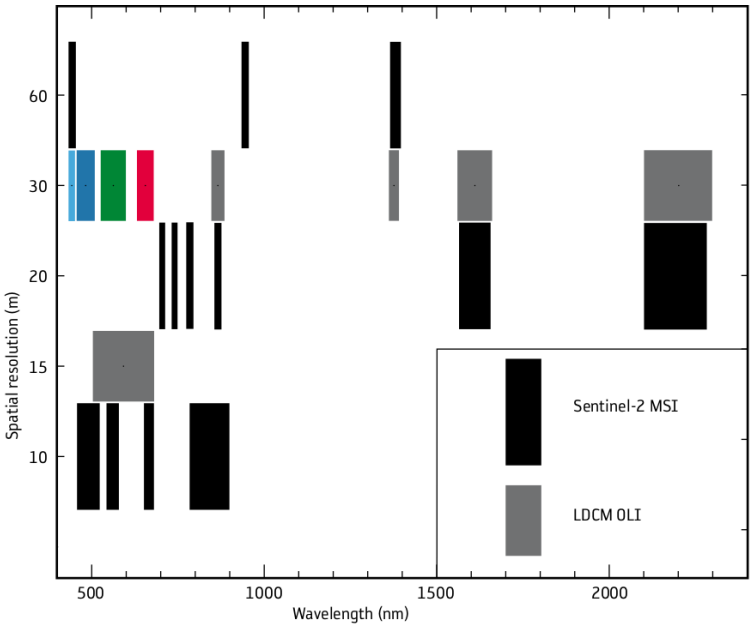


Figure 2.8: Comparison between the 13 spectral bands of Sentinel-2 (in black) and the 11 spectral bands of Landsat 8 (dark grey and colour) [Fletcher, K., 2012]

.

mapping work on the data is expected to be carried out at resolutions of 10m and 20m. A summary of the general characteristics of the systems is given in Tab.2.6.

Satellite sensor	Landsat-5 / 7	Sentinel-2
Measurement principle	Scanner	Pushbroom
Multispectral resolution (m)	30-120/60 m	10-20-60 m
Multispectral bands	7	13
Repeat cycle (days)	8 (2 satellites)	5 (2 satellites)
Swath (km)	185	290
Coverage	All sites	All sites
Launch date (year)	1984 (Landsat-5)	2015 (Sentinel-2A)
	1999 (Landsat-7)	2016(planned) (Sentinel-2B)

Table 2.6: General characteristics of the satellite data sources used in this study and of the twin-satellite mission Sentinel-2 [ESA, 2010] [Hagolle et al., 2010] [Fletcher, K., 2012]

2.2.2 Training and validation data

Land cover validation data have been generated based on the Registre Parcellaire Graphique (RPG) and the following datasets of the Institut Géographique National (IGN) [IGN, 2014]: the French national forest inventory, the hydrological resources database DB CARTHAGE ® (Base de Données sur la CARtographie THématique des AGences de l'Eau) [IGN, 2006], and the DB TOPO ® for buildings and transportation systems [IGN, 2014]. J-F.Dejoux and M.Kadiri at CESBIO performed the collection and the grouping of the data into a unified dataset, respectively. M.Arias also at CESBIO undertook the production of an final dataset, which has been applied as reference

data throughout in this work, and of a refined version that has been applied in Chap.4 to evaluate the impact of ground truth accuracy on results.

The RPG is a geographical information system allowing the identification of agricultural plots, administered by the Agence de Services et de Paiement (ASP), and used for the management of European funds. It contains around 6 million plots and more than 27 million hectares annually declared by nearly 400000 farmers [ASP, 2013]. It is updated based on the IGN database DB ORTHO ®, an orthophoto dataset with resolutions from 20 to 50cm [IGN, 2011], for the 94 departments of metropolitan France and the 4 overseas departments. Fig.2.9 illustrates the definition of plots for the RPG on an orthophoto from the IGN dataset.

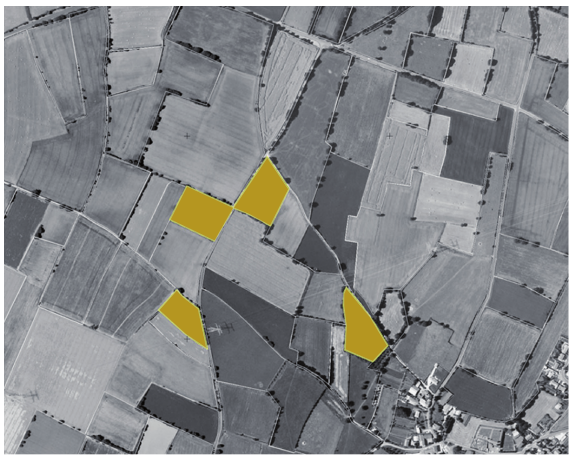


Figure 2.9: RPG plot definition on DB ORTHO ® imagery

The French national forest inventory is also a service of IGN, in charge of the permanent listing of the national forest resources independently of property issues. It applies a minimum mapping unit of 0.5 Ha (extension determined by the United Nations Food and Agriculture Organisation (FAO) definition of *forest* [FAO, 2013]). Its update frequency was up to ten years before 2004, due to a regional update policy; currently, annual inventories are produced based on a moving systematic area sampling for aggregation with previous annual campaigns [IGN, 2014]. This allows the production of national results every 5 years, and determines a variable cycle-specific accuracy for each area, depending on the sampling. DB CARTHAGE ® describes the hydrographic features of the French territory for a scale ranging from 1:50000 to 1:100000, with decametric precision [IGN, 2006], and an annual regional update.

Finally, the DB TOPO ® contains a 3D vectorial description of the elements of the territory and infrastructures with a metric output, including buildings, administrative structures and road, train and transportation networks. Geometrical precision is variable depending on the acquisition mode, the planimetric precision varying from 0.5m to 30m and the altimetric precision ranging from under 1m to 20m [IGN, 2014]. The DB Topo presents variable update ranges: from under 6 months to 4 years for roads, and from 3 to 5 years for any information derived from aerial photography, such as building areas [IGN, 2014].

The data sources heterogeneity explained above, with different update rates, collection methods, product resolutions, and accuracy estimates, affects the reliability of the ground truth applied for training and validation, as it is later discussed in Sec.2.4. Tab.2.7 summarises the number of available ground truth pixels for training and validation that have been used as a basis in this work. Each of the columns lists the corresponding pixels for each of the Landsat tiles and each of the classes that have been applied. Nomenclature is explained in Sec.2.3. The number of classes that are represented per tile is also specified.

Classes/Tiles	0362	0363	0462	0463	0562	0563	0662	0663	Total
Corn	0	34188	51	188772	0	29449	0	4223	256683
Sunflower	0	921	0	10939	0	96450	11	41545	149866
Other summer crops	0	950	0	5228	0	25202	221	7872	39473
Wheat	0	942	0	16975	0	53324	27	18426	89694
Barley	0	103	0	1610	0	11290	106	8207	21316
Rape	0	136	0	3036	0	18165	0	5285	26622
Other winter crops	0	1211	0	6233	0	223074	166	67088	297772
Deciduous broadleaf forest	10695	186140	43055	279280	239037	357271	90126	192447	1398051
Deciduous needleleaf forest	0	1846	0	383	0	0	0	211	2440
Persistent broadleaf forest	0	849	0	8	0	0	214804	44685	260346
Persistent needleleaf forest	67	50755	114047	146523	55309	53701	63891	81032	565325
Mixed forest	867	2063	1205	1589	1882	9328	11558	459	28951
Woody plants	0	0	0	302	0	1527	0	136	1965
Vineyards	0	627	0	2149	0	1036	4104	5552	13468
Fruit/nut/olive trees	0	200	0	486	0	861	350	815	2712
Permanent herbaceous crops	145	17937	20216	37935	12132	5373	2217	46144	142099
Temporary meadows/fallows	0	12201	575	28755	536	34127	6129	15124	97447
Herbaceous moorlands	7028	22805	41592	6722	7803	4410	2457	1449	94266
Shrub-populated moorlands	56	861	143	340	248	4392	1308	1094	8442
Rocky moorlands	0	0	7736	1047	0	0	0	0	8783
Water bodies	348	8960	3136	17010	2147	13756	9659	4630	59646
Impervious surfaces	0	694	186	701	136	1777	1991	2152	7637
Total	19206	344389	231942	756023	319230	944513	409125	548576	3573004
Represented classes:	7	20	11	22	9	19	17	21	

Table 2.7: Number of available ground truth pixels for each of the Landsat tiles, per class, with nomenclature as in Sec.2.3. The number of represented classes per tile is also given.

2.2.3 Exogenous information: topographic and meteorological data

The use of ancillary information is central for some of the contributions presented in this work, in particular the training data extraction approach for the targeted selection of training samples for supervised learning presented in Chap.5 and the non-linear temperature-driven sampling approach for an improved characterisation of land cover classes for classification described in Chap.6. This exogenous information includes digital elevation, meteorological and snow presence datasets. Sources and characteristics of these datasets are are briefly described in this section.

2.2.3.1 Digital elevation model

The digital elevation model used in this work is derived from the Shuttle Radar Topography Mission (SRTM) data and provided by the Consultative Group for International Agricultural

Research (CGIAR). The original SRTM, first released in 2000 and made available by USGS, constituted the first global high-quality digital elevation model at resolutions of one and three arcseconds [Mouratidis et al., 2010]. It was generated using single-pass synthetic aperture radar (SAR) interferometry, covering the Earth between latitudes 60°N and 57°S, and was jointly performed by NASA, the German Aerospace Center (DLR) and the Italian Space Agency (ASI). Vertical accuracies were defined as +-16m for 90% of the data, and existent no-information areas where water or lack of texture prevented calculation, were processed further by application of interpolation methods, as described by [Reuter et al., 2007] [Jarvis et al., 2008]. CGIAR's 90m-resolution SRTM digital elevation model has been resampled at CESBIO in order to match the 30m spatial resolution of Landsat. Fig.2.10 presents the used model over the study area.

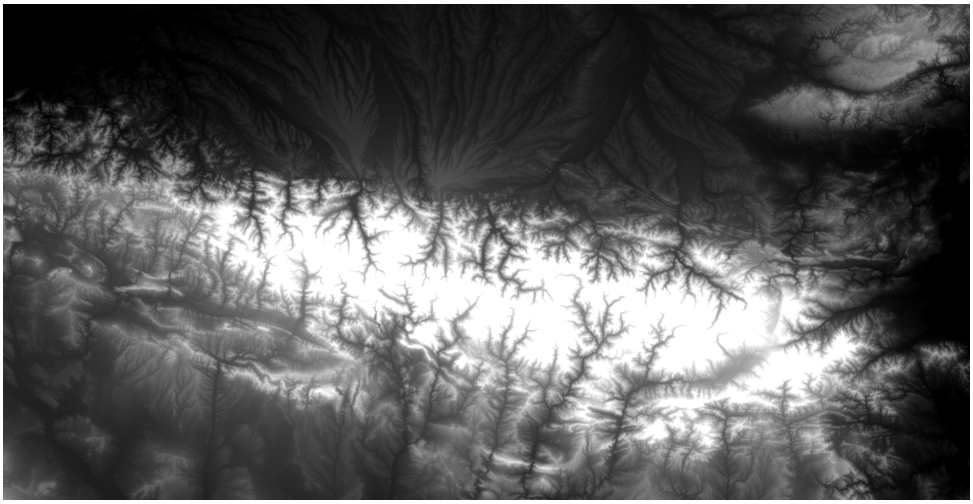


Figure 2.10: CGIAR 90m SRTM digital elevation model for the study area resampled at 30m.

2.2.3.2 SAFRAN meteorological data

Temperature data applied in this study have been provided by SAFRAN (*Système d'Analyse Fournissant des Renseignements Adaptés à la Nivologie*), an atmospheric analysis system for surface variables by Météo-France, yielding ground data observations with an hourly time step. These observations include air temperature at 2m, wind speed at 10m, relative humidity at 2m, quantity and phase of precipitation (both solid and liquid), and radiation in the visible and the infrared.

They are based on a method of interpolation integrating the observations of the Météo-France network of stations, information on climatically homogeneous zones, and data on height differences [Quintana-Seguí et al., 2008], their outputs being provided for a grid of $8km$ resolution covering the entire France. With an irregular distribution of stations, as illustrated in Fig.2.11, the method of optimal interpolation [Gandin, 1965] is used in order to integrate observations, geographical distances and correlations using homogeneous climatic regions derived from the Symposium II sectioning of France, as illustrated in Fig.2.11. These have irregular shapes and generally an area inferior to $1000km^2$. The vertical component is taken into account by applying altitude cuts every

$300m$, so that SAFRAN can estimate values for each parameter, region and level of altitude, the source observations being inside or outside the regions. Measurement frequencies vary from $24h$ for precipitation to $6h$ for temperature and wind, and then temporal interpolation is applied to obtain 1h calculations [Quintana-Seguí et al., 2008]. Once values are estimated for each region, the analysis is also interpolated, this time spatially, to a regular grid at $64km^2$. The temperature observations of this final product have been applied in this work.

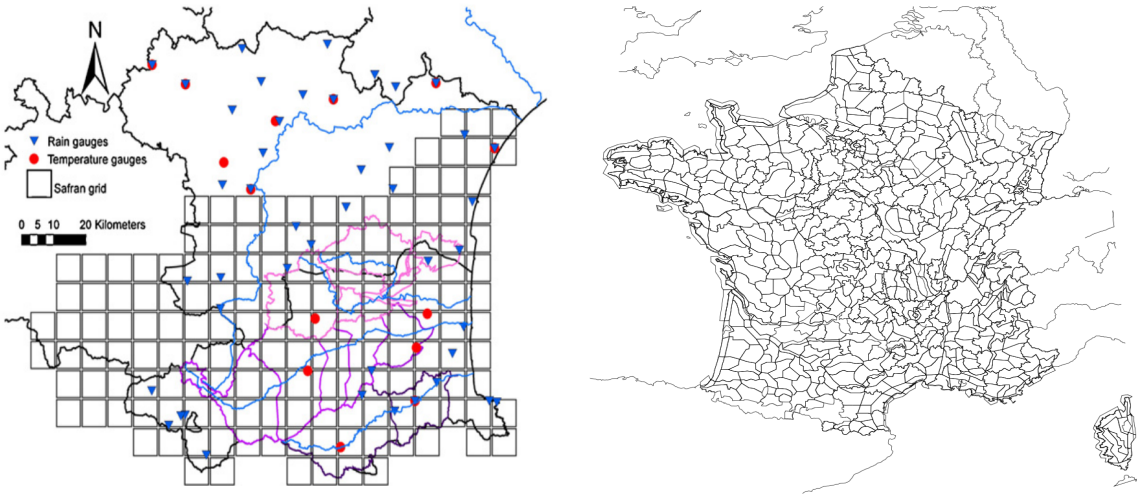


Figure 2.11: SAFRAN final grid and distribution of rain and temperature gauges for the southeast of France [Chaouche et al., 2010] and SYMPOSIUM II sectioning of France in 615 areas [Moigne, 2002]

2.2.3.3 MODIS snow information

The MODIS sensor, onboard both the Terra and the Aqua satellites, images the entire Earth every one to two days with 36 spectral bands between violet and thermal infra-red (0.405 to 14.385μm) and with resolutions of 250m, 500m and 1km. Many geophysical products are derived from collected data, offering information about land, oceans and atmosphere, such as snow cover and sea ice data products. The National Snow and Ice Data Center (NSIDC) Distributed Active Archive Center (DAAC) makes available MODIS snow and ice products since September 13, 2000 [Hall et al., 2002].

The MODIS snow datasets MOD10A2 and MYD10A2 [Hall et al., 2006] were processed at CESBIO by Simon Gascoin [Gascoin, 2013]. The initial 8-days snow cover syntheses at 500m resolution from Terra (MOD10A2) and Aqua (MYD10A2) for the period 01/01/2009 to 01/01/2012 were processed for gap-filling over the study area. A summary of the characteristics of the Terra and Aqua EOS satellites carrying the MODIS sensor is given in Tab.2.8.

As the snow mapping algorithm used to generate snow maps is based on the use of a Normalized Difference Snow Index (NDSI) exploiting the ratio between the short-wave infrared (Band 4) and the visible (MODIS Band 6), and due to the fact that 70% of the Band 6 detectors on MODIS Aqua failed shortly after its launch, the Terra product was used as reference for combining both products [Parajka and Blöschl, 2008]. Temporal adjacent deduction [Parajka and Blöschl, 2008] was applied considering periods of 24 days, and a spatial absolute majority filter was used, using the 8 adjacent

pixels, for temporal and spatial gap filling. Finally, a classification tree was created based on predictors (altitude, aspect, northing, easting) and applied to predict missing values. A final 8-days gap-filled snow cover synthesis was thus obtained.

MODIS sensor	Terra EOS	Aqua EOS
Launch Date	December 18, 1999	May 4, 2002
Status	operational (MODIS service dismissed from 14 April 2014)	operational (MODIS service dismissed from 14 April 2014)
Altitude	705km	705km
Inclination	98.1°	98.2°
Orbit	near-polar, sun-synchronous	near-polar, sun-synchronous
Equatorial Crossing Time	10:30AM local time (descending node)	01:30PM local time (ascending node)
Period of Revolution	98.88min; ≃14.5 orbits/day	98.8min; ≃14.5 orbits/day
Swath	2330km	2330km
Repeat Coverage	16 days	16 days

Table 2.8: Features of the Terra and Aqua satellites carrying the MODIS sensor [Parkinson et al.(Eds.), 2006] [NASA, 2014a].

2.3 Nomenclature

Tab.2.9 offers a base nomenclature of general land cover classes with their corresponding description. The final nomenclature applied in this work is based in this nomenclature but attains a finer level of detail for some of the labels and also adds a mixed class. It consists of 22 classes, including the following: wheat, corn, barley, rape, sunflower, other winter cereals, other summer cereals, woody plants, herbaceous moorlands, tree-populated moorlands, rocky moorlands, permanent grasslands, leys, orchards/nut trees/olive trees, vineyards, persistent coniferous forests, persistent broadleaved forests, deciduous coniferous forests, deciduous broadleaved forests, mixed forests, water areas, and impervious surfaces. Equivalences between nomenclatures are given in Tab.2.10.

2.3.1 Remarks

Several aspects need to be taken into account regarding the nomenclature here applied. Firstly, this work presents a methodology of land cover map generation at large scales, and the approach presented aims at generality: it aims to be applicable at any scale and so that any land cover targeted by users can be readily accommodated. Thus, the nomenclature that has been used is solely intended to be an example of legend to be aimed for.

Secondly, a combination of land cover and land use has been targeted for agricultural areas. Land cover refers to the observed biophysical cover of the Earth [Gregorio and Jansen, 2005], while land use has to do with human action on a land cover, and can encompass different land cover types; i.e. “recreation area” can refer to land cover types such as “sandy area” or “built-up area”.

Thirdly, it is left to the user to decide what legend to target depending not only on the purpose of the land cover map to be generated, but also on two other determinant factors: the spatial resolution of the satellite imagery input dataset, and the type of imagery itself. For example, roads can be well distinguished with very high resolution imagery such as that of Quickbird or

Class name	Description
Seasonal summer crop	Cultivated agricultural land with summer growth - corn, silage, sunflower, soy, peas, sorghum.
Seasonal winter crop	Cultivated agricultural land with springtime growth - wheat, barley, rape, forage.
Deciduous forest	Canopies of deciduous phenology - beech, deciduous oak, poplar.
Permanent forest	Needle-leaved forest cover of evergreen phenology -pine, fir, cedar, cypress, juniper- and broad-leaved forest cover of evergreen phenology -evergreen oak, holly, laurel.
Permanent woody crop	Ligneous crops with production during several years - vines, fruit trees, nut trees, berry trees, olive trees, arboriculture.
Meadows, fallows and herbaceous crops	Grasslands for grazing or forage - legumes and grasses- and lands left to lie idle -unproductive lands.
Moorlands	Uncultivated open land covered with herbage and low shrubs, common in high altitudes.
Water bodies	Permanent water surfaces - seas, oceans, streams, lakes.
Impervious surfaces and barren land	Buildings, pavements -roads, driveways, parking areas-, beaches, sandy areas, bare rock.

Table 2.9: Base nomenclature: general land cover classes and their description.

Class name	Subclass name	Code
Seasonal summer crop	Corn	COR
	Sunflower	SFL
	Other summer crops	OSC
Seasonal winter crop	Wheat	WHE
	Barley	BAR
	Rape	RAP
	Other winter crops	OWC
Deciduous forest	Deciduous broadleaf forest	DBL
	Deciduous needleleaf forest	DNL
Persistent forest	Persistent broadleaf forest	PBL
	Persistent needleleaf forest	PNL
	Mixed forest	MFO
Permanent woody crop	Woody plants	WOO
	Vineyards	VIN
	Fruit/nut/olive trees	FNO
Meadows, fallows and herbaceous crops	Permanent herbaceous crops	PHC
	Temporary meadows and fallows	TMF
Moorlands	Herbaceous moorlands	HMO
	Shrub-populated moorlands	SMO
	Rocky moorlands	RMO
Water bodies	Water bodies	WAT
Impervious surfaces and barren land	Impervious surfaces and barren land	IMP

Table 2.10: Equivalences between general classes (base nomenclature) and applied nomenclature of 22 classes (with a finer level of detail), with the corresponding classification codes used in the following chapters.

Pleiades, but difficulties arise with medium spatial resolutions such as that of Landsat at 30m. Another example of the impact of spatial resolution is the fact that it determines the possibility of using texture features. These can be effective when the texture of the targeted classes matches

the resolution of the applied imagery; for example, the contribution of texture analyses for crop classification may be advantageous with very high resolution imagery, although not so obvious with imagery at 30m [Ota et al., 2011] [Asner et al., 2002] [Lu and Weng, 2005].

The time span covered by the available input time series determines extraction feasibilities, as the occurrence in time of spectral responses characterising vegetation types depends on their phenological cycles. Thus, summer crops for example present a cycle starting in spring and finishing at the end of the summer. Characterisation requires thus satellite acquisitions during such months of phenological activity, and would not be possible if only winter acquisitions were available. In this way, users need to consider the available time span for target land cover class identification.

Also, the spectral bands of the satellite acquisitions and available exogenous information determine possible outcomes: for example, availability of data in the 500 to 1200 nm region allows for the detection of iron oxide content, otherwise difficult. Similarly, the availability of canopy height information or altitude data as a complement to spectral time series can help the identification of certain classes. In summary, it can be said that the legend needs to be sensibly set by the user based on needs but also considering scale dependency, and source dependency.

2.4 Data-related difficulties

Data-related difficulties stem generally from the characteristics of satellite imagery time series, presenting a considerable amount of artifacts and missing information, and also from the particular features of the chosen datasets, as well as from the inaccuracies of the training and validation data that are used to automatically generate land cover information and ultimately, to validate it. Also, preprocessing errors contained in the exogenous data that has been applied have an impact on the obtained results.

Generally, input satellite imagery is characterized by an important presence of missing information due to clouds, cloud shadows and satellite swath borders. These are impactive factors when dealing with non-synthetic datasets, which either hinder or prevent the production of land cover maps, depending on the map generation strategy. In addition to this, the future availability of great volumes of data with the next generation missions has processing cost implications, and prevents approaches still applied today based on manual or semiautomatic strategies. Also, due to the operation over large areas, a considerable spectral variability due to the large extent of the zones considered needs to be addressed.

More particularly, each satellite image time series presents different peculiarities and sources of error. For example, the Landsat-5 and Landsat-7 acquisitions that have been used in this work come from different sources, USGS and ESA, as explained. The superposition error after orthorectification for the USGS images has a standard deviation between 0.2 and 0.8 depending on images and cloud presence (the greater the cloud contamination, the bigger the error); that of the ESA images varies between 0.3 and 0.9 pixels [Hagolle and Huc, 2011], for the eight neighbours. This lack of perfect registration of the images results in a non-correspondence of pixels along the time series that necessarily impacts the classification, especially along border areas. Fig.2.12

illustrates the pixel non-correspondence that can be found in the dataset derived from this orthorectification error; for a small building area, the same pixel can appear as corresponding to different locations along the series. Also, Fig.2.13 shows examples of corregistration error between ground truth and land cover affecting water areas; both water and other types of land cover are contained in the water-labelled regions, which introduces error in both training and classification validation processes.

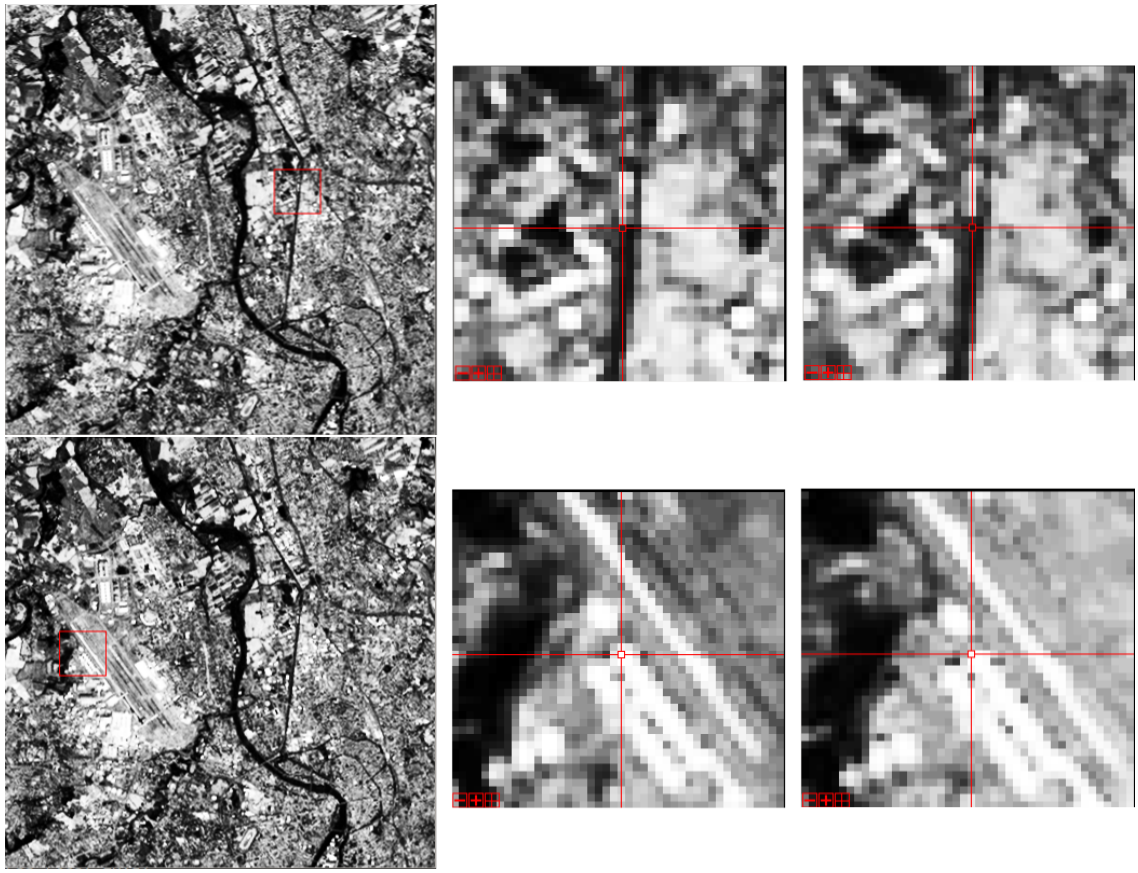


Figure 2.12: Orthorectification error illustration with different land cover types: water areas (water canal intersected by bridge, top images) and impervious surfaces (plane parking area, bottom images). Different relative locations depicted by the same corregistered pixel

Also, and mainly affecting the Landsat-5 images, the presence of lineage can be observed in transition areas east-west between bright and dark areas, a noise appearing due to a memory effect of the sensor as it scans side to side [Hagolle and Huc, 2011]. Another type of noise is impulse or "salt and pepper" noise, consisting of single-pixel bright or dark spots that do not correspond to acquired data, and observable specially in some ESA images. It corresponds to single-pixel value shifts due to transmission errors or problems in the ground processing systems [USGS, 2014]. Fig.2.14 exemplifies the two types of noise.

In relation to exogenous data, topographic data is derived from CGIAR's 90m-resolution SRTM DEM, further resampled to 30m in order to match Landsat's resolution, which introduces height

Figure 2.13: Corregistration error illustration with water-labelled areas overlayed on Google Earth imagery (covering most land at 15m resolution): labelled regions for training and validation present displacement in relation to land cover regions

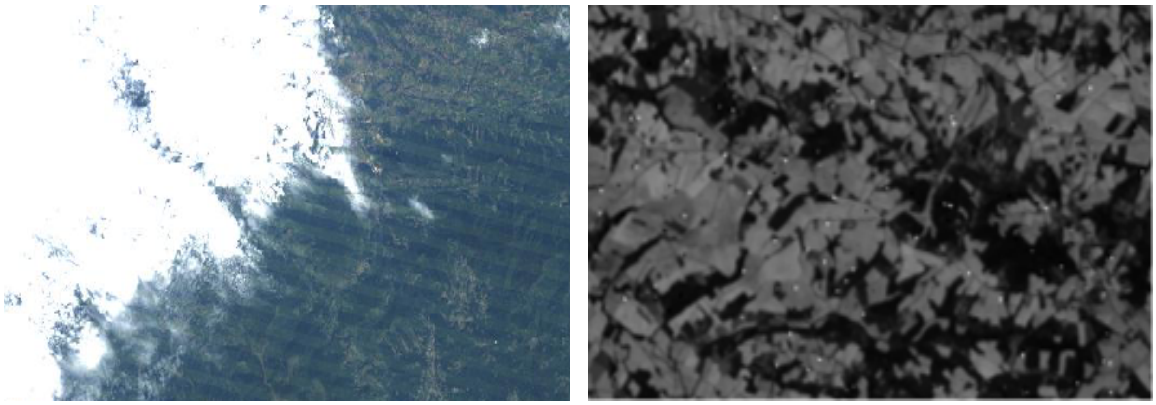


Figure 2.14: Examples of east-west transition between bright and dark areas, producing a lineage effect, and of "salt and pepper" noise

information mismatch. Indeed, resampling from a coarser resolution with a cell size of 90m to a finer resolution with a cell size of 30m does produce an output at 30m, but with an accuracy of 90m. This mismatch will be further impactive, as temperature information from SAFRAN data is corrected with height information. Also, cloud and cloud shadow masks obtained from the Landsat time series are calculated at 240m resolution in order to obtain short computation times and then resampled to 30m; this fact can have an effect on the detection of small clouds and passage border areas.

Regarding SAFRAN data, three sources of inaccuracies can be detected. Firstly, those related to the measurements themselves, which are in some cases biased. Indeed, SAFRAN data are sensible to the density of the network of observations (even more markedly in the case of rain and snow precipitation, where there is a trend to underestimate where the number of observations is

smaller, and also radiation in the visible and the infrared are nearly systematically underestimated [Quintana-Seguí et al., 2008]). Secondly, those derived from the application of spatial interpolation into a grid of $64km^2$; finally, inaccuracies appear as well due to the fact that some measurements are obtained every 6 or 24 hours, and then temporally interpolated into regular one-hour slots.

Finally, the available ground truth for training and validation is significantly problematic. As explained in Sec.2.2.2, the data sources (RPG, the French national forest inventory, DB CARTHAGE, DB TOPO) present a high heterogeneity at different levels: different update rates, collection methods, product resolutions, and accuracy estimates. In particular, resolution mismatch, source accuracy differences and the existence of non-updated information affects the accuracy of the ground truth itself and thus all derived products and undertaken validation.

One of the sources, the IGN forest inventory, applies a minimum mapping unit of 0.5 Ha, with variable update ranges and annual inventories based on a moving systematic area sampling and aggregation with previous annual campaigns [IGN, 2014]. Similarly, DB Carthage yields a regional yearly update [IGN, 2006]. Another of the sources, DB Topo, presents variable update ranges: from under 6 months to 4 years for roads, and from 3 to 5 years for any information derived from aerial photography, such as building areas [IGN, 2014]. Finally, the accuracy of the RPG, the inventory of crop region blocks in France, is based solely on the exactitude of annual farmer declarations [ASP, 2013], which may be non-exact for a variety of reasons. Thus, accuracy estimates are bounded to be affected by the presence of error, which occurs at different degrees depending on the sources.

This heterogeneity of inaccuracies has inevitably a certain impact on the quality of all products trained and validated with the available ground truth, which is difficult to quantify in the absence of an ideal training and validation dataset. Thus, accuracy scores need to be interpreted bearing in mind the lack of an ideal reference dataset such as that to be found in a synthetic product. However, real datasets are non-optimal and thus the applied ground truth exemplifies that to be found in a realistic land cover map generation scenario.

Chapter 3

Automation of learning and classification for large area land cover with time series

Contents

This chapter introduces the problem area of automatic classification for large area land cover generation and offers a review on the literature in relation to land cover classification and large area land cover classification, giving an overview of current global projects and their automation efforts. A fully automatic approach is then proposed for the generation of land cover maps across large areas with large-volume time series such as those to be provided by the coming satellite mission Sentinel-2. The methodology allows for the production of results in operational times (in some hours) and the managing of great volumes of information. Based on the exploitation of multitemporal and multispectral information, it uses a combination of automatic sample extraction, missing information processing, characterizing feature extraction, sampling, supervised model learning, and ultimately classification based on the learnt model. A description of the proposed scheme is given, with information on class characterisation, feature extraction, and the contributions of the approach; novel aspects, such as missing information processing through local interpolation in the time series and temporal resampling with virtual points, spatial sampling techniques application for training data extraction, and non-linear temporal sampling based on temperature, are further developed in the following chapters.

3.1 The problem of automatic classification

Land cover information is essential for a great diversity of areas, as described in Chap.1, ranging from environment studies, resources management and urban planning to sustainable development [Chen et al., 2015]. Indeed, land cover and land cover change analysis are determinant for the understanding of global scale processes and characteristics, such as biodiversity, hydrological and biogeochemical cycles [Giri et al., 2013]. Thus, land cover products are needed in a variety of thematic areas, as a product in itself, for example for deforestation, urban extension or crop rotation monitoring, or as input data for other scientific fields, such as the definition of biophysical models. In all cases, desired features include accuracy, large-area applicability, and rapid generation.

The earlier efforts to produce large area land use were manually undertaken during decades by geographers and ecologists, using field campaigns and the compilation of a variety of expert literature, such as atlases and maps at different scales [CSAEOS-NRC, 2008]. Later, with the revolutionary arrival of satellite data, characterization of land use and land use change patterns has been made possible, initially at coarse resolutions, ranging from 1° to 500m. With the new availability of Landsat imagery, now global land cover mapping at 30m has been made possible with results such as China's GlobeLand global land cover classification [Chen et al., 2015], as it is further detailed in Sec.3.2.2.

Still, large-area mapping initiatives to the present day have often lasted several years, distributing work among many workgroups, applying different working methods, and requiring coordination between a variety of institutions and workforces. With the arrival of the new generation sensor datasets, such as those of the twin-satellite constellation Sentinel-2, unprecedented data volumes will require new processing methodologies in order to cope with the associated computation and memory requirements. An operational methodology will require the elimination of traditional approaches based on manual operator-dependent tasks, having automation at its core, and thus reducing time delays, associated costs, and subjective interpretation.

Also, this new operational methodology will need to be able to adapt to the various complexities of satellite remote sensing data. The problem of automatic map generation over large areas defines some main necessities: firstly, flexibility regarding the presence of missing information related to clouds and cloud shadows; secondly, adaptation to spectral and spatial heterogeneity; thirdly, flexibility to be able to incorporate and exploit new data; in the fourth place, ability to adapt to the existence of temporal irregularities; and in the fifth place, adaptation to the absence of ideal input data. To this aim, a variety of well validated data mining and pattern recognition methods exists in the literature for the purpose of classification; however, these methods need to be compliant with the specific particularities of the new remote sensing data, and with the specific need of achieving automation.

3.2 Background

A variety of attempts at the generation of large area land cover classification exists dating back from the 1980s, mostly at coarse spatial resolutions. For example, some of the earliest global projects include the Mathews Global Vegetation Land Use (1983) and the Willson and Henderson-

Sellers Global Land Cover (1983), with resolutions of 1°, as well as the DeFries-Townshend Global Land Cover (1995), at 10km resolution. The GLCC (IGBP DISCover) (1997), the UMD Land Cover (2000) and the GLC-2000 (2003) improved resolution to 1km, and the MODIS Land Cover (2003) and the Vegetation Continuous Fields (2003, 2008) evolved from 1km to 500m. Then, GLOBCOVER (2009) provided coverage at 300m [Giri et al., 2013]. At a smaller scale, the European CORINE Land Cover (1990, 2000, 2006, 2012) was a large area attempt covering European countries, with a resolution of 250m and a minimum mapping unit of 25 Ha [Pérez-Hoyos et al., 2012] [EEA, 2007].

With a recent data policy change, open access has been granted to the LANDSAT data archive of the United States Geological Survey (USGS), and similar plans have been established for the coming LANDSAT missions [Hansen and Loveland, 2012]. Due to these factors, as well as to the characteristics of the programme itself, which embraces global acquisition and long-term archiving [Wulder et al., 2008], opportunities have been offered for long-term monitoring at large scales [Wulder et al., 2012], thus awakening an interest for large area land cover mapping at 30m resolution. Indeed, acknowledging a need for finer resolutions than the previously existent, the first projects to create a land cover database at 30m resolution have already been presented, such as the Chinese GlobeLand, using LANDSAT imagery in combination with MODIS, HJ-1 and FY3A data) [NGCC, 2010].

Indeed, improved land cover products can nowadays be generated, which were not possible some years ago. Technical advances have much shaped current-day possibilities with improved spatial and temporal resolutions, spectral coverages, calibration and geolocation accuracies. These technical advances, with the free open access to Landsat data and that to be offered by the coming satellites such as Sentinel-2, make the production of land cover maps with better resolutions and quality possible.

Also, much research has been conducted in the field of digital analysis for land cover mapping, although the application of these techniques when large areas are concerned is not straight-forward. The challenge remains for the development of methods capable of processing large areas, and able to manage the problems of variability, data volume, and missing information. The next two sections address land cover mapping methods, and land cover mapping methods for large areas, respectively, showing the urgent need for the development of processing strategies and methodologies for land cover mapping over great regions, in order to be able to respond to the growing need for timely information over large areas.

3.2.1 Land cover mapping methods

Land cover mapping is one of the most widely studied subjects using satellite data. However, most of the early studies based on high and medium resolution datasets were mostly methodological, testing pattern recognition and machine learning techniques over limited areas [Cihlar, 2000]. Indeed, two main trends can be identified, determined by their spatial and temporal resolutions: mapping with frequently obtained coarse resolution data (such as MERIS, AVHRR or MODIS GLI imagery), often permitting the use of compositing procedures in the absence of atmospheric effects, and mapping with relatively infrequent finer resolution data (for example Landsat TM datasets), where cloud contamination and seasonal effects have stronger impacts on the time series,

resulting in longer data gaps not allowing compositing approaches. Initially, large-area mapping with satellite imagery has been mostly undertaken with the former group, due to data availability and manageable computational demands [Cihlar, 2000].

Numerical techniques for image classification have a tradition that can be traced back to the early 70s [Cihlar, 2000]. Broadly, two main statistical classification techniques have been used in the literature for land cover mapping with satellite information, supervised and unsupervised classification. These approaches have different assumptions: in the former, a priori knowledge of all targeted informational categories is assumed and used for signature definition, while in the later classification does not use knowledge-based control and is entirely data driven, based on the statistics of the data distribution. A variety of methods have been developed over the years for both branches, all of them presenting advantages and limitations.

Some example methods for supervised classification include Support Vector Machines (SVM), Neural Networks, Decision Trees, Random Forests (RF), or Naive Bayes. The details of these methods have been repeatedly described in the literature and they are not reproduced here; a comprehensive analysis is given in [Richards, 2006]. An advantage of supervised methods is the fact that informational categories can be defined for precise targeting, the classification being controlled by user knowledge. As disadvantages, the need for the existence of reference data, the fact that defined classes may not match the naturally existing classes in the data, or may not be distinct with reference to spectral properties only, and also that classification can be misguided by inaccurate or incomplete training datasets.

Unsupervised classification methods include single pass and iterative clustering techniques, such as the K-Means and ISODATA approaches. Advantages of this type of classification include the fact that no prior knowledge on the data is required, and that user input error is minimized. As disadvantages, the fact that only limited control can be exerted on the menu of classes and their identities and that a posterior phase of assignment of the spectral clusters to targeted land cover classes is needed for land cover mapping. Also, spectrally homogeneous classes do not necessarily coincide with the targeted informational categories.

Application of the pattern recognition and machine learning techniques for land cover mapping has often been based on the exploitation of the different spectral reflectances of different land cover classes by use of spectral indices. To classify vegetated targets, the Normalised Difference Vegetation Index (NDVI) [Tucker, 1996] is the most widely used index in current literature using optical data, based on green vegetation high reflectance at near-infrared wavelengths and low reflectance at red wavelengths. Other vegetation indexes have been developed in an attempt to minimise soil background reflectance in low density areas, for example, or the impact of soil moisture. These include the Perpendicular Vegetation Index (PVI) [Richardson and Wiegand, 1977], the Soil Adjusted Vegetation Index (SAVI) [Huete, 1988], and its modifications, such as the Transformed SAVI (TSAVI) [Baret et al., 1989] and the Modified SAVI (MSAVI) [Qi et al., 1994]. A detailed study further investigating both broadband and hyperspectral vegetation indexes can be found in Broge and Leblanc [Broge and Leblanc, 2001]. Also, work in the literature has evolved from one-image analysis to time-series analysis, allowing the exploitation of a further data dimension.

Much literature can be found on specific crop and forest classification, both groups presenting distinct phenological evolution patterns, and generally targeting distinction between determinate subclasses in defined geographic areas. In general these approaches need to be defined based on expert knowledge on the targeted classes. In [Doraiswamy et al., 2007], MODIS time series are applied to classify corn and soybean crops by use of a decision tree approach based on NDVI values to separate crop and non-crop areas, and the application of second derivatives between two fixed days of the year. The authors report a large north-south variation in the development, which is attributed to different planting dates; the method is based on expert knowledge, as a period of 11 days of the cycle are targeted to distinguish between the two types of crops. In [Zhu et al., 2008], maximum likelihood supervised classification is applied on a single Landsat 7 image, and assessed against the same procedure using two images corresponding to the leaf fall season in order to demonstrate a reduction in the confusion between evergreen and deciduous forest and an overall increase in accuracy. Again, expert knowledge on the phenological response of the target classes determines the selection of input data.

[Wang and Tenhunen, 2004] applied not only NDVI time series values, but also NDVI-derived metrics, such as mean, maximum, amplitude, thresholds, total length of the growing season, fraction of growing season during green-up, rate of green-up, rate of senescence and others, for vegetation mapping in the North Eastern China Transect. The authors offered a comparative study between supervised classification with minimum distance and unsupervised classification with k-means with different combinations of these metrics, the study results seeming to support unsupervised classification on NDVI profile metrics only, although the need for accurate atmospheric and geometric correction was acknowledged. In [Itzerott and Kaden, 2006], crop classification in the Havel river catchment, Germany, is undertaken by analysis of the specific NDVI curves for each target crop, and comparison of supervised classification with maximum likelihood classification, superior in accuracy to a hierarchical classification including an iterative application of the parallelepiped method with final decision based on majority.

More recently, [Cheema and Bastiaanssen, 2010] offer a land use classification based on phenology information by performing unsupervised clustering and applying expert knowledge and ground information to label classes, for further water management analysis. In [Julien et al., 2011], classification of agricultural areas into crop types is undertaken on a one-year Landsat 5 time series in the Barrax area, Spain, with the Yearly Land Cover Dynamics approach (YLCD), based on annual behaviours of NDVI and land surface temperatures (LST). The YLCD parameters are extracted by plotting the LST and NDVI data for all dates for a given pixel, and applying linear regression to obtain three main descriptors, and then supervised maximum likelihood classification is applied. Good results are reported for distinction of thirteen targeted classes, when distinction between irrigated and non-irrigated areas is introduced, and cereals aggregated in a single crop due to similarities in annual NDVI and land surface temperature evolutions.

Also, different methods have been proposed for the detection of phenological events, mostly always following noise and missing data processing algorithms [Cong et al., 2012] as previously described. [White et al., 2009] offered a comparison of ten start-of-spring determination methods with AVHRR data. [Zhang et al., 2003] applied logistic function fitting to MODIS NDVI data in order to represent four transition dates determining annual phenology (green-up, maturity, senes-

cence and dormancy), discovering a green-up onset progressively moving north in north-eastern United States, and the contrary effect for senescence onset. The same results were confirmed by [Li et al., 2010] when south-eastern regions were included in the analysis for both NDVI and EVI series. [Cong et al., 2012] reported spring green-up onset north of 30°N China being delayed towards high latitudes and with an increase of elevation by previously applying and comparing different noise-reduction techniques. The authors identify correlation with temperature and precipitation, but widely varying across regions and vegetation types.

The reviewed literature on phenological studies and their application for land cover analysis reveals the potential of using time series of vegetation indexes like NDVI for map generation of land use. However, several weaknesses can be identified. Firstly, many studies are largely geographically dependent and not scalable to large areas, that is to say, tailored to spectral behaviours studied for determinate classes in determinate regions, and thus not affected by the variability that is to be encountered when large areas are considered. Secondly, much research is methodological in nature, based on the classification of user-determined mask-delimited areas (such as for example particular agricultural areas, or several selected crops, or broad types of forest) which requires a priori knowledge on the existent landcover. Thirdly, missing and noise-affected data are a common property of satellite data that is sometimes overlooked, for example by selection of cloud-free acquisitions. This approach simplifies the problem but results in the loss of cloud-free observations surrounded by clouds. Finally, the generation of land cover maps has frequently been accomplished by systems heavily dependent on user expertise. With the coming large volume datasets of multispectral multitemporal information, and when considering large areas, the user-based delineation of interest regions or the elimination of noise-corrupted data become too costly, and automation becomes a priority.

3.2.2 Large area land cover mapping methods

Mapping of land cover over large areas is considerably more difficult than mapping smaller regions [Ban et al., 2015]. As stated by the authors, for small areas, data collection and processing is easier, as a single-date image can possibly cover the target region; manual processing is simpler, with different degrees of operator involvement being possible; validation and refinement of results is also easier. However, when increasing the target region to a regional or global scale, a diversity of challenges arises, as described in Sec.1. As a result, and despite both the widely acknowledged need for timely and accurate land cover mapping and the large amount of existent research on land cover mapping, as shown in Sec.3.2.1, only a handful of land cover maps have been produced at a global scale [Ban et al., 2015].

Projects aiming at producing land cover maps at regional and global scales have traditionally lasted several years, distributing work among numerous work groups, and involving a variety of institutions. These delivery times are far from operational (with delivery in a matter of hours) for a variety of reasons, including the inherent difficulty of the task, the amount of image analyst work in the classification process, and the need for work coordination among involved institutions. For example, CORINE Land Cover was generated by operator photointerpretation, with more than 30 teams involved [Pérez-Hoyos et al., 2012], and delivery times of 10, 4 and 1.5 years for three of its products CLC1990, CLC2000, and CLC2006, respectively [EEA, 2007]. Another instance of coordinated production is GLC2000, with a delivery time of 4 years, and

involving 30 research groups coordinated by the European Commission's Joint Research Centre (JRC) [Bartholomé and Belward, 2005]. Tab.3.1 offers a timeline of main global land cover (GLC) products by reference year and public release year [Grekousis et al., 2015].

Reference year	CLC product	Public release date
1993	GLCC 2.0	2000
	ISLCP II	1999
	UMD	1998
2000	GlobLand30	2014
	CCI-LC	2014
	GLC2000	2004
2001	CLC 250	2014
	MODIS Yearly Collection 5.1	2013
2003	GLCMNO V1	2008
2005	CCI-LC	2014
	GeoWiki	2014
	GlobCover v2.2	2008
2008	GLCMNO V2	2013
2009	GlobCover	2013
2010	GlobLand30	2014
	FROM-GLC	2013/2014
	CCI-LC	2014
	GLC250mCN	2014
2014	GLC SHARE	2014

Table 3.1: Timeline of global land cover products by reference year with indication of their public release year. From [Grekousis et al., 2015].

Methods and data sources employed by current large area land cover projects are of diverse nature in terms of not only delivery times, as shown in Tab.3.1, but also of output resolution, employed sensor types, working methods, and characteristics and number of targeted land cover types. Tab.3.2 summarizes the characteristics of the main ten recent GLC products, ordered by output resolution, and indicating name, producer, source imagery, output resolution, number of classes and classification system, and stated accuracy. The features and methods applied by these projects are further described in the next paragraphs.

The GLC 2000 (Global Land Cover) product provides land cover at 1km resolution with 23 classes using the Food and Agriculture Organization of the United Nations Land Cover Classification System (FAO UN LCCS) with a stated accuracy of 68.6%. It was released in 2004 using SPOT-4 VEGETATION daily data acquired between November 1999 and December 2000. Different techniques were applied for its production depending on the partners [Pérez-Hoyos et al., 2012]. It involved 30 research groups coordinated by the European Commission's Joint Research Centre [Bartholomé and Belward, 2005], and used a "regionally tuned" approach where the lead scientists in each region took responsibility for the choice and implementation of image classification methods. In general, unsupervised classification was preferred, and labelling was based on existing vegetation maps, NDVI statistics and visual analysis.

GLC SHARE (Global Land Cover-SHARE) was published by FAO UN in 2014 as the result of a harmonization effort of already existing LC information into 11 FAO LCCS LC classes at

Name	Producer	Sensor	Resolution	Classes level I	Accuracy
GLC 2000	JRC	SPOT4 VGT	1km	FAO LCCS 22	68.6
GLC SHARE	FAO UN	Various	1km	FAO LCCS 11	80.2
GLCNMO V1/V2	ISCGM	MODIS	1km/500m	FAO LCCS 20	76.5/77.9
MODIS MCD12Q1 5.1	NASA	MODIS	500m	IGBP 17	71.6
GlobCover 2005 v2.2/2009	ESA	MERIS FR	300m	FAO LCCS 22	73.1/67.5
GeoWiki Hyb.1/Hyb.2	GeoWiki	MODIS, SPOT4, MERIS FR	300m	FAO LCCS 10	87.9/82.8
CCI-LC 2000/2005/2010	ESA	MERIS FR-RR, SPOT VGT	300m	FAO LCCS 22	In validation
GLC250mCN	CAS	MODIS	250m	11	74.93/75.17
FROM-GLC -/seg	CESSC	Landsat TM, ETM+	30m	9	63.69/64.42
GlobeLand30 2000/2010	UN NASG	Landsat TM, ETM7, HJ-1A/b/	30m	10	78.6/80.3

Table 3.2: Main global recent land cover products characteristics. From [Grekousis et al., 2015]. Notation: JRC, European Commission's Joint Research Center; FAO UN, Food and Agriculture Organization of the United Nations; ISCGM, International Steering Committee for Global Mapping; NASA, National Aeronautics and Space Administration; ESA, European Space Agency; CAS, Chinese Academy of Sciences; CESSC, Centre for Earth System Science China; NASG, National Administration of Surveying, Mapping and Geoinformation.

1km resolution with an accuracy of around 80% [Latham et al., 2014]. The classes are: artificial surfaces, cropland, grassland, tree covered areas, shrubs covered areas, mangroves, sparse vegetation, baresoil, snow and glaciers, water bodies, and finally herbaceous vegetation, aquatic or regularly flooded. The GLC products Globcover 2009, Modis 2010 and Cropland 2012 are used in areas where no regional or national products exist; when different databases are available, they are organised by assigning priority levels based on a variety of variables. The more detailed the dataset is considered, the highest the level assigned. Finally, a decision tree algorithm is used to do adjustment based on a set of criteria.

GLCNMO (Global Land Cover by National Mapping Organizations) by ISCGM (International Steering Committee for Global Mapping) accounts for two versions based on supervised classification and analyst work: V1, delivered in 2008, at 1km resolution, with 20 FAO LCCS classes and based on MODIS 2003 acquisitions, and V2, delivered in 2013, at 500m resolution and based on 2008 acquisitions. Stated accuracies are 76.5% and 77.9%. The map production process consisted of [Tateishi et al., 2014]: (a) use of existent GLC products GLCC, UMD LC (University of Mariland Land Cover), GLC2000, MOD12Q1, GLCNMO and GlobCover to generate potential reference maps; (b) analyst training data selection by division of the classes into subclasses based on NDVI and geographic locations (c) decision tree method with the commercial software See5 (retained after comparison with MLM with the ENVI software) (d) individual mapping of 6 classes by specially tailored processes (for example, wetlands limited to those larger than $1000km^2$; for each remaining wetland, best index and period for identification manually selected with Google Earth images; and selection of threshold values).

MCD12Q1 (MODIS Land Cover Type collection) used Terra and Aqua MODIS imagery and ancillary information in a supervised classification scheme [Friedl et al., 2010] to generate 5 classification schemes, the primary identifying 17 LC IGBP classes at 500 m resolution with a

yearly temporal coverage and a 71.6% overall accuracy. The base algorithm is a decision tree (C4.5), using boosting to generate ensemble classifications. Class conditional probabilities are derived for each class at each pixel, and postprocessing is applied to modified classification results for biases inherent to the decision tree algorithm caused by the characteristics of the training sample [Friedl et al., 2010]. Finally, classification results are stabilized across three-year windows.

The GlobCover project was developed by European Space Agency in a large partnership including the following institutions: JRC, FAO, EEA (European Environment Agency), UNEP (United Nations Environment Programme), GOFC-GOLD (Global Observation for Forest Cover and Land Dynamics) and IGBP (International Geosphere-Biosphere Programme) [Bontemps et al., 2010]. It was the first 300m GLC map, developed based on MERIS and ENVISAT data, applying supervised and unsupervised clustering on bi-monthly composites. It was delivered in 2005 for acquisitions between December 2004 and June 2006, and in 2013 for acquisitions between January and December 2009 [ESA-UCL, 2011] with accuracies of 73.1% and 67.5%, respectively. Classification was based on the following steps [Giri, 2012]: (a) what is called "stratification" (splitting) of the world into 22 areas, in order to allow for regional tuning of the extraction parameters; (b) per-pixel classification with supervised classification (to identify small-area LC classes) and unsupervised classification on the rest of pixels with a varying number of clusters; (c) for each cluster, computation of annual minimum and maximum of vegetation average; (d) cluster merging based on minimum and maximum of vegetation averages to obtain 50 to 70 classes; (e) class labelling based on GLC database, compiled from the GLC2000 product and some national and regional maps; (f) gap filling with reference land cover database; (g) correction of selected underestimated or inaccurately delineated classes (flooded forest, water bodies) directly from the reference land cover database and the SRTM Water Body Data.

GeoWiki provides a different approach to LC mapping, as it is based on volunteered geographic information validated by users around the world based on the previously described products GLC 2000, MODIS 2005 and GlobCover 2005. With this approach, two Hybrid maps have been created for 2005, by application of two methods based on geographically weighted regression. Stated accuracy was 82.8% for Hybrid map 2 [Grekousis et al., 2015].

The CCI-LC (Climate Change Initiative-Land Cover) product at 300m was released by ESA in 2014 for three acquisition periods, 1998-2002, 2003-2007 and 2008-2012, based on MERIS Full and Reduced Resolution and SPOT VEGETATION imagery. It is based on temporal syntheses obtained over a 7-day compositing period, applying world "stratification" (splitting) into a number of regions, and undertaking 4 main processing steps [Bontemps et al., 2015]: supervised classification, unsupervised classification, merging of classification based on rules, and post-classification edition. The typology contains 22 UN LCCS classes, three of which are identified mostly using an external dataset ("tree cover, flooded, saline water", "water bodies" and "snow and ice"). A weighted-area overall accuracy of 74.1% is estimated [Bontemps et al., 2015].

GLC250mCN (Global Land Cover 250 m China) is the only 250m spatial resolution GLC product, released by the Chinese Academy of Sciences (CAS) in 2010 and 2014. It is based on MODIS data for the periods 2000-2001 and 2009-2011 and has accuracies of 74.93% and 75.17%, respectively. The product contains 11 Level 1 classes (croplands, forests, grasslands,

shrublands, water bodies, barren lands, snow/ice) and applied random forests on FROM-GLC extracted training samples to obtain class labels and probabilities. An additional spatial-temporal consistency model was applied to improve results, and finally relabelling was undertaken based on prior knowledge or auxiliary data using the previous probabilities to estimate new labels [Wang et al., 2015].

As a result of the new free policy of Landsat imagery, China produced the first two 30m resolution GLC maps: FROM-GLC (Finer Resolution Observation and Monitoring of Global Land Cover) and GlobeLand30, released in 2013/14 and 2014, respectively, and with reference date 2010. The first one was produced with four per-pixel image classification approaches, with Landsat TM and ETM+ imagery acquired circa 2010, and efforts were made to select images from the "wettest" and "greenest" seasons [Ban et al., 2015]. A considerable mosaic effect was obtained due to the use of different acquisition dates across regions, and overall accuracies for Level-1 (10 classes: cloud, snow/ice, water, cropland, forest, grass, shrub, bareland, impervious) range from 53.88% to 64.89% with the best overall accuracy for Level-2 (29 classes) being 52.76% [Ban et al., 2015]. Several classes were also poorly classified [Gong et al., 2013], partly due to lack of exploitation of the temporal series due to the use of single-date Landsat images [Yu et al., 2013].

GlobeLand30 responds to an increasing awareness of the limitation of the previous product, offering a classification of 10 major classes based on Landsat TM, ETM7 and HJ-1A/b/ acquisitions. A pixel-object-knowledge based classification approach was developed for the production, based on pixel-wise classification using supervised classifiers and decision trees, segmentation applied at different scales (using the eCognition (v8) software, and different scale parameters depending on the landscape), object-based labelling (manual or automatic, depending on decision rules) and a final stage of manual verification and refinement based on web service technology [Chen et al., 2015]. Products were derived for two baseline years, 2000 and 2010, and released in 2014, with a claimed overall classification accuracy of 80.33% [Chen et al., 2015]. It must be noted that the knowledge-based refinement process is labour intensive and time consuming [Ban et al., 2015], and that accuracies need to be interpreted with caution, as the use of different reference data prevents comparability between projects. For example, Globeland30 and FROM-GLC were validated using different validation datasets [Ban et al., 2015].

After reviewing some of the main characteristics and constraints of the classification processes applied by global land cover projects, the difficulty of mapping land cover over large areas is made clear. Even when much research is devoted to land cover mapping techniques, mostly tested on simplified or reduced regions, the reviewed projects show how the application of such techniques becomes much more difficult when the problem extends to larger regions, and how the application of mapping techniques is simply not possible without addressing the particular challenges posed by large areas. The same is observable not only in GLC products, but also in regional and national products, which become non-operational. For example, CORINE Land Cover, coordinated by the European Environmental Agency (EEA), was generated by operator photointerpretation of one-date LANDSAT data (period 1986-1998 for CLC1990, and 1999-2001 for CLC2000) and two-date SPOT-4 imagery combined with IRS LISSS III (period 2005-2007 for CLC2006) [EEA, 2007] [Feranec et al., 2007], with more than 30 teams involved [Pérez-Hoyos et al., 2012], and delivery times of 10, 4 and 1.5 years for the three products [EEA, 2007], respectively. Other examples or

regional products include AFRICOVER by FAO in Africa, NALCMS (North American Land Cover Monitoring System) by NRCan/CCRS (Natural Resources Canada/Canadian Centre for Remote Sensing), USGS (United States Geological Survey) and INEGI (Instituto Nacional de Estadística y Geografía) in North America and South America 30 m of 2010 by USGS/EROS (Earth Resources Observation Systems).

Limitations are thus made clear, with large area attempts still not attaining full operational status, in the sense of being able to produce results in a matter of hours rather than months. Also, the minimization of the analyst role can be identified as an urgent objective, in order to increase the objectivity of classification results, and as a means to increase efficiency and reproducibility [Ban et al., 2015]. Full automation would undoubtedly have a positive impact on outcomes in terms of quality and delivery times. If the true dynamic nature of land cover is taken into account, with land cover types disappearing under the snow at high latitudes in the winter, and crop rotation practices transforming one type of crop into another in a matter of months, and fires and floods and other disasters producing fast changes in the landscape, then the need for a truly operational approach undoubtedly appears.

Obstacles for the operational production of land cover maps in the past include high image costs (which determined the operation on only a limited set of images), a common single-date approach (which did not profit from the richness of temporal series), coarser resolution of available imagery (with much spectral mixing), and particularly the difficulties associated to large areas, as described in Chap.1: the lack of processing approaches capable of processing large areas, managing variability, volume, and missing information. Nowadays, some of these obstacles have disappeared: the new free data policy of Landsat and the new Sentinel-2 projects constitutes an authentic milestone for the Remote Sensing community, and rich high resolution time series are now available, permitting to exploit spectral information in the temporal axis for image analysis. Only new adequate processing strategies and methodologies are needed in order to fully take advantage of all these new possibilities, so that users can generate good quality land cover maps tailored to their needed nomenclatures in operational times.

3.3 Proposed method

This work presents an approach for the estimation of large-area land cover with high spatial, spectral and temporal resolution Earth observation imagery. The next subsections describe its general characteristics, constituting the framework in which the main contributions of the work, presented in the following chapters, are to be understood. Fig.3.1 visually summarizes the proposed methodology, highlighting main contributions, and indicating corresponding chapters where these constributions are further described.

3.3.1 Description

The proposed approach responds not only to the need for new methodologies able to cope with the characteristics of the new high volume datasets that the new satellite missions will provide, but also to the acknowledgement of the dynamic nature of land cover, which leads towards the recognition of a need for truly operational approaches to land cover map generation. In particular,

it focuses on the problem of applicability on large areas, full automation, and the problems of missing information and spatial variability. It is intended as a contribution not only to research in the domain of land cover mapping, but also as a tool for users who are not specialists of the field but that rely on land cover maps as inputs for their own pursuit areas. The reader is referred to sections 1.2.2 and 1.3 for further explanations on the constraints and objectives of the work presented here.

The approach is generalisable to different landscapes and nomenclatures, and offers operational computation times with high volume and high resolution satellite datasets. Indeed, a combination of Landsat-7 and Landsat-5 time series covering the chosen Pyrenees study site have been used in order to simulate the soon-available Sentinel-2 twin-satellite data, which will feature higher temporal and spatial resolutions. Broadly, the method includes several processes, each of them further described in the chapters of this manuscript: preprocessing (Ch.2), feature extraction (Ch.3), missing information processing with local interpolation in the time series and linear temporal resampling with virtual datapoints (Ch.4), spatial stratified sampling for training set extraction (Ch.5), non-linear temporal sampling based on temperature data (Ch.6), and supervised model learning and classification (Ch.3). A scheme of the proposed methodology is given in Fig.3.1; novel contributions are highlighted in blue, and corresponding chapters are indicated.

As it can be seen in Fig.3.1, the general framework is similar to other classification schemes based on feature extraction, learning and classification, as follows. Given different multispectral time series, cloud, passage border and cloud shadow masks are obtained. Missing information is processed by applying local interpolation in the time series, and two types of temporal sampling are applied: linear temporal sampling, and non-linear temporal sampling based on ancillary temperature data. Feature vector construction is undertaken and spatial stratified sampling is applied for training sample set extraction based on ancillary topographic and bioclimatic information. Finally, model learning can be applied for classification.

However, the proposed approach introduces significant differences that allow its application on temporal series acquired over different areas and with different acquisition dates, that permit fully automatic and fast processing, and that manage the problems of data volume, data variability over large areas, and missing information. The novelty of the approach lies in the inclusion of three different methods that have not been explored in the literature so far to the best of the author's knowledge. These methods are: (a) the use of local interpolation methods in the time series in order to extract virtual points (in the temporal or temperature spaces) to reduce data volume and to allow for the application of single machine learning algorithms over large areas; (b) the application of stratified sampling methods for training set extraction based on topographic and bioclimatic information in order to allow for faster machine learning and classification, and for the intelligent selection of samples in order to enhance classification; (c) the use of non-linear temporal sampling based on temperature in order to extract significant features for land cover class characterization and to enhance classification. Blocking points apparent in the literature of land cover mapping over large areas such as operator interaction and subjectivity, long delivery times, and lack of full temporal series exploitation are in this way overcome.

An important consideration in this work is to attain reproducibility, that is to say, the yielding

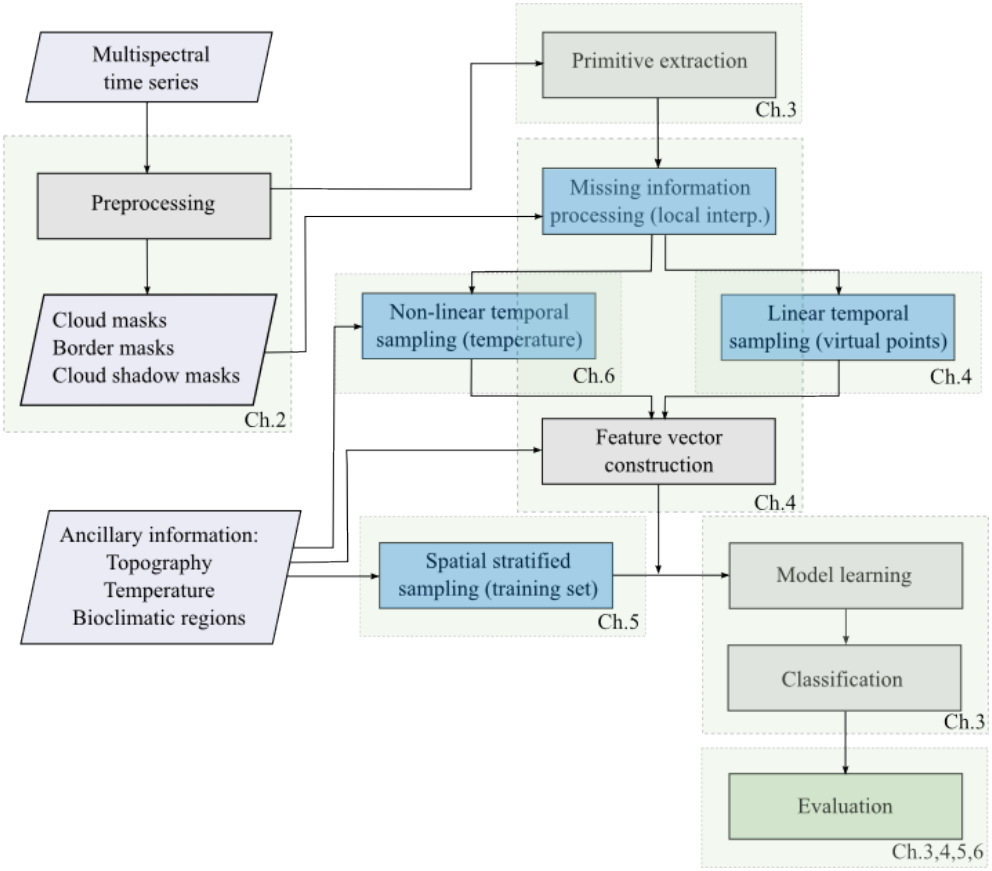


Figure 3.1: Proposed methodology. Corresponding chapters of this work are indicated, and novel contributions are shown in blue and presented in Chaps.4, 5 and 6.

of equal results by different operators given the same input data, procedure and targeted region, avoiding the introduction of subjectivity, as results become non-reproducible when the operator subjective input is distributed throughout the classification procedure. However, even if operator-independent procedures have been implemented, the operator cannot be totally excluded from the process because classifications are in themselves a human construct overlapping a non-natural scheme on the real world [Cihlar, 2000].

Thus, the following considerations have determined the characteristics of the proposed method, in order to comply with the specific necessities described in Sec.1.3:

(1) Purpose of the land cover map. Different applications have different needs in terms of targeted informational classes: the method is flexible to accommodate any number of classes and types of land cover information.

(2) Data. The availability and quality of data limit the accuracy that can be obtained: the method is flexible to work with any type of raster satellite time series, the spectral characteristics of the data determining feature extraction feasibilities. The proposed scheme has been developed

for use with Sentinel-2 imagery, using simulation with a combination of Landsat-5 and Landsat-7 datasets.

(3) Spatial resolution. Feature extraction possibilities are significantly determined by spatial resolution, which dictates operator choices such as the use of texture or shape features. Fig.3.2 illustrates the different information extraction feasibilities for different satellite sensor resolutions. Features used in this work are tailored to ETM and ETM+ sensor data with a spatial resolution of 30m, and targets applicability with Sentinel-2 data at 10m and 20m resolution. The approach remains flexible and permits the integration of other features when higher spatial resolutions are available. Indeed, multisensor information can be integrated with proper resampling and corregistration.

(4) Temporal resolution. Although targeting a temporal revisit of 5 days, data of different temporal resolutions can be applied. As the method can work with any user specified length of data, temporal strategies can manage data gaps of any length (with a possible decrease of accuracy for long gaps), and spectral evolution curves of any shape, including vegetation index series of more than one peak period.

(5) Spectral wavelengths. The similitude between six spectral bands of Landsat and Sentinel-2 has been exploited in this work. The rest of bands incorporated by Sentinel-2 can be easily integrated in the system as either further land cover features (for example the vegetation red-edge band) or applied in the preprocessing stage (such as the cirrus or aerosol detection bands).

(6) Generality. As described, the proposed approach has been developed as a processing chain for the coming Sentinel-2 datasets, aiming to establish a general framework applicable to other satellite data or to multisensor datasets duly corregistered and resampled. Landsat ETM and ETM+ data have been used to simulate the Sentinel-2 time series, thus determining operator choices described in (3), (4) and (5) related to spatial resolution, temporal resolution, and spectral wavelengths, respectively.

(7) Automation. The method is totally automated and generic. In order to be able to accommodate different types of data and targeted informational categories, it restrains the operator input to initial decisions subordinated to available data and desired output, such as for example the targeted land cover types, type of features to be used, or decisions on the parameters of the processes described above, such as number of timepoints used for temporal resampling with virtual points, number of training samples to be used, or study temporal period. In this way, the reproducibility and consistency of the method can be assured, and the impact of operator-related subjectivity and associated costs are minimised.

(8) Processing costs. The method ensures a timely production of land cover maps, so that information can be updated rapidly. It is designed for large areas and large-volume multitemporal datasets, and able to cope with the computation requirements, in terms of operational processing and memory management.

(9) Specific strategies. The particular difficulties of land cover mapping over large areas with satellite data, such as the spectral variability linked to spatial diversity in extended regions, the

presence of missing information (due to clouds, cloud shadows, snow, or passage borders) and temporal irregularities, the volume of time series and their processing costs, or the phenological (and thus spectral) variability of vegetation in relation to climate, have been targeted by specific strategies operating in both the spatial and temporal spaces, described in the contribution chapters of this work.

3.3.2 Time series feature extraction

In the context of machine learning, the extraction of features for retrieval and separation between targeted classes is a common processing step for the application of machine learning methods such as those discussed in Sec.3.2.1. The extraction of image features has been significantly explored in machine vision and pattern recognition, with objects or pixels represented by a vector of numerical extracted features, such as spectrum-derived values, or others based on texture, shape, or topology. Here, feature extraction is performed in the time series for a characterisation refecting phenological response in time. In particular, both linear and non-linear approaches to temporal resampling for feature extraction are undertaken, as further described in Chap.4 and Chap.6, in order to allow for an efficient land cover mapping when large areas are considered.

3.3.2.1 Sensor-dependent factors

Transformations derived from multispectral data in the form of indices are the most common form of extracted features to characterise land cover types in the field of remote sensing. However, sensor spatial resolution determines the applicability of texture, shape and topological features, and sensor spectral bandwidths establish the available properties that can be exploited in order to enhance class characterization and class separability. Depending on the available satellite imagery, a thermal band can be present to facilitate the detection of clouds, or a red edge band can be included to aid in the identification of vegetation types.

Comparably, the use of shape analysis, texture descriptors and topological features is determined by the spatial relationship between the targeted objects and the sensor resolution. At coarse scales, objects are lost as they get incorporated into neighbouring pixels, thus preventing appropriate extraction of contour, shape or topological information, which need appropriate delineation. Fig.3.2 illustrates the different degrees to which extraction of such information is feasible, showing the same area at resolutions of 1m, 10m and 30m.

Application of contour-based shape features (such as measures of compactness, eccentricity or solidity, or Fourier-based descriptors) has been common in the literature with good results with very high resolution sensors, as well as the use of region-based shape descriptors based on moments, as described in [Teague, 1980]. Due to the spatial resolution of the imagery used in this work (30m), and the targeted Sentinel-2 imagery (20m, although resolutions will range between 10m and 60m), this type of information was not applied in this work, although the proposed method can readily incorporate them if higher resolution imagery is used.

Analogously, the applicability of texture analysis is also greatly dependent on sensor spatial resolution and its ability to capture the spatial organization of spatially varying spectral values for the

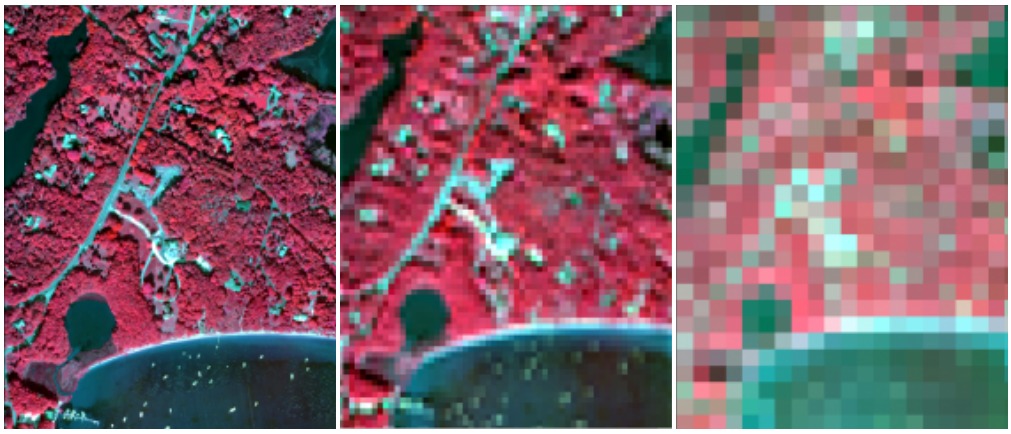


Figure 3.2: Different feature extraction possibilities are possible at different scales. Resolutions of 1m, 10m (Sentinel-2) and 30m (Landsat) are given for the same imaged region.

objects of interest. A wide variety of techniques in the research literature have been proposed to describe image texture. These include statistical methods such as the Local Binary Pattern (LBP) operator [Wang and He, 1990] or the Gray Level Coocurrence Matrix (GLCM) [Haralick et al., 1973], geometrical methods, model-based methods such as Markov random fields and fractals, and other signal processing techniques such as wavelet transforms. For example, different experiments have been conducted in the literature in order to include texture in LANDSAT-based classifications [Asner et al., 2002] [Berberoglu et al., 2007] [Ota et al., 2011] [Lu and Weng, 2005], by inclusion of Haralick features computed on grey-level co-occurrence matrices, such as measures of homogeneity, contrast, dissimilarity, entropy or correlation. Problems related to resolution where identified in [Berberoglu et al., 2007], and results indicate that the analysis of texture would be more beneficial when used on higher resolution imagery [Ota et al., 2011] [Asner et al., 2002] [Lu and Weng, 2005]. For this reason, texture analysis has not been applied here, although again the method can easily incorporate this type of information with higher spatial resolution data such as Sentinel-2 imagery.

3.3.2.2 Selected features

Feature vectors have been extracted by combined discretization of height-related information and temporally sampled extracted features based on LANDSAT multispectral information for a chosen period. Only 6 spectral bands have been considered; the thermal infra-red region (10.40 – 12.50μm, 60m resolution) has been discarded due to lack of correspondence with any of the Sentinel-2 bandwidths, as shown in Tab.2.3.

A digital elevation model being available, height and derived slope and aspect information have been computed and included as feature vector data. Slope has been calculated as gradient ∇f on the available digital elevation model (DEM) as in Eqs.3.1 and 3.2 and normalised, with $\frac{\partial f}{\partial x}$ being the gradient in the x direction and $\frac{\partial f}{\partial y}$ the gradient in the y direction:

$$\vec{\nabla} f = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \tag{3.1}$$

$$slope = \| \vec{\nabla} f \| = \sqrt{\frac{\partial f^2}{\partial x} + \frac{\partial f^2}{\partial y}} \tag{3.2}$$

Aspect has been calculated and normalised with respect to a North-South orientation, as in Eq.3.3.

$$aspect_{N-S} = \cos \left(\text{atan} \left(\frac{\frac{\partial f}{\partial y}}{\frac{\partial f}{\partial x}} \right) \right) \tag{3.3}$$

This responds to a vegetation dissymetry which is commonly encountered in north and south slopes due to different sun radiation regimes and resulting temperature conditions, as south-facing slopes receive more solar radiation for a given surface area than north-facing slopes. Chapter 6 provides further information on the impact of temperature on vegetation phenology and distribution. An illustration of the obtained aspect and slope information for the study area is given in Fig.3.3 and closer views of initial height information and derived aspect and slope data are shown in 3.4.

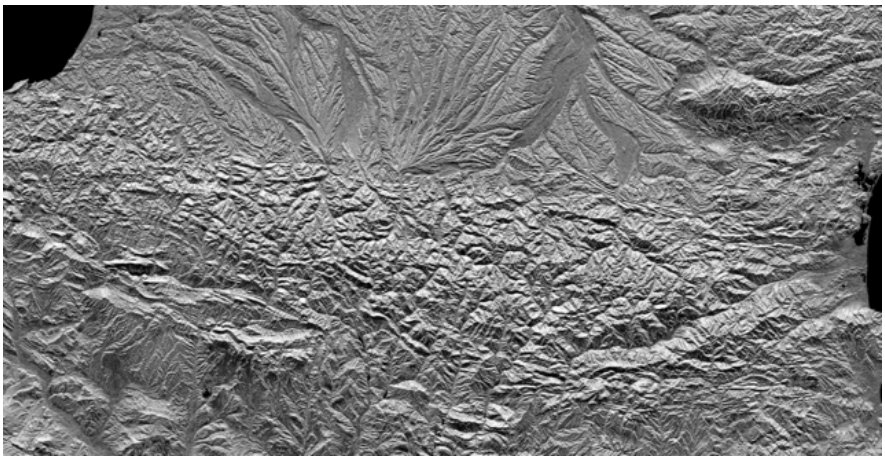
Spectral feature extraction operations make use of physical principles and the application of ratioing, as it constitutes an effective technique for selectively enhancing particular spectral features and thus particular targets. Also, the use of ratio is effective for suppressing topographic shadows due to the relative angles between land surface and incident solar radiation. Spectral indices have been mostly developed based on ratio and differencing operations, such as the well-known Normalised Difference Vegetation Index (NDVI) [Tucker, 1996] that has been applied here. It is given in Eq.3.4.

$$NDVI = \frac{NIR - R}{NIR + R} \tag{3.4}$$

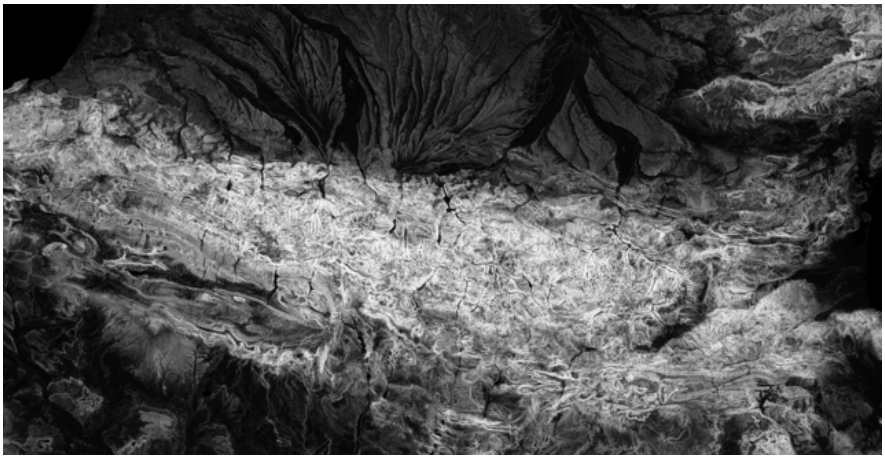
The application of spectral indices such as the NDVI allows the exploitation of biophysical features in order to differentiate land cover classes. The NDVI makes use of a characteristic high reflection peak in the near-infrared and an absorption in the red that correspond to photosynthetic vegetation. Red-light absorption is due to photosynthesis and near-infrared light reflection is related to leaf structure [Tucker, 1996]; in this way, the NDVI constitutes an indicator of vegetation photosynthetic activity.

Similarly, the Normalised Difference Water Index (NDWI) enhances the presence of water features and eliminates the presence of soil and terrestrial vegetation features in digital imagery. It makes use of the near-infrared and green band reflectances, as formulated by [McFeeters, 1996] and given in Eq.3.5. A variant of this index applying SWIR instead of green light was formulated by [Gao, 1996] as a vegetation index complementary to NDVI and sensitive to liquid water content of vegetation canopies.

$$NDWI = \frac{NIR - G}{NIR + G} \tag{3.5}$$



(a)



(b)

Figure 3.3: Aspect and slope information extracted from the SRTM digital model elevation available for the Pyrenees area. The aspect image (a) shows North-South orientations of the relief (in the range [0,1]; bright values correspond to North exposures). The slope image (b) reveals considerable slope differences (in the range [0,81.092]; bright values correspond to greater slopes) between mountainous regions and surrounding plains.

In the same way, the Normalized Difference Built-up Index (NDBI) has been applied in order to enhance built-up areas. Proposed by [Zha et al., 2003] and given in Eq.(3.6), it was initially applied on TM imagery for the mapping of urban land by using it in combination with extracted NDVI information.

$$NDBI = \frac{NIR - SWIR}{NIR + SWIR} \tag{3.6}$$

Finally, a normalised Brightness Index (IB) has been computed, following [Nicoloyanni, 1990] as given in Eq.(3.7). Two normalised ratios in the visible (NDGR and NDBR) have also been computed, in order to include information from all spectral bands, as given in Eq.(3.8) and Eq.(3.9).

(a) (b) (c)

Figure 3.4: Detail of SRTM digital model elevation available for the Pyrenees area (a), and extracted aspect (b) and slope information (c).

$$IB = \sqrt{\frac{(R^2) + (G^2) + (NIR^2)}{3}} \tag{3.7}$$

$$NDGR = \frac{R - G}{R + G} \tag{3.8}$$

$$NDBR = \frac{R - B}{R + B} \tag{3.9}$$

The use of primitives with SVM classification has advantages long exploited by the remote sensing community; indeed, the previously enumerated primitives enhance different identifying aspects of the targeted classes, such as photosynthetic activity, water content, build-up characteristics, brightness or greenness. For a linear kernel SVM, margins are built based on linear combinations of the parameters, so the inclusion of linear combinations of the existent parameters do not bring new information. Contrarily, the introduction of non-linear combinations does provide new degrees of freedom, which is favorable to the improvement of classification accuracies. Numerous other features can be extracted depending on sensor characteristics, as explained, and a variety of spectral indices exists, including for example the Ferrous Minerals ratio or the Iron Oxide ratio [Segal, 1982], and numerous other soil, vegetation, water and built-up indexes, which the user can set for use depending on the targeted informational classes. A well-known transformation is that of the Tasseled Cap function, which performs data reduction and extraction of three uncorrelated indices called "brightness", "greenness" and "wetness" that have popularly been applied in the literature. Originally developed for Landsat MSS data [Kauth and Thomas, 1976], it has been extended to Landsat TM [Crist and Cicone, 1984] and ETM+ [Huang et al., 2002b] and other sensors. In preparation for Sentinel-2 data series, and the transformation being sensor-dependent, these indices have not been applied here, but remain an option for land cover mapping with Landsat data.

3.3.3 Supervised learning

Due to its good performance among the machine learning algorithms for high-dimensionality dataset classification in terms of accuracy, stability and computation time costs

[Melgani and Bruzzone, 2004], land cover classification has been undertaken by application of support vector machines (SVMs). With most supervised approaches, dimensionality generally penalizes performance. As the feature dimension increases, the number of training samples needs to increase exponentially, so that the training sample set is representative of the input space dimension [Carrão et al., 2008]. With a fixed number of training samples, predictive power firstly grows and then declines as dimensionality increases, which is known as the Hughes effect. The solution of increasing the number of samples can overcome this effect, although incurring in a risk of overfitting [Carrão et al., 2008]. With SVM classifiers, performances are mostly unaffected by oversized dimensions of input feature spaces. In order to work on high dimensional outputs, and given the good performance of SVM on a wide variety of previous land cover classification studies, an SVM approach has been selected.

The mathematical formulation of SVMs can be found in multiple publications, such as [Vapnik, 1998], [Burges, 1998] and [Huang et al., 2002a]; the following description is based on [Lennon et al., 2002]. Broadly, the algorithm solves a two-class classification problem by looking for the optimal separating hyperplane between two classes given N training samples $(y_i, \vec{x_i})$ with y_i being a class label, $\vec{x_i} \in \Re^n$ being a feature vector, n being the dimension of the feature space, and $i = 1...N$. The classifier is represented by the function $f(\vec{x}, \alpha) : \vec{x} \rightarrow y$ with α being its parameters. The SVM finds the optimum separating hyperplane so that samples with labels +1 and -1 are located on each side, and it maximizes the distance (margin) of the closest vectors for each class (support vectors) to the hyperplane. A hyperplane can be written as satisfying $\vec{w} \cdot \vec{x} - b = 0$,where $\vec{w}$ is its normal vector and $\cdot$ the dot product. Vectors not on the hyperplane satisfy either $\vec{w} \cdot \vec{x} - b > 0$ or $\vec{w} \cdot \vec{x} - b < 0$, the classifier thus becoming $f(\vec{x}; \vec{w}, b) = sgn(\vec{w} \cdot \vec{x} + b)$, and the support vectors lying on $\vec{w} \cdot \vec{x} + b = \pm 1$.

If the training samples are linearly separable, the following optimization problem needs to be solved:

$$\begin{aligned} min \left\{ \frac{1}{2} \parallel \vec{w} \parallel^2 \right\} \\ y_i(\vec{w} \cdot \vec{x_i} + b) \pm 1, i = 1...N \end{aligned} \tag{3.10}$$

If the training samples are not linearly separable, error variables ξ_i and a regularization parameter C are introduced:

$$\begin{aligned} min \left\{ \frac{1}{2} \parallel \vec{w} \parallel^2 \right\} + C \sum_{i=1}^{N} \xi_i \\ y_i(\vec{w} \cdot \vec{x_i} + b) \geq 1 - \xi_i, i = 1...N, \xi_i \geq 0 \end{aligned} \tag{3.11}$$

Both cases are illustrated in Fig.3.5. The optimisation problem can be solved with Lagrange multipliers λ_i, which are non-zero for support vectors, and thus becoming:

$$\begin{aligned} max\{\sum_{i=1}^{N} \lambda_i - \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_i \lambda_j y_i y_j (\vec{x_i} \cdot \vec{x_j})\} \\ C \geq \lambda_i \geq 0, i = 1...N \\ \sum_{i=1}^{N} \lambda_i y_i = 0, i = 1...N \end{aligned} \tag{3.12}$$

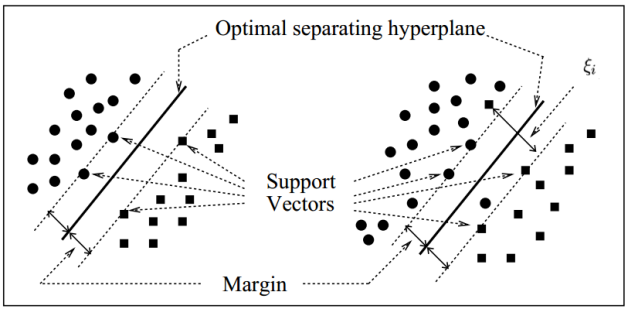


Figure 3.5: SVM optimal separating hyperplane and related concepts. Left: linearly separable case. Right: non-linearly separable case. From [Lennon et al., 2002].

Non-linear decision surfaces can also be used with SVMs by embedding the data in a higher dimension space where they are linearly separable. Similarly, m-class problems when $m > 2$ can also be solved by SVMs by applying one of these two approaches: (a) "one against all", when m SVMs are applied, one for each class against all the others, and the vote from the classifier generating the highest value from its decision function is selected; and (b) "one against one", when $m(m-1)/2$ SVMs are applied on each pair of classes and a voting scheme retains the most voted label. The former approach has been applied here; it requires the application of more support vectors, but it allows for a reduction in computation time, as the complexity of the approach depends on training sample number [Lennon et al., 2002].

Supervised classification with SVMs has been profusely used during the last years for multispectral and hyperspectral remote sensing data classification. Research has focused on a multiplicity of topics, ranging from comparative studies to other pattern recognition approaches [Melgani and Bruzzone, 2004] [Huang et al., 2002a] and investigation on kernel optimization [Mercier and Lennon, 2003] to different aspects of land cover mapping. [Knorn et al., 2009] applied SVMs on selected LANDSAT images for binary forest/non-forest classification by masking of clouds and cloud shadows and chain classification; the same binary approach and processing of clouds was chosen by [Kuemmerle et al., 2009] for forest cover change mapping in time.

Also working on selected LANDSAT images, [Huang et al., 2008] used SVMs for forest/non-forest classification by masking of non-vegetated dark objects such as water, topographic shadow, and dark soils. [Huang et al., 2002a], [Dixon and Candade, 2008] and [Candade and Dixon, 2004] performed a comparative study of SVMs against other classifiers on limited ground truth data, verifying their strengths in terms of computing times and dimensionality. [Cao et al., 2009] used SVM classification as part of an iterative method for the detection of burnt scars in grasslands identifying a constraint by clouds [Mountrakis et al., 2011]. A comprehensive review on the use of SVMs in remote sensing is provided in [Mountrakis et al., 2011].

The study of the literature on land cover classification using remote sensing data and supervised classification reveals several common pitfalls, such as a recurrent use of image date selection by analysis, or of classification region reduction through the use of operator-selected masks (in order to avoid cloud, cloud shadow or other types of impact on the imagery), a common application of binary classification for the evaluation of automatic techniques, training on great numbers of pixels,

validation on reduced numbers of pixels, or generally a high dependence on operator manual work.

The presented approach overcomes these limitations: no date selection is performed, as the method makes use of all available information for each acquisition and no valid information is discarded; no operator-selected masks are used, and classification is applied on whole areas; land cover mapping is done with a reference ground truth of 22 classes, which may or may not be present on the considered regions; training is undertaken with a very small number of pixels (under 0.07% of the region pixels with 4 tiles, for example); validation is performed over large numbers of pixels (756023 for tile L930463, 2593501 when 4 tiles are considered and 3573004 when 8 tiles are studied, see Tab.2.7 for details), and the method does not require operator interaction or subjective interpretation.

It must be pointed out, SVM classification has been applied for the evaluation of the application of the techniques described in this section, the focus not lying on the optimization of the SVM application itself, but on the effectiveness of the methodology for the processing of the new kind of satellite high resolution data that will soon be available, and other classification approaches can indeed be employed. Indeed, recent research studies seem to indicate that the performance of Random Forest (RF) classifiers may be superior to that of SVMs for land cover classification, although the characteristics of the datasets used for comparison (single or multitemporal, outlier presence, imbalance of training or validation classes, distribution of data, present variability) and the choice of classifier parameters may have an impact on results. For example, [Pal, 2012] compares the performance of SVMs and RF on a multispectral (one image of 307x330) and a hyperspectral dataset (one image of 515x515), concluding the latter is more advantageous in terms of ease of use and computation cost, with similar accuracy results depending on classifier variables, with training on 400 pixels per class on the multispectral image (7 classes) and 100 pixels per class on the hyperspectral one (7 classes). With a larger scope, in [Inglada et al., 2015] two strategies for crop type map production (SVM and Random Forests) are compared over 12 sites around the world, indicating a good performance of RF for most sites, and in particular improvements in terms of main class for most tested sites in comparison with SVMs. The technique has not been applied here, and it remains as a future research topic to investigate.

3.3.4 Class characterisation

Class characterisation for model learning has been based on temporally sampled extracted features as described in Sec.3.3.2.2, and on the inclusion of non-spectral information in a variety of approaches, as explained in Chaps. 4, 5 and 6. For illustration purposes, Figs.3.6 to 3.11 present a total of 44 graphs, two for each of the targeted land cover types, based on the available temporal series for one of the applied features. In the left column, average available values of NDVI for each of the classes are presented, together with the associated variance for each of the satellite acquisition dates. As the crop year October 2009 to September 2010 has been considered, the day-of-year range is [276:636], referenced to 1st January 2009. NDVI values are shown when information is available on an acquisition date for any of the selected samples of a class; for illustrative purposes, zero values have been used to indicate a total absence of valid information. Variances vary significantly from one class to another, and from one acquisition date to another, both as a result of the natural variance of the samples of the classes across locations, but also due to the existence of different types of noise and to the varying robustness of the characterization, depending on the

amount of valid information. Available source data is illustrated here; class-characterising profiles will be obtained after the application of the proposed methodology for the processing of missing information due to clouds, cloud shadows and snow, as it will be shown in Chap.4.

In the right column, corresponding levels of missing information are presented, in percentages of pixels. A total lack of valid information in the left column corresponds to a maximum amount of missing information (100%) in the right column. In other cases, even when approaching 100% of missing data and thus based on a small number of pixels relative to the amount of pixels of the class in the reference ground truth, corresponding average values and variances can be computed. For the interpretation of the graphs, it must be taken into account that, in the amount indicated by the right column corresponding timepoint, values corresponding to areas labelled as containing missing information have not taken into account for the computation of average values. Thus, greatly variable levels of information constitute the basis for the generated profiles of average values and variances shown in the left column. The application of the proposed method will allow for the recovery of missing information and the exploitation of all valid data, so that no valid source is disregarded no matter how small the percentage for a given acquisition. Obtained class-characterising temporal profiles are given in the next chapter.

Although with considerable differences in variance (as shown on the left column graphs), and with high levels of corresponding missing information (as shown on the right column graphs), general trends can be identified for each of the 22 targeted land cover classes. Crop characterization is given in rows 1 to 4 in Fig.3.6, 1 in Fig.3.7, and 2 and 3 in Fig.3.9. Two main phenological curve trends can be observed. On the one hand, a trend typical of winter crops, with curves increasing over 0.5 from day of year (doy) 400 (approximately beginning of February), peaking at around 500 (around middle of May) and descending below 0.5 around doy 540 (towards the end of June). These graphs correspond to the land cover classes "wheat", "barley", "rape" and a mixed category of "other winter cereals".

On the other hand, rows 2 in Fig.3.6 and 1 in Fig.3.7 exhibit a trend typical of summer crops, with curves increasing over 0.5 around doy 540 (towards the end of June), peaking at around doy 575 (end of July), and descending again at approximately doy 625 (middle of September). These graphs correspond to the classes "corn" and "sunflower". Another related land cover class is represented in graph 3 in Fig.3.9, corresponding to "other summer crops". Its trend is similar, although with an extended peak period comprising the spring and summer months.

Forest characterization is given in graphs 1 to 4 in Fig.3.10 and graph 1 in Fig.3.11 corresponding to "deciduous broadleaf", "persistent broadleaf", "deciduous needleleaf", "persistent needleleaf" and "mixed forest". Unavailability of valid information for a considerable amount of timepoints is visible for the classes "persistent broadleaf", "deciduous needleleaf" and "mixed forest", for which a significant lower number of samples is available (particularly in the case of "persistent broadleaf", with explains the levels of noise associated with the land cover class) in relation to the more common "deciduous broadleaf" and "persistent needleleaf". For these classes, a common high NDVI trend is visible during the summer months from June to the end of September.

Graphs 1 in Fig.3.8 and 4 in Fig.3.9 show the land cover classes "vineyards" and "fruit/nut/olive

trees", respectively, having in common fairly similar low variation NDVI time series with values of around 0.6. Graph 2 in Fig.3.7 shows "woody plants", with a phenological curve approaching that of summer crops. Graphs 3 and 4 in Fig.3.7 correspond to "permanent herbaceous crops", with high missing information ratios, and "temporary meadows and fallows", showing a lower amplitude NDVI curve. Also, graph 2 in Fig.3.8, corresponding to "water bodies", presents an average NDVI curve with great variances and includes negative values. The observed variability may be related to the characteristics of the various types of water bodies, such as lakes and oceans, and to the fact that the turbidity of water can cause dramatic reflectance changes.

Fig.3.8 includes graph 3 for "impervious surfaces and barren land", characterised by constant low NDVI values at around 0.2 thoroughout the crop year. Finally, graphs 4 in Fig.3.8, 1 in Fig.3.9 and 2 in Fig.3.11 show the land cover classes "herbaceous moorlands", "shrub-populated moorlands" and "rocky moorlands". Despite missing information, the curves seem to present a loosely similar phenological curve reaching maximum values between 550 to 610 (June and August), and with lower overall values for the "rocky moorlands".

As explained, for each of the studied land cover classes, characterization is based on the temporal analysis of the extracted features described in this chapter, as well as the inclusion of non-spectral information in a variety of approaches. Only NDVI has been illustrated here, and the presented average NDVI curves show a great variability. The accuracy of these values depends not only on the number of available acquisitions, but more importantly on the proportional number of acquisitions having non-missing and non-corrupted information for the class area regions. Processing of missing data is undertaken in the next chapter, in order to minimise the impact of missing values in the characterisation of land cover classes.

3.4 Conclusions

The problem of automatic classification for large-area land cover mapping has been analysed, and a review on the topic of land cover mapping techniques and large area land cover mapping techniques has been given, giving an overview of current global projects, their classification approaches and automation efforts.

A method for the automatic production of land cover maps with high resolution time series has been proposed in preparation of the Sentinel-2 constellation time series, allowing for the production of results in operational times. It is based on the generation of temporally and spatially continuous HR time series, the exploitation of multitemporal and multispectral information, and the use of machine learning techniques for the creation of a final land cover map product. Based on a common land cover classification scheme, the proposed approach introduces significant differences that allow its application on temporal series acquired over different areas and with different acquisition dates, that permit fully automatic and fast processing, and that manage the problems of data volume, data variability over large areas, and missing information.

The novelty of the approach lies in the integration of three contributions: (a) the use of local interpolation techniques in the time series in order to extract virtual points (in the temporal or temperature spaces) to reduce data volume and to allow for the application of single machine

learning algorithms over large areas; (b) the application of stratified sampling methods for training set extraction based on topographic and bioclimatic information in order to allow for faster machine learning and classification, and for the intelligent selection of samples in order to enhance classification; (c) the use of non-linear temporal sampling based on temperature in order to extract significant features for land cover class characterization and to enhance classification. Identified blocking points in the literature of land cover mapping over large areas such as operator interaction and subjectivity, long delivery times, and lack of full temporal series exploitation can in this way be overcome, as it is demonstrated in the following chapters.

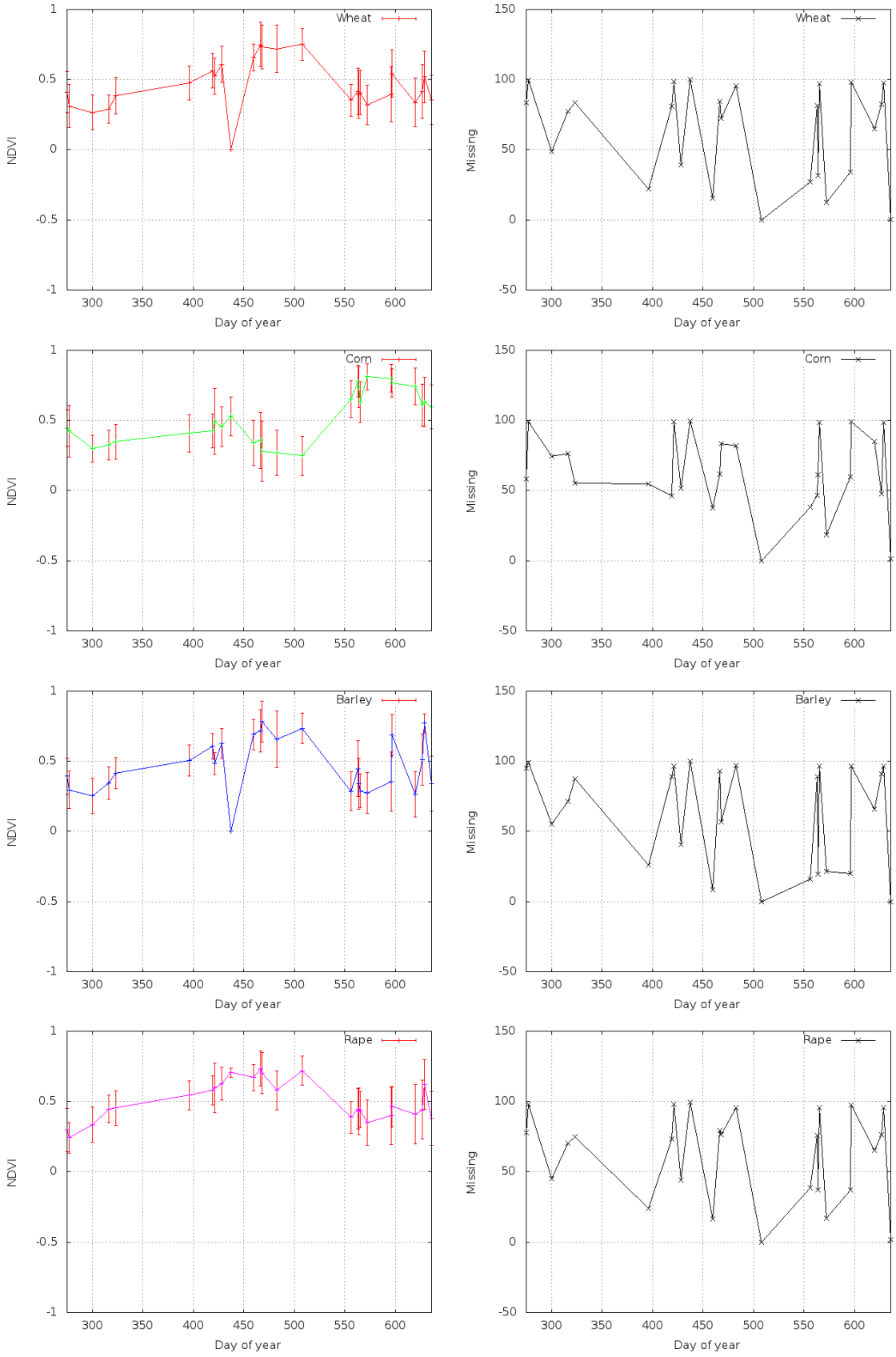


Figure 3.6: Example of available time series spectrum-derived information for the targeted land cover classes (left column). NDVI average values and variances for each class correspond to varying amounts of available valid information for each acquisition (Oct 2009 - Sep 2010, day of year referenced to Jan 2009). Zero values in the NDVI series indicate total absence of valid information. Corresponding percentages of missing information are also shown (right column).

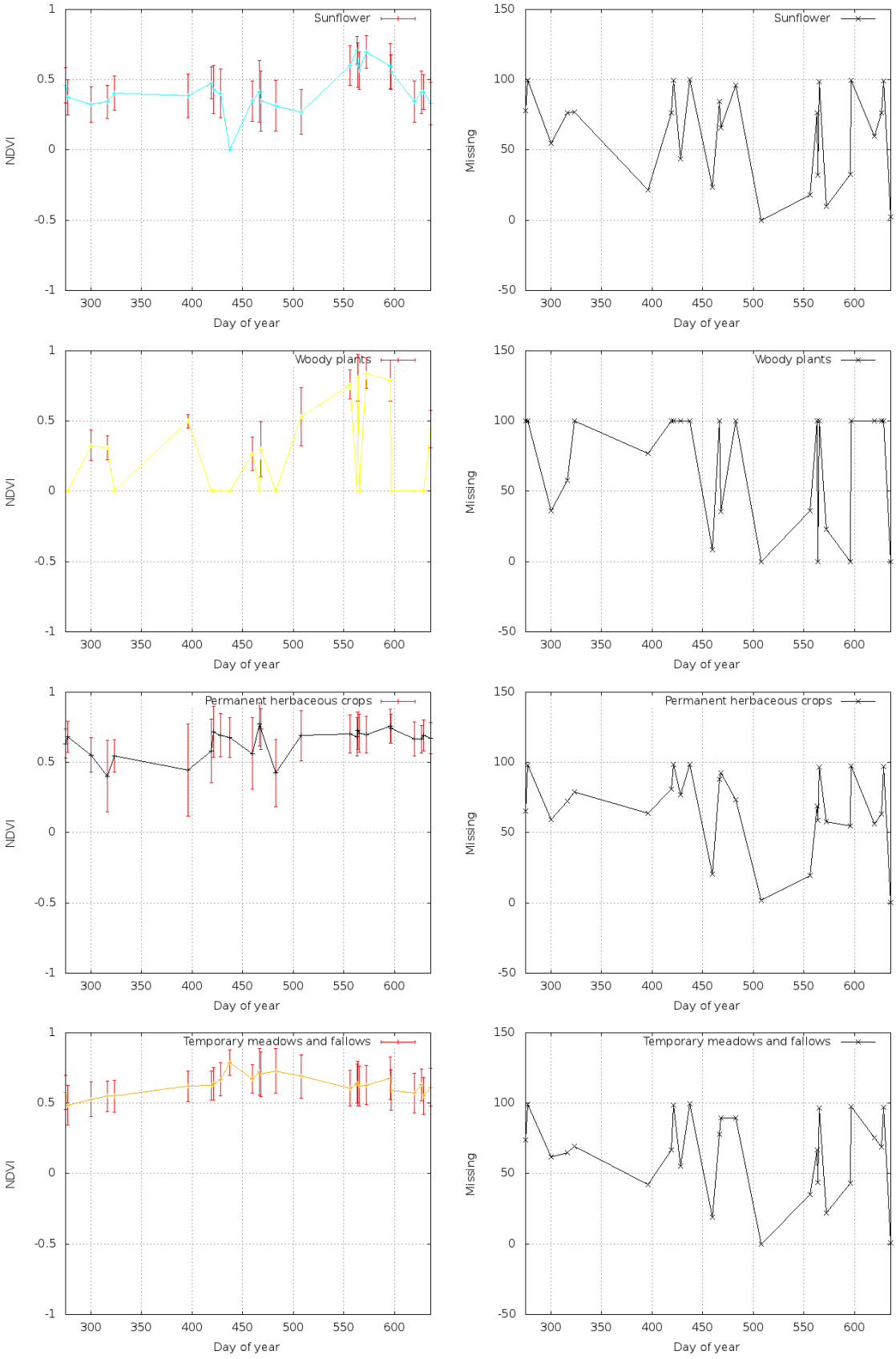


Figure 3.7: As in Fig.3.6.

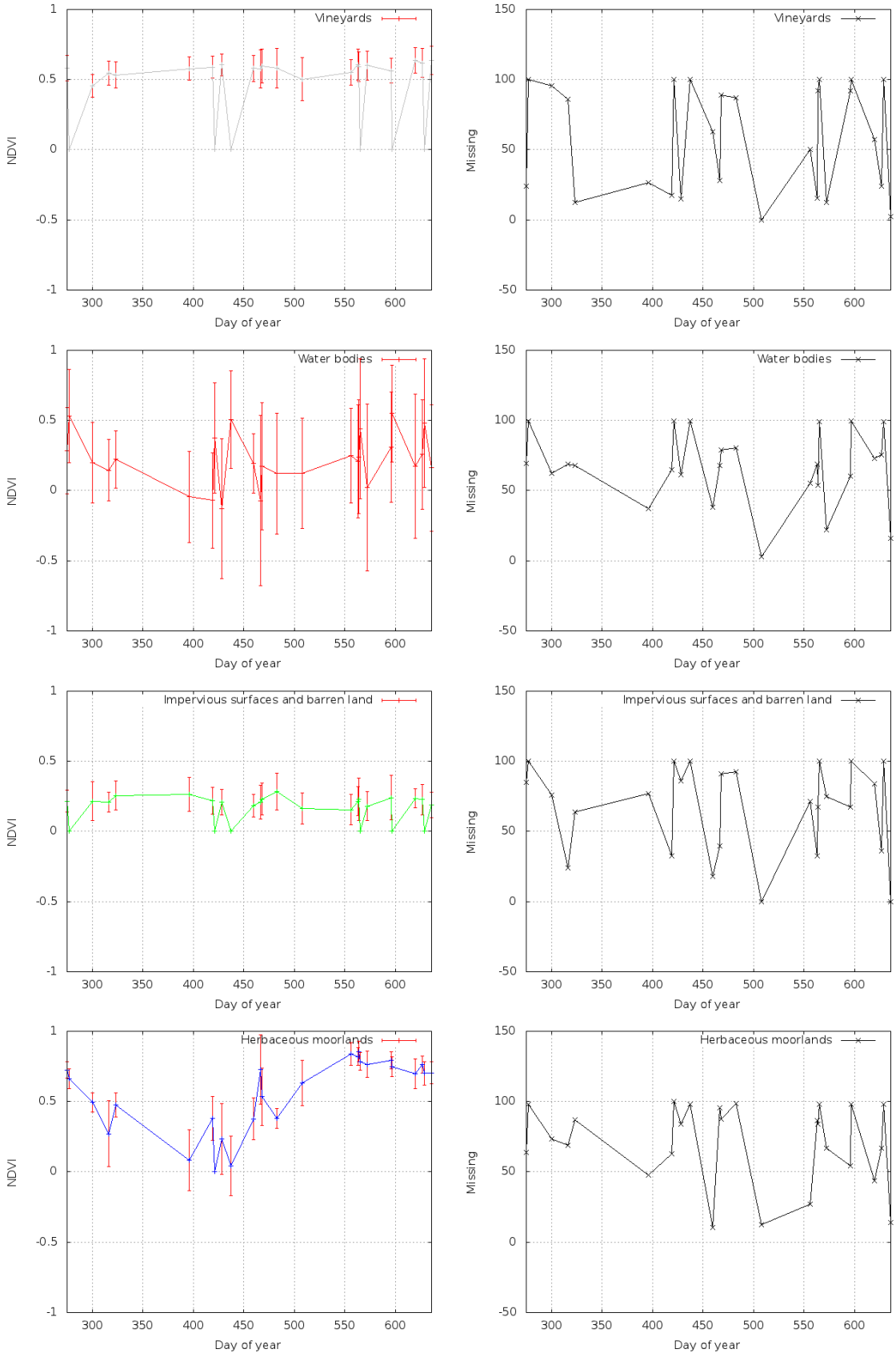


Figure 3.8: As in Fig.3.6.

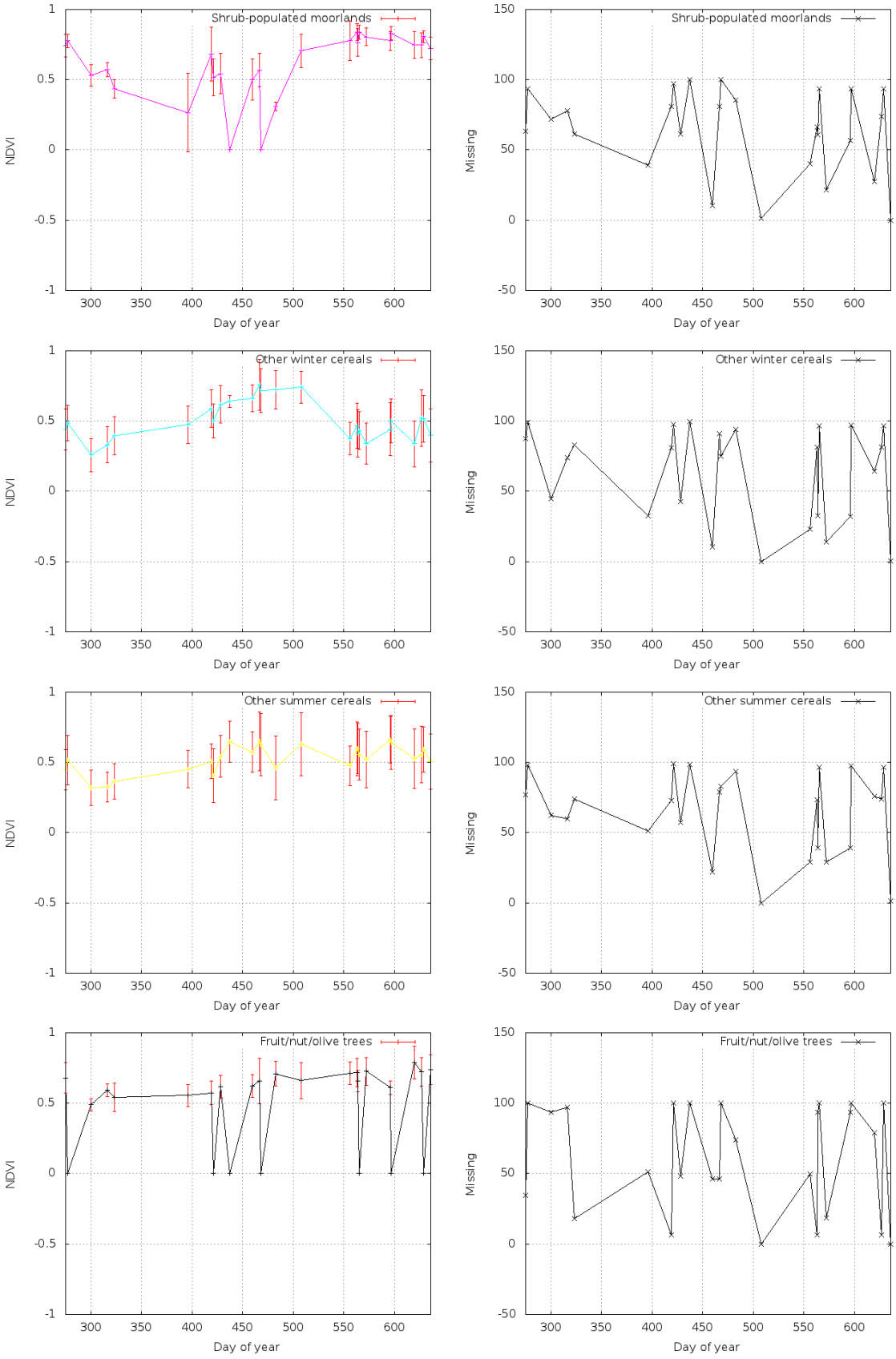


Figure 3.9: As in Fig.3.6.

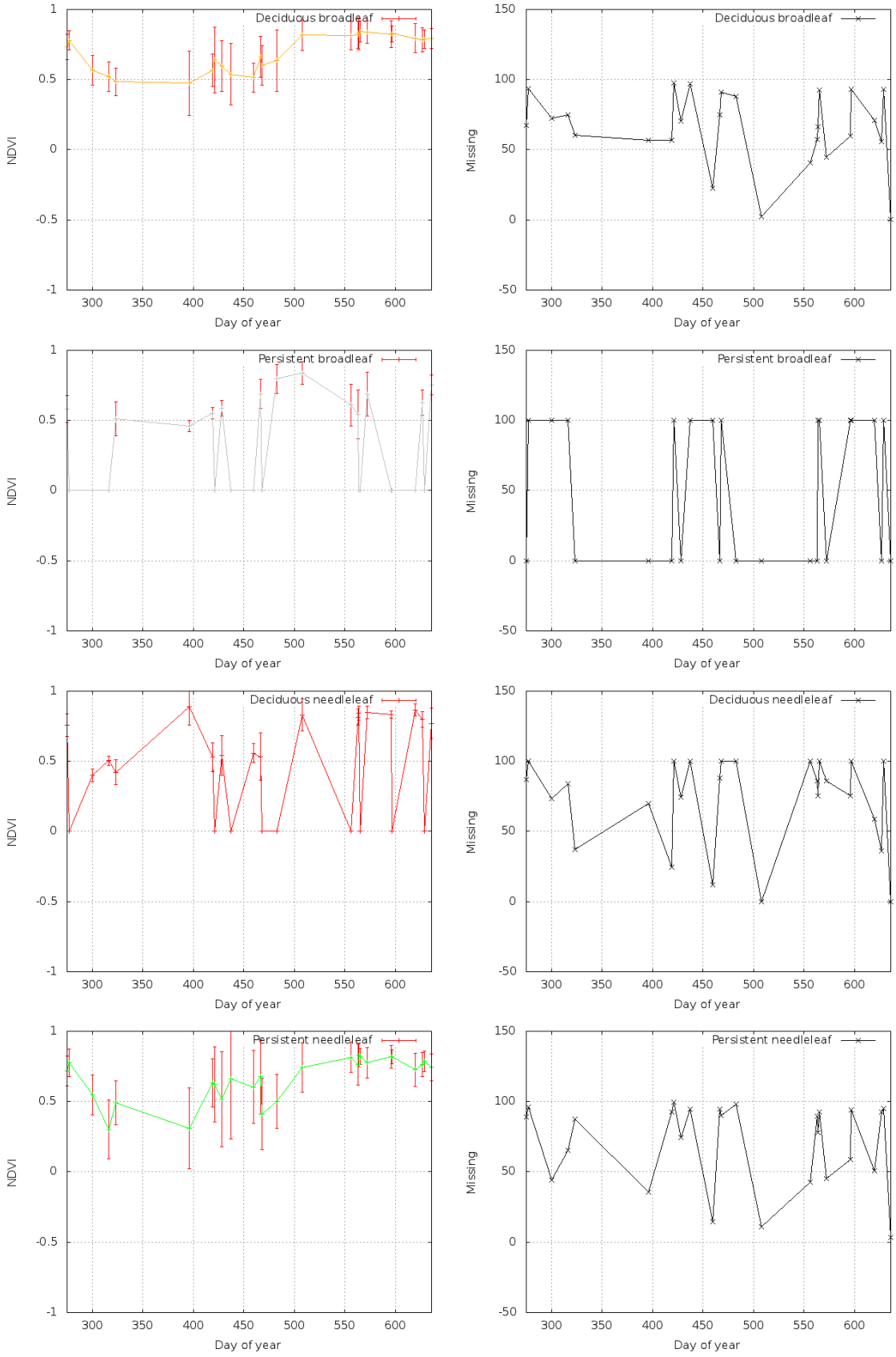


Figure 3.10: As in Fig.3.6.

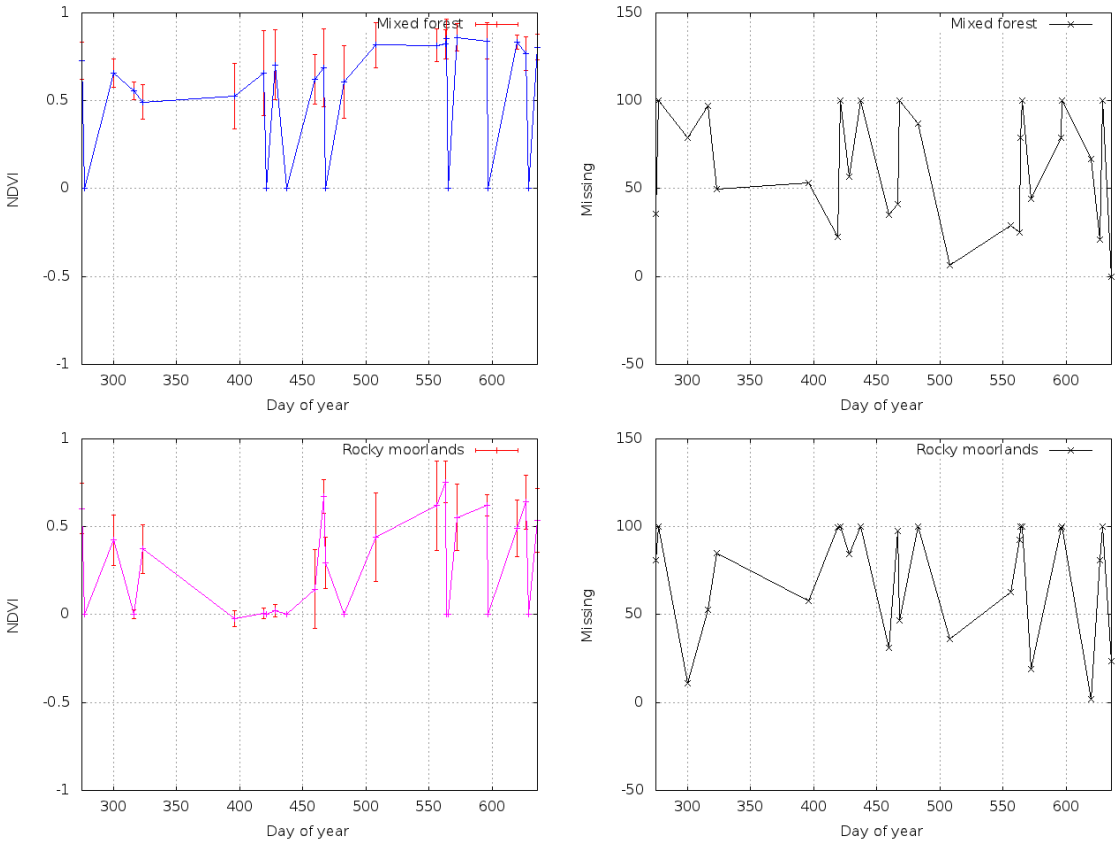


Figure 3.11: As in Fig.3.6.

Chapter 4

Time series missing information processing for large area mapping

Contents

This chapter introduces the problem of missing information in satellite time series and offers an overview of the literature for missing information processing. A multi-temporal method to produce temporally and spatially continuous high resolution time series for large-scale processing is presented, based on the application of local kernel interpolation techniques on the temporal series. It allows for the processing of missing information due to the presence of clouds, cloud shadows, snow, and satellite acquisition borders, and for the generation of continuous series over large areas based on virtual timepoints that can be then processed by standard machine learning algorithms. Different approaches to the proposed method are implemented and compared. Assessment is undertaken on simulated missing data, by analysis of temporal profiles, and based on classification results over large regions.

4.1 The problem of missing information

The presence of missing valid information in datasets constitutes one of the main problems for machine learning tasks. Data irregularities such as gaps or noisy information can confuse the process of learning, leading to a poor discovered pattern that cannot be applied successfully on new samples. The application of algorithms for the processing of missing valid information becomes thus important for the improvement of the accuracy and reliability of the process of knowledge extraction.

When dealing with remote sensing time series, invalid information is commonly found, for example due to atmospheric effects and sensor anomalies, with different degrees of noise impact. Most significantly, it is found in the form of border areas outside satellite passage zones, and of cloud, cloud borders and cloud shadow contaminated areas when optical wavelength sensors are used, as it is the case here. Fig.4.1 exemplifies a Landsat time series, which have been applied in this work, where the presence of missing information and noisy data contaminated with clouds and cloud shadows can be observed.

Tab.4.1 shows the percentages of cloud, shadow, snow and acquisition border presence in one of the annual Landsat datasets for tile L930463 that have been applied in this work. High percentages of affected pixels can be observed for all variables: for the crop year 2009-2010, average percentages attain 15.490% for cloud, 6.816% for cloud shadows, 3.162% for snow, and 47.730% for no-data outside border areas. Thus, an average of 73.198% of each image is contaminated by noise or presents missing information and, for the year, 100% of the pixels are affected at some point in the time series.

The presence of missing and noisy data described above results in the sparseness of valid observations in the data series and the existence of important data gaps. The application of machine learning algorithms on such sparse and irregularly sampled series presenting significant volumes of missing valid data is a problematic issue, particularly when targeting fast mapping over large areas. In this case, the automatic production of consistent, spatially and temporally complete time series may improve data extraction for the learning of knowledge models and their application for the generation of land cover maps.

When large areas are considered, a need for uniformisation arises, both in terms of obtaining a robust knowledge model applicable to new samples over whole regions, but also in relation to augmented disparities for data acquisition dates, which constitutes a problem different to that of missing information. Fig.4.2 shows the first four acquisition dates for three of the considered Landsat tiles. The disparity of acquisition dates prevents the learning of a unique model that can be applied on whole study regions, which indicates the need for the generation of a spatially and temporally complete time series.

The extent to which temporal irregularities (in the sense of variety of patterns for the presence of valid information) can be found in the input data can be seen in Tab.4.1, with for example three consecutive-day acquisitions for adjacent areas in July, but no data in June. Indeed, the distribution acquisitions and missing information in each dataset has a clear impact on learning and classification tasks performed on the data; for example, cloud cover, with greater yearly presence on some areas, can significantly difficult land cover mapping. Also, this impact is related

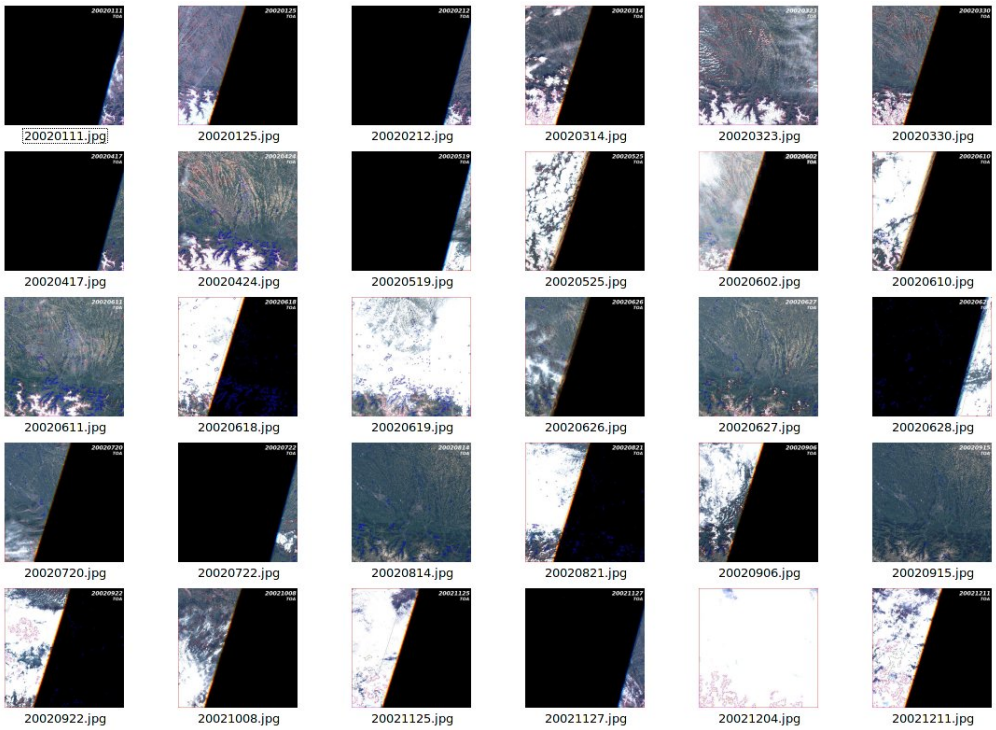


Figure 4.1: Example of Landsat time series for tile L930563 extracted from scenes 199-030, 198-030 and 197-030 (30 acquisitions from 11/01/2002 to 11/12/2002). The presence of missing information due to clouds, cloud shadows and passage borders can be observed. From [Hagolle and Huc, 2011].

.

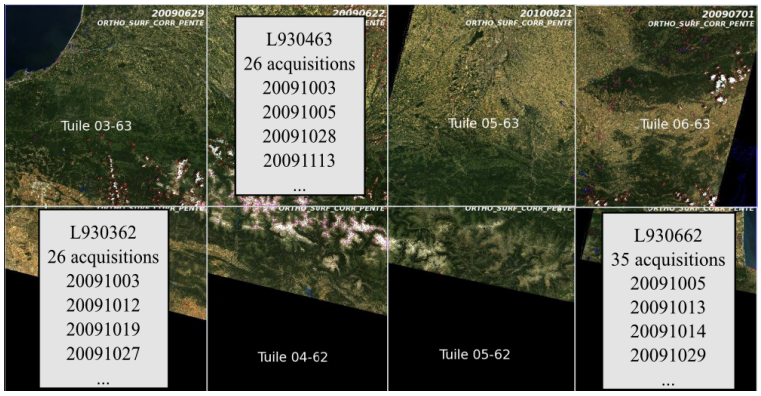


Figure 4.2: First four acquisition dates for the crop year 2009-2010 of three of the considered Landsat tiles, showing the existent disparity for different areas in the selected region.

to the phenomenon to be observed; for phenological cycle analysis, missing information has a stronger impact when there is coincidence with the main stages of vegetation, such as green-up or senescence. All these facts highlight a need for signal reconstruction allowing for resampling in a uniform time grid.

Date	Cloud %	Shadow %	Snow %	Border %	Missing %
2009/10/03	3.339	9.750	0.000	57.789	70.878
2009/10/05	0.198	0.005	0.000	96.689	96.892
2009/10/28	68.468	1.473	0.274	0.000	70.215
2009/11/13	68.236	6.876	13.893	0.000	89.005
2009/11/20	8.409	4.098	3.098	55.221	70.826
2009/11/22	6.001	0.189	0.021	93.428	99.639
2010/02/01	40.763	1.355	20.799	0.000	62.917
2010/02/24	5.592	0.834	3.616	57.965	68.007
2010/02/26	2.332	0.718	0.390	95.988	99.428
2010/03/05	50.856	5.838	12.678	0.000	69.372
2010/03/14	2.650	2.164	2.590	91.498	98.902
2010/04/06	23.381	3.937	11.887	0.000	39.205
2010/04/13	4.775	13.609	3.718	57.697	79.799
2010/04/14	20.261	10.580	3.546	54.296	88.683
2010/04/29	30.744	0.160	1.397	58.230	90.531
2010/05/08	36.041	38.383	11.068	0.001	85.493
2010/05/17	4.169	0.993	0.118	94.719	99.287
2010/05/24	0.745	1.402	7.430	0.004	9.581
2010/07/11	29.284	13.278	0.461	0.000	43.023
2010/07/18	0.890	5.832	0.086	57.455	64.263
2010/07/19	0.150	0.875	0.000	64.101	65.126
2010/07/20	0.141	0.033	0.000	91.668	91.842
2010/07/27	9.362	25.164	0.162	0.001	34.689
2010/08/04	39.420	0.563	0.043	54.959	94.985
2010/08/05	6.846	0.281	0.000	92.242	99.369
2010/08/20	0.927	3.839	0.029	54.620	59.415
2010/08/21	0.017	0.008	0.000	95.737	95.762
2010/09/13	13.397	54.782	0.000	0.000	68.179
2010/09/20	0.970	3.830	0.000	59.836	64.636
2010/09/22	0.142	0.168	0.000	94.527	94.837
2010/09/29	1.522	1.261	0.511	0.000	3.294
Average %	15.490	6.816	3.162	47.730	73.198
Pixel %:					100.00

Table 4.1: Percentages of cloud, shadow, snow and acquisition border presence in each of the images of the LANDSAT TM and ETM+ dataset for tile L930463 for the crop year 2009-2010. High percentages of affected pixels in the input data can be observed.

Furthermore, the new spatial missions such as Sentinel-2 will yield high resolution datasets, with increased temporal, spatial and spectral capabilities, and thus much larger data volumes to be processed. In the quest for an operational methodology, traditionally used algorithms and approaches cannot be longer used, in particular those relying on manual operator-dependent tasks, such as the selection of areas for which sufficient information is available for the direct application of standard processing algorithms, or the experimental setting of local thresholds for classification. Also, the presence of significant percentages of invalid data should not justify the elimination of full acquisitions, where still some data remains valid. Instead, the development of new algorithms and methods that do not discard valid information, while being able to cope with great volume datasets in an automatic fashion, is needed. Recent research has highlighted this fact, such as [Petitjean, 2012].

Thus, the problem of missing valid information over large areas defines some main necessities:

firstly, consistent, spatially and temporally complete time series may improve the reliability of data extraction for the learning of knowledge models and their application for the generation of land cover maps; secondly, uniformisation becomes necessary in order to be able to process large areas not covered by the same satellite passages; thirdly, automation is also needed for an operational methodology to be capable of processing the new volumes of data that will be available with the new generation satellite missions.

It can be asserted that there is no standard product giving a solution to the stated needs at the targeted resolutions. A processing chain for the production of Sentinel-2 Level 3A (composites of Level 2A with pixels free of clouds and cloud shadows) exists; however, it requires long input periods (a duration of 50 days is recommended for cloudy sites) in order to ensure the presence of valid pixels to allow computation. The use of composites has proven to be adequate for high temporal resolutions, but can be disadvantageous, as explained before, when lower temporal resolution series are used, or for time series with significant levels of missing information, unless a previous signal reconstruction approach is applied. Thus, a different multi-temporal method to produce temporally and spatially continuous HR time series for large-scale processing based on the application of local interpolation techniques on the temporal series is proposed. To this purpose, Landsat data is used as a simulator of the coming Sentinel-2 time series. Different temporal interpolation procedures are tested and compared, and a method of interpolation allowing for temporal resampling strategies is proposed, in order to obtain temporally and spatially continuous time series that can be processed by standard classification algorithms.

4.2 Background

This section gives an overview of the literature for missing information processing with remote sensing imagery. Although not developed in this work, some references are given for cloud and cloud shadow detection, justifying the methodology applied for preprocessing. Then, an overview of the existent methods for missing data processing with different sensors is offered, ranging from single-date to multitemporal approaches.

4.2.1 Cloud and cloud shadow detection

Cloud and cloud shadow contamination constitute a significant problem for the analysis of optical time series, as they limit the amount of valid surface information for an acquisition. The presence of clouds and their shadows, with their respective brightening and darkening effects, is indeed a problem for subsequent atmospheric correction tasks, aerosol retrievals, biophysical parameter extraction, and ultimately land cover classification [Sedano et al., 2011] and land cover change estimation.

In particular, optical images are much affected by clouds, especially in the tropical regions [Asner, 2001]. Figs. in 4.3 illustrate the variable impact of clouds in available time series for the years 2009 and 2010, respectively, of the Landsat tile L930563, corresponding to a region around Toulouse in France. Also, source Landsat scene characteristics determine the amount of available information. The number of non-missing and non-noise-affected timepoints is observable in the illustration. Blue areas correspond to lower coverage regions between Landsat5 and Landsat7

adjacent scenes affected by satellite passage borders or cloud cover.

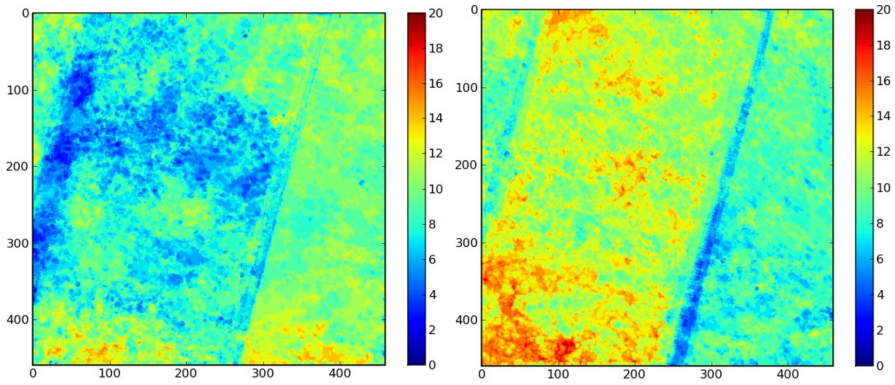


Figure 4.3: Number of available temporal points per pixel not affected by clouds for Landsat tile L930563 for years 2009 and 2010. From [Hagolle and Huc, 2011].

The global annual mean of cloud cover has been estimated at 66% by the International Satellite Cloud Climatology Project Flux Data (ISCCP-FD) [Zhu and Woodcock, 2012]. Indeed, only applications requiring one Landsat cloud-free observation per year are largely unaffected by cloud presence, with applications requiring two cloud-free acquisitions per year or more being affected by cloud data constraints [Ju and Roy, 2008], and significant low global probabilities for cloud-free acquisitions in particular seasons or with a frequency of more than once per year. These facts highlight the need for new approaches to mitigate cloud effects in Landsat imagery [Ju and Roy, 2008].

The Automated Cloud Cover Assessment (ACCA) system has traditionally been used for cloud screening in Landsat data, applying spectral filters and the thermal infrared band to estimate overall percentages. However, it does not yield sufficiently precise boundaries and locations to be applied in automated analysis of Landsat time series [Zhu and Woodcock, 2012]. Given the relevance of the problem, a variety of methods have been developed for cloud and cloud shadow detection, going from the empirically tuned thresholds combined with temperature (approaches for the moderate resolution sensors AVHRR and MODIS) to more sophisticated alternatives using for example time series analysis [Sedano et al., 2011]. Succint reviews of methods are given in [Goodwin et al., 2013] and [Zhu and Woodcock, 2012].

Some of the latest methods include [Zhu and Woodcock, 2012], [Goodwin et al., 2013] and [Hagolle et al., 2010]. [Zhu and Woodcock, 2012] presents a method called Fmask for cloud and cloud shadow detection in Landsat imagery based on reflectance and temperature information. For a globally distributed set of reference data, accuracies reported were high at 96,4% with Landsat imagery. However, the new missions such as Sentinel-2 or Venμs do not offer a thermal band. Using a multitemporal approach, [Goodwin et al., 2013] proposes a hierarchical approach using spectral and contextual rules for the state of Queensland in Australia, with some improvement in cloud and specially cloud shadow detection in comparison to the Fmask method.

The Multi-temporal Cloud Detection (MTCD) method of [Hagolle et al., 2010] combines the de-

tection in the time series of sudden increases of reflectance in the blue wavelength, with a comparison of reflectance variations between the blue and the red wavelengths, and a test of linear correlation of pixel neighbourhoods between successive images. It does not make use of the typically used thermal band, as neither Sentinel-2 nor Venμs offer such a band, and it has been conceived to be globally applied, showing better results than traditional methods. Additionally, with Sentinel-2 a spectral band at 1.38μm will enhance the detection of high clouds. The MTCD will be used for Venμs processing and proposed for Sentinel-2 level 2 processing [Hagolle et al., 2010], and is thus part of the preprocessing steps (although not developed in the frame of this work) that have been applied on the satellite data time series used here. Figs. in 4.4 show two examples of cloud detection by the MTCD method for two different dates on two regions of the study area.

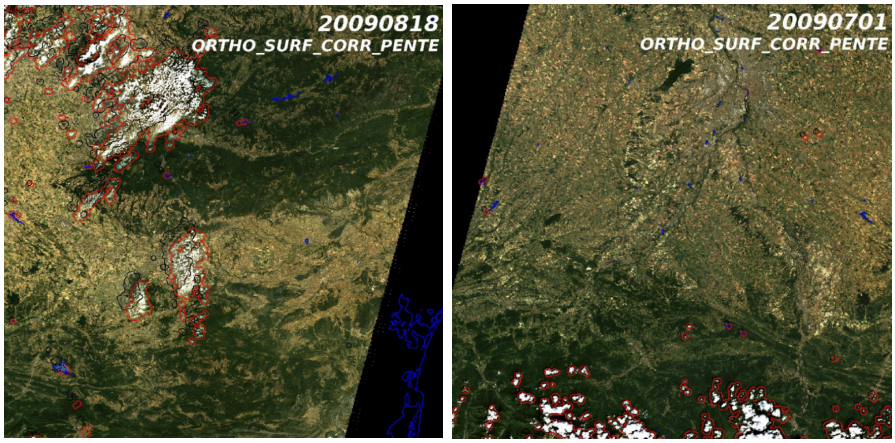


Figure 4.4: Cloud detection by the MTCD method for Landsat tiles L930563 and L930663 [Hagolle and Huc, 2011].

4.2.2 Missing information processing

Cloud and cloud shadow effects do not constitute the only sources of missing information in satellite time series, as explained. Indeed, the sparseness of valid observations in a time series is also determined, aside from the available acquisitions, by the presence of snow and satellite path border areas. Noise is also present due to atmospheric effects and sensor anomalies, although this problem has not been treated here; the presence of these variables is not easy to anticipate. Similarly, the impact of clouds and cloud shadows is not easy to predict, as there is dependence on the spatio-temporal dynamics of clouds and satellite acquisitions, and also on the period of interest for the analysis and the seasonality of the phenomenon which is studied [Ju and Roy, 2008], as explained in Sec.4.1.

Until recently, land cover analyses were commonly based on a limited number of images, due to cost concerns of satellite imagery before the advent of the free data policy with Landsat, which will be continued with missions such as Sentinel-2 and Venμs. In this context, scenes with a cloud coverage greater than a certain threshold were typically rejected, in this way also rejecting valid information areas contained in the acquisitions. Also, with a limited dataset, a common approach was to discard areas with missing information manually, in order to be able to apply standard processing algorithms on the remaining regions. However, with the imminent high resolution time

series this approach becomes unfeasible, with great time consumption and other costs, and severely limiting machine learning tasks. In order to exploit dense multitemporal datasets, automation of the processing of missing information is needed, so that no manual selection of valid area regions is further required.

A variety of methods for cloud, cloud shadow and noise removal can be found in the literature. Gap imputation approaches can be summarised as: (i) merging information from two or more sensors, with the same or different resolutions, (ii) using information from neighbouring pixels in a spatial window (spatial processing), or (iii) using information from the same pixel in a temporal window (temporal processing) [Poggio et al., 2012]. Corregistration of the images of different sources is an issue for (i), while gap filling of large spatial gaps can result in unacceptable smoothing effects for (ii) [Rulloni et al., 2012]. A review on spatial-based methods for missing information reconstruction including interpolation, propagated diffusion, variation-based and exemplar-based methods is given in [Shen et al., 2015], with negative conclusions when large areas are considered. A comparison of spatial interpolation techniques is given by [Hartkamp et al., 1999].

Multitemporal approaches in (iii) are commonly based on global function fitting or on compositing. Global function fitting, either by interpolation or regression analysis, consists on the formulation of the curve best fitting a series of datapoints. It requires at least some knowledge on the land cover classes and their temporal behaviour for the selection of an appropriate target function, such as logistic or polynomial of a particular degree, in order to perform an appropriate curve fit that describes the seasonal behaviour of vegetation. Compositing is typically based on the selection of cloud-free information in a temporal window, and was first developed with AVHRR data using the maximum NDVI criteria, although other criteria have been developed, such as the maximum difference between near-infrared and red reflectances or the minimum red reflectance.

For composite creation, as only one observation in the temporal window is selected, temporal biases are common [Ju et al., 2010]. A major limitation for the use of composites is the need for very close acquisition dates, so that the dynamics of the area that is studied is slower than the temporal window considered [Rulloni et al., 2012]. Indeed, compositing is generally applied with high temporal resolution time series, such as those of MODIS or AVHRR. For lower temporal resolution series such as those of Landsat 5 or 7 and Sentinel-2, the application of composites becomes problematic. Indeed, temporal biases are most prominent when available information is sparse, and in some locations it may not be possible to obtain sufficient cloud-free observations during particular times of the year [Watmough et al., 2011].

Furthermore, for any compositing method, further time series denoising is necessary to remove remaining artifacts, such as temporal averaging, median smoothing, or the application of Fourier transforms or wavelet decomposition. These last two methods eliminate changes based on a frequency threshold, which determines the drawback that small variations may still be rejected even when they correspond to land surface changes [Latif et al., 2008]. Also, they are generally applied after preprocessing for artifact removal to series with limited remaining artifacts [Latif et al., 2008].

A variety of approaches have been designed for gap-filling with sensors with short revisit periods, both working on reflectances or on extracted primitives. An example are MODIS products,

often degraded by atmospheric conditions such as dust, cloud cover, and other aerosols, and sensor related artifacts [Poggio et al., 2012]. For instance, [Gao et al., 2008] presented an algorithm for MODIS Leaf Area Index (LAI) time series based on the use of the TIMESAT software [Jönsson and Eklundh, 2004] with the help of seasonal variation curve fitting. The approach required work on three-year periods and knowledge on the targeted land cover types. Similarly, [Hird and McDermid, 2005] presented a comparison between algorithms for noise reduction in 16-day MODIS NDVI composite time series, where the TIMESAT asymmetrical Gaussian function fitting [Jönsson and Eklundh, 2002] and the double logistic function fitting [Beck et al., 2006] performed better than other filters such as the Savitzky-Golay filter and the mean-value iteration filter. Again, work was based on three-year periods.

Other common approaches include threshold-based methods, such as the Best Index Slope Extraction (BISE) algorithm [Viovy et al., 1992], requiring the definition of a sliding period and a threshold of percentage increase, which is subjective and operator dependent [Chen et al., 2004]. Asymmetric function fitting methods are also common in the literature [Jönsson and Eklundh, 2002], having as drawbacks the complexity of the fitting, which makes it time-consuming, and the difficulty of identifying consistent sets of maxima and minima to perform the fit, especially with noisy data [Chen et al., 2004].

More recently, an iterative method for NDVI time series smoothing with the Savitzky-Golay filter is given by [Chen et al., 2004] using 10-day Maximum Value Composite (MVC) SPOT outputs after linear interpolation of cloudy values. Superior results are reported against those of the BISE algorithm and a Fourier-based fitting method. Other methods include non-local means methods, applying a weighted mean of pixels in the image based on similarity to the target pixel, and other approaches for the search of similar pixels, which may be problematic when considering time series with a high amount of missing data. Also, [Borak and Jasinski, 2009] presented a comparison of different simple interpolation techniques with MODIS LAI products, including per-class mean (PDM) substitution (requiring knowledge on the target classes), linear averaging using a time window (presenting difficulties with large time gaps), or a hybrid between spatial interpolation using a space window and previous temporal observation substitution (sensitive to the spatial window and the size of spatial gaps), with variable results, as each technique presents different drawbacks.

A learning scheme to recover contaminated data with a Kohonen Self-Organizing Map (SOM) has been proposed by [Latif et al., 2008], combined with a previous outlier detector, and applied on low resolution time series to estimate missing values. Results are positive using visual investigation and correlation estimates of recovered information for a selected region. [Rulloni et al., 2012] proposed three algorithms for missing data imputation, two based on merging information with lower resolution sensor information, and a third one using knowledge on the map of classes. Finally, [Poggio et al., 2012] proposed a spatio-temporal method for gap filling of MODIS EVI time series, with varying performances and moderate processing times, both depending on the space-time pattern of missing data. A technical review on missing information reconstruction with temporal-based approaches including temporal replacement methods, temporal filter methods and temporal learning model methods is given in [Shen et al., 2015].

4.3 Proposed method

The proposed method has been developed for application with the new generation Earth Observation missions in the context of the literature overview presented in the previous section and the described drawbacks of different techniques. Thus, assumptions on the characteristics and distribution of land cover classes and noise patterns, as well as on vegetation dynamic shapes, have been avoided. Input temporal series time spans are not limited nor designed for a particular length of time, the approach can be applied on any desired time span, and subjective and operator dependent decisions have also been avoided.

The expected availability of acquisitions, with 1 or 2 images per month (rather than 1 per day with satellites such as MODIS), has also been taken into account, which determines the unfeasibility of the application of composites that are common with high temporal resolution time series. The presence of clouds is considered as granted in mostly all images (in comparison to the use of almost cloud-free images and the manual detection of clouds with other satellites such as SPOT), and the automatic processing of large regions is understood as necessary, as opposed to the traditional highly supervised processing of one-image datasets. Finally, simplicity and the reduction of computational load have been aimed for.

4.3.1 Description

With these premises, this work is based on a signal reconstruction approach using local kernel interpolation in order to obtain spatially and temporally seamless time series. Furthermore, when very large areas are considered, the methodology allows for temporal resampling in a uniform time grid as illustrated in Fig.4.5, which in turn permits the global application of standard classification algorithms to the generated products, and ultimately the automated generation of land cover maps over large areas.

The proposed method is based on locally varying kernel interpolation and temporal resampling in the time series for the extraction of features for classification in order to permit the exploitation of information on the spectral response of land cover classes in time, and particularly of the phenological variation of vegetation, to attain good separability among land cover types. For vegetation, this variation should reflect green-up, peak condition and senescence stages, for example. In all cases, noise and missing information should be minimized. A variety of methods exist in the literature with this aim, as seen, such as the use of composites or global fitting methods. As explained, methods based on compositing generally present temporal biases, especially when available information is sparse, and are thus not used here. Global fitting methods, such those applying asymmetric Gaussians or double logistic functions, require some a priori knowledge on the expected spectral response of classes (as explained, in order to define appropriate functions to perform fitting) and are consequently also discarded. An alternative is the application of local interpolation methods; in this way, signal reconstruction is targeted rather than just the generation of cloud-free images, and no previous knowledge on land cover is required.

Partial acquisitions are determined in spatial extent and temporal frequency not only due missing and corrupted data related to meteorological conditions such as cloud cover and shadows, but also due to different satellite passages. These factors make it impossible to obtain measurements

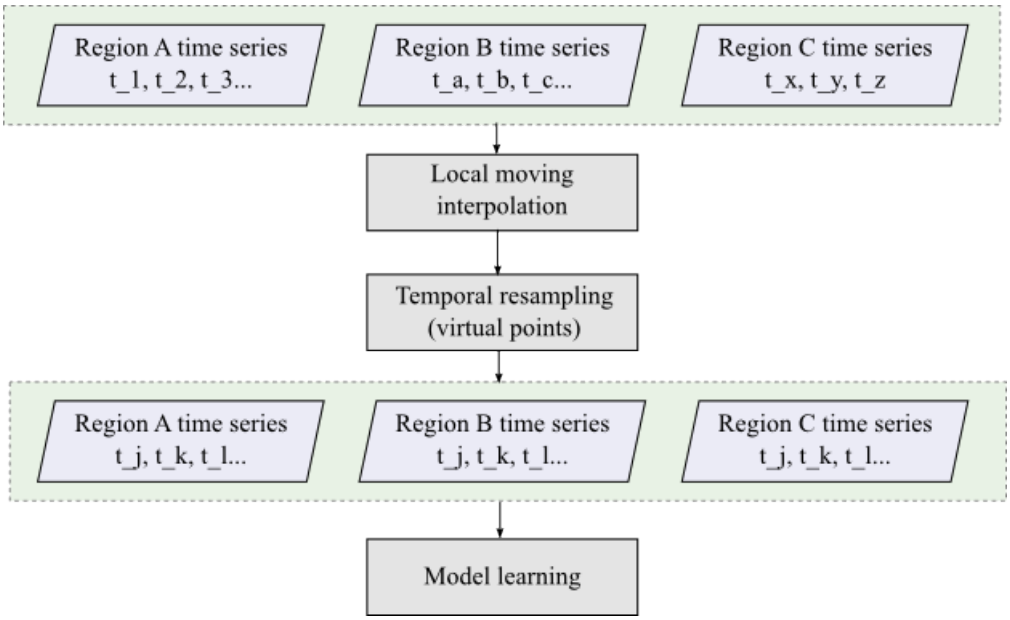


Figure 4.5: Proposed method for for the generation of uniform and temporally and spatially continuous time series very large areas. See Fig.3.1 for integration in the proposed general framework

.

for spatially distant points that are directly comparable, as the acquisitions are unequally spaced in time and also non-coincident. The approach here undertaken takes these acquisitions and makes predictions on missing values by applying local interpolation methods in the temporal series.

When very large areas are considered, the use of interpolation allows also for the generation of continuously sampled data by means of temporal resampling. As shown in Fig.4.5, given multispectral series for different regions with irregularly-spaced acquisitions at times t_1, t_2, t_3, t_a, t_b, t_c, and t_x, t_y, t_z, the application of local intepolation and resampling allow for the generation of the corresponding values at regularly spaced timepoints t_i, t_j, t_k. This allows for the production of a temporally and spatially continuous dataset that can be subsequently processed by standard data processing algorithms such as model learning methods for classification on the whole available dataset, based on different passage series and acquisition dates, rather than on a traditionally used single location series with common acquisition dates.

Three different approaches with kernel functions have been implemented and tested for comparison, and they are described in the following subsections. Kernel functions are non-negative even functions presenting a maximum value at zero distance and decaying as the the distance increases. A variety of kernel functions have been proposed, such as the boxcar kernel [Stone, 1977], the Gaussian kernel [Deheuvels, 1977] and the Epanechnikov kernel [Epanechnikov, 1969] [Atkeson et al., 1997]. Here, a linear or triangular kernel, a Gaussian kernel and a polynomial kernel have been applied in conjunction with an adaptive radius; that is to say, the radius of search for valid values around missing data is adapted to the existent dataset particular timepoint and pixel, so that a locally varying kernel interpolation is applied in a moving window fashion along the time series.

4.3.2 Triangular kernel interpolation

The first of the approaches consists of the application of local linear interpolation for missing values in the time series. Each function is fitted to valid values in a variable neighbourhood of missing data, and the interpolator is successively shifted until all missing values have been estimated. In practice, given two known values y_0 and y_1 at times x_0 and x_1, any interpolated value y in the interval (x_0, x_1) can be geometrically derived from Eq.4.1:

$$y = y_0 + (y_1 - y_0)\frac{x - x_0}{x_1 - x_0} \tag{4.1}$$

An advantage of this approach is its simplicity and rapidity; however it incurs in extrapolation where the unknown value is situated outside an interval determined by known samples, that is to say, when no valid data is to be found on both sides of missing datapoints. In these cases, common approaches are extrapolation according to Eq.4.1 (which may produce values greatly distant to the reference curve, and even outside the primitive ranges, especially when the searched value is located far from the interval) or replication of the closest known value (which ignores the curve trend). In order to overcome these limitations, Gaussian curve fitting has been applied in these cases, in order to avoid values greatly distant to the reference curve and to remain within the primitive range, as given in Eq.4.2, where a is the amplitude of the bell curve, b the center of the peak, c the standard deviation, and d the asymptote that the curve approaches as it reaches farther:

$$y = a\exp\left(-\frac{(x-b)^2}{2c^2}\right) + d \tag{4.2}$$

Given the last known values in the time series y_1 and y_2 at times x_1 and x_2, two different fittings can be obtained. Then, $(y-d)/a = \exp\left(-(x-b)^2/2c^2\right)$ and applying natural logarithms, $ln(y-d)/a) + (x-b)^2/2c^2 = 0$. As two different equations are derived for x_1, y_1 and x_2, y_2, renaming $A_1 = ln(y_1 - d)/a$ and $A_2 = ln(y_2 - d)/a$, and equating, $A_1 + (x_1 - b)^2/2c^2 = A_2 + (x_2 - b)^2/2c^2$ is obtained. The two possible centers b_1 and b_2 can be computed, given $A_1/A_2 = ((x_1-b)^2/2c^2)/((x_2-b)^2/2c^2)$, with Eq.4.3:

$$b_1 = (A_1x_2 - A_2x_1 \pm A_2x_1(A_1/A_2)^{(1/2)} \mp A_2x_2(A_1/A_2)^{(1/2)})/(A_1 - A_2) \tag{4.3}$$

Given b_1 and b_2, two possible standard deviations c_1 and c_2 can thus be calculated, with Eq.4.4:

$$c = sqrt(((x_2 - x_1)(x_2 + x_1 - 2b))/(2(A_1 - A_2))) \tag{4.4}$$

Only the solutions corresponding to the greatest standard deviation are adopted, as both valid points are assumed to be on the same side of the curve, in order to obtain the Gaussian fit.

The approach is simple and fast; however, it cannot make predictions above maximum or under minimum reference values, which results in the flattening of peaks and valleys when they are not included in the sample values. Furthermore, error increases with the distance between known points. As the availability of valid samples in optical satellite time series is variable, the effects on interpolation error can also be different depending on the dataset. In addition to this, its output being linear, only approximations to curves can be obtained.

4.3.3 Gaussian kernel interpolation

A neighbourhood approach has been chosen for the second compared method, combined with a weighting technique in order to ponderate the importance of the neighbourhood pixels based on their distance to the targeted missing information. Gap-filling using a moving window with a Gaussian kernel gives a solution to the problems posed by traditional interpolation methods such as the nearest-neighbour interpolation, consisting of the assignation to unknown points of the nearest valid available data value, or the uniform kernel interpolation, where an averaging window is applied. These techniques produce very poor results when temporal distances between target and valid datapoints are considerable.

Thus, gap-filling using a moving window with a Gaussian kernel as formulated in Eq.4.5 has been used, with the weighting function w achieving its maximum at zero distance, and decaying smoothly with the increase of distance:

$$w = exp(-d^2)/\sigma^2 \tag{4.5}$$

For a position in time t, for $i = 1...n$, where n is the number of data points y, and w are the weights, the unknown is calculated as given in Eq.4.6:

$$y(t) = \sum_i (w_i y_i) / \sum (w_i) \tag{4.6}$$

Weight functions approaching zero at a finite distance allow faster implementations, since points farther from the target point than that distance can be ignored with no error. Thus, a confidence radius can be established by applying a moving window of a safe size to ensure acceptable results; in this case a wide radius of 60 days has been applied, given the temporal resolution of the given dataset, although a wider radius can also be used. Also, a variance $\sigma^2 = 1$ has been chosen as ponderation adjustment [Atkeson et al., 1997] in order to give a higher weight only to the closest valid timepoints, with no dependency on either the distribution or the number of available timepoints around missing timepoints, which is highly variable and can be highly biased. In this way, the kernel presents a constant behaviour, with no distribution-dependent normalisation in the range.

The effect of flattening of peaks and valleys affecting the previous linear interpolation method is also to be found here. Indeed, when no high or low points are available as sample points, no estimates can be made above or below the given maximum and minimum samples. Also, only data points with relatively close valid timepoints can safely be estimated, as confidence in the estimation decreases with the distance between the predicted point and the reference values.

4.3.4 Polynomial kernel interpolation

The third of the approaches consists of the application of an interpolation based on the Savitzky-Golay interpolation method, consisting of a moving least squares fit along the temporal axis of data, and improved with the use of an adaptive radius.

The digital filtering method was introduced by [Savitzky and Golay, 1964] in order to simultaneously smooth and differentiate chemistry-related time profiles. The original algorithm provided a simplified method to replace a least-squares technique by an equivalent and simple convolution

[Gorry, 1990] with the property of performing noise reduction while preserving higher order moments of the original profile [Ruffin and King, 1999]. For a given filter length, polynomial order and derivative order the authors obtained coefficients that proved to be generic for application as convolution coefficients, and were thus published in the form of tables that could be directly applied for the processing of data time series. Corrections to these tables were later published by [Steinier et al., 1972] and [Madden, 1978].

The convolution coefficients were only applicable in cases of equally spaced data points, with an analytical solution to the least-squares problem. Then, for a set of n points (x_j, y_j) where x_j is the independent time variable, $j = 1, \ldots, n$, $i = 1, ..., m$, and m the number of convolution coefficients c_i, the solution at central point of each subset can be computed with Eq.4.7:

$$y_j = \sum_{i=-(m-1)/2}^{(m-1)/2} c_i y_{j+i} \text{ where } (m+1)/2 \leq j \leq n-(m-1)/2 \tag{4.7}$$

Having time series with unequally spaced data acquisitions, the premise of equally spaced data is not adopted here. Thus, the filtering method consists of the fitting of consecutive sets of adjacent points with a low-degree polynomial by applying the original least squares technique. Generally, a low order polynomial is applied, in order to approximate the low frequencies of the spectrum with the polynomial, while loosing the high frequency noise in the approximation error [Ruffin and King, 1999]. As the polynomial of first order is unsuitable for smoothing data with curvature over m data points, a second order has been applied.

The use of least squares allows also for the application of weights depending on the input noise. This option has not been applied here, as no assumptions on types of noise have been made, although it could be applied in the future in order to better tune the interpolation when the impact of each type of noise is known.

At this point, two main disadvantages of the method, both the original of [Savitzky and Golay, 1964] and the variant applying unequally spaced data and a moving least squares approach with weights, are observable. Firstly, no computation is possible at the beginning and end of the series for the values preceding and following the first and last computed values in a neighbourhood of m values, that is to say, for values $j < (m+1)/2$ and $j > n-(m-1)/2$. Secondly, a radius m needs to be imperatively chosen for the standard application of the algorithm, which needs to be adequate to the available frequency of acquisitions and to the presence of missing information. For example, for a given radius in a time series, areas with cloud presence in the interval would not present enough data points for the least squares calculation to be undertaken.

In order to overcome both problems, an adaptive radius has been applied. The moving window performing the fit is in this way of variable size, depending on the necessary number of valid points. This is in turn defined by the order of the least squares fit: for a degree k (here $k = 2$), $k+1$ points are needed to estimate a result. In this way, instead of using a fixed radius, this variant looks for at least the minimum number of valid points needed to undertake the interpolation, admitting the application of different radius lengths, and allowing calculation for values $j < (m+1)/2$ and

$j > n - (m-1)/2$.

In this way, the main drawbacks of the standard method are overcome. Firstly, there is no need for a priori radius selection, as it is variable for each neighbourhood and determined by the available information. Secondly, the use of equally spaced timepoints is not required any more, as no coefficients are applied. Thirdly, values in the extremes of the time series can be calculated in this way. In the fourth place, calculation is ensured at every timepoint and no missing information is generated due to the failing of the constraint of enough data points in a given radius. In general, advantages include the fact that curves can be approximated, as a second degree polynomial is used, and that the method remains simple and fast.

4.4 Experiments and results

Experiments have been undertaken in order to evaluate the feasibility of using signal reconstruction with local kernel interpolation methods and temporal resampling in a unified time grid, as proposed and presented in the previous section, for missing information processing and land cover classification over large areas.

The three approaches for missing information processing described in Sec.4.3 have been implemented and compared by establishing quantitative measures of error in the reconstruction with synthetic data, and also in terms of the impact on classification results, for three datasets: the first two covering a region of $12100km^2$ each, and the third one comprising the whole study site of $84000km^2$.

4.4.1 Experimental setting and evaluation metrics

In order to assure comparability and to demonstrate the stability of results, a classic cross-validation approach is applied for the evaluation of the impact on classification results. The experimental setting is the following: n independent sets of learning samples have been randomly produced for each desired number of samples and used in conjunction with each of the compared approaches. Given a sample set, a complementary set excluding all training samples and incorporating all the rest of available ground truth has been created and applied for validation.

In this way, for each set of training examples all other pixels in the ground truth data are used as validation samples, so that training and validation are always performed on non-overlapping sets. N has been set to 20. For each independent set, a classification has been generated and evaluated; over 500 full training, classification and validation tasks have been undertaken; for each configuration only summary results averaging classification scores for 20 generated test sets are presented. The process applied for evaluating results is summarized in Fig. 4.6.

Quantification of classifications is based on a confusion matrix for ground truth and produced classifications, where accuracy is assessed in terms of average global accuracy, average percentages of well-classified pixels per class, average standard deviation of the classification results, and average kappa statistic for 20 learning sets.

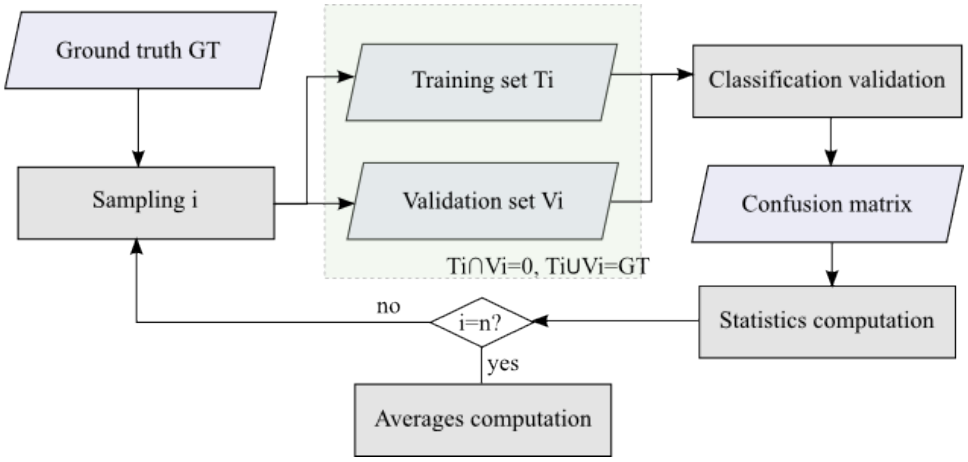


Figure 4.6: Illustration of the validation procedure

.

Overall accuracy is calculated as the total number of well classified pixels, the trace of the matrix, divided by the total number of validation pixels. For a confusion matrix M, given r the number of classes, n_{ij} the matrix position at coordinates $M(i,j)$, and $i,j = 1,...,r$ (with lines being reference classes and columns being those estimated by the classifier) it is formulated in Eq.4.8.

$$A = \frac{\sum_{i=1}^{r} n_{ii}}{\sum_{i=1}^{r} \sum_{j=1}^{r} n_{ij}} \tag{4.8}$$

Accuracy corresponds to the fraction of well-classified pixels of a classification class with regard to the total number of pixels of that class. It is calculated as given in Eq. 4.9.

$$A_i = \sum_{j=1}^{r} \frac{n_{ii}}{n_{ij}} \tag{4.9}$$

Averages for all statistics computed on the 20 independent sample sets are computed. For overall accuracy measures, the standard deviation σ is also calculated as in Eq.4.10 in order to quantify the amount of variation found in the classification results. N is the number of undertaken classifications, $i = 0...N$, and x_i corresponds to the overall accuracy statistic of classification N.

$$\sigma = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N} (x_i - \mu)^2} \text{ where } \mu = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{4.10}$$

The kappa coefficient (κ), a statistical measure of inter-rater agreement taking into account agreement occurring by chance, has also been computed. It estimates the degree of matching for a classification scheme, determining whether it is better than a random allocation. It is calculated as expressed in Eqs. 4.11, where P_o is the proportion of observed agreements, P_c is the proportion of agreements expected by chance, and $n_{i.}$ and $n_{.i}$ correspond to marginals.

$$\kappa = \frac{P_o - P_c}{1 - P_c}$$
$$P_o = \frac{1}{n}\sum_{i=1}^{r} n_{ii}$$
$$P_e = \frac{1}{n^2}\sum_{i=1}^{r} n_{i.}n_{.i} \tag{4.11}$$

Ranging between -1 and 1, different formulations for the evaluation of kappa exist. A well-known one is that of [Landis and Koch, 1977], where standards of strength are proposed as summarised in Tab.4.2.

≤ 0	poor
$0.01 - 0.20$	slight
$0.21 - 0.40$	fair
$0.41 - 0.60$	moderate
$0.61 - 0.80$	substantial
$0.81 - 1$	almost perfect

Table 4.2: Standards of strength of kappa proposed by [Landis and Koch, 1977].

Several remarks need to be made on the use of kappa. The magnitude of the coefficient is affected by a variety of factors, such as the number of categories. The higher the number, the greater the potential for disagreement; thus, kappa decreases with a greater number of categories [Sim and Wright, 2005]. Also, prevalence is influential when the proportion of agreements on a class differs from that on another. When prevalence is high, kappa is reduced, especially for large values of kappa [Sim and Wright, 2005]. Comparably, bias, the proportion of disagreements on classes by the raters has also an effect on kappa. When the bias is large, kappa is enlarged, especially for small values of kappa. Thus, kappa is influenced by distribution; for two confusion matrices with raters agreeing on the same number of cases, kappas can be greatly divergent. In [Pontius and Millones, 2011], five different kappa indices and their limitations are examined. Thus, kappas are seldom fully comparable across studies or populations, and need to be interpreted accordingly.

4.4.2 Interpolation assessment

Previous to classification, interpolation assessment has been undertaken in order to decide on the appropriateness of the considered variants of the proposed approach. In order to compare the different implementations, quantitative measures of the reconstruction error have been calculated. To this aim, a set of tests has been devised, where the analysis of the performance is based on the calculation of the mean squared error between synthetic reference temporal profiles and their corresponding reconstructed profiles when different amounts of random error are incorporated. The mean squared error (MSE) is calculated as follows. $\hat{Y}_i$ being a vector of n predictions, and Y the

corresponding vector of ideal values, then the MSE is obtained with Eq.4.12:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_i - Y_i)^2 \tag{4.12}$$

Twenty different synthetic temporal profiles have been created, similar to those expected from phenological curves, by producing curves with different sinusoidal amplitudes and frequencies for 100 temporal points. For each of these curves, six different levels of noise have been evaluated: 1%, 5%, 10%, 30%, 50%, and 70%. For each of these levels of noise, 20 different sets with randomly located noise have been produced. Finally, all sets have been tested for the three different implementations of the proposed approach. Thus, a total of 7200 experiments of profile reconstruction and evaluation have been undertaken. Tab.4.3 summarizes the obtained results, by offering an average for 20 curves for each of the different levels of noise, and for each of the approaches.

Noise	Tri.K. MSE (10^-5)	Gau.K. MSE (10^-5)	Pol.K. MSE (10^-5)
1%	9.9407	0.4562	0.0025
5%	4.1739	8.3517	0.0022
10%	2.7415	16.6483	0.0048
30%	2.5478	56.7697	0.0353
50%	6.0164	120.6006	0.1373
70%	18.2191	193.2450	0.6942

Table 4.3: Average MSE (10^-5) per level of randomly introduced missing information and interpolation approach, based on 20 sets for each curve (20 curves). Tri.K.: triangular kernel interpolation; Gau.K.: Gaussian kernel interpolation; Pol.K.: Polynomial kernel interpolation.

Results are given visually in Fig.4.7. The Gaussian kernel interpolation offers the worse results, with a curve greatly increasing as the level of noise increases too, and clearly with worse MSE values than the other two implemented interpolations. The MSE curve for the triangular kernel interpolation decreases and then increases with noise. The reason for this is the fact that, as explained, linear interpolations cannot make predictions above maximum or under minimum reference values, which results in the flattening of peaks and valleys when they are not included in the sample values; that is to say, higher MSE values are generated when missing points correspond to extrema. The smaller the number of pixels on which results are evaluated, the higher the weight that extrema can have, if randomly set as missing, per assessment sample. As more pixels are included in the evaluation, the influence of extrema or near extrema points on results becomes milder. Then, as the included noise further increases, the MSE of the triangular kernel interpolation also does so.

Finally, the polynomial kernel interpolation presents lower MSE values throughout the evaluation range as expected, as both curves and extremum regions can be better approximated. In particular, these facts can be observed in the low and high noise regions in comparison to the results obtained with the other two interpolation methods. It must be reminded, however, that even if a clear trend is observable, the order of the difference between interpolations is very low (10^{-5}). Fig.4.7 gives a visual summary of the obtained results.

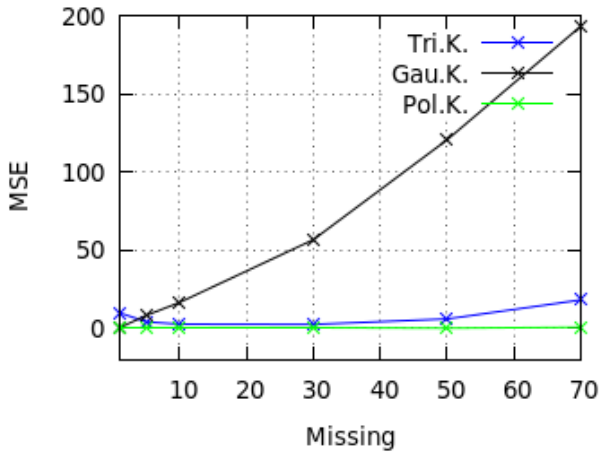


Figure 4.7: Average MSE (10^{-5}) for different percentages of added random missing information for the three tested interpolations. Averages computed on a total of 7200 tests (2400 tests per interpolation).

Given the rate of increase of the MSE scores obtained with the Gaussian kernel approach, the interpolation method has been dismissed and only the triangular and polynomial kernel interpolations are selected for further testing. Differences between the outputs of both interpolation methods are present, as the triangular kernel method cannot fully approximate curves, thus incurring in a greater error, as explained, in regions including extrema. Only when the distribution of missing information determines long spans without valid data, differences in interpolation results between the interpolation approaches become significant. Fig.4.8 presents synthetic data with large gaps of missing information, and the disadvantages of the triangular kernel interpolation in front of the polynomial approach are made clear. However, the higher order applied with the latter may also generate greater errors when noisy real data is used for interpolation, in contrast to the ideal synthetic datasets used here. The next section, indeed, analyses the impact of the two approaches based on classification tasks performed on interpolated real data.

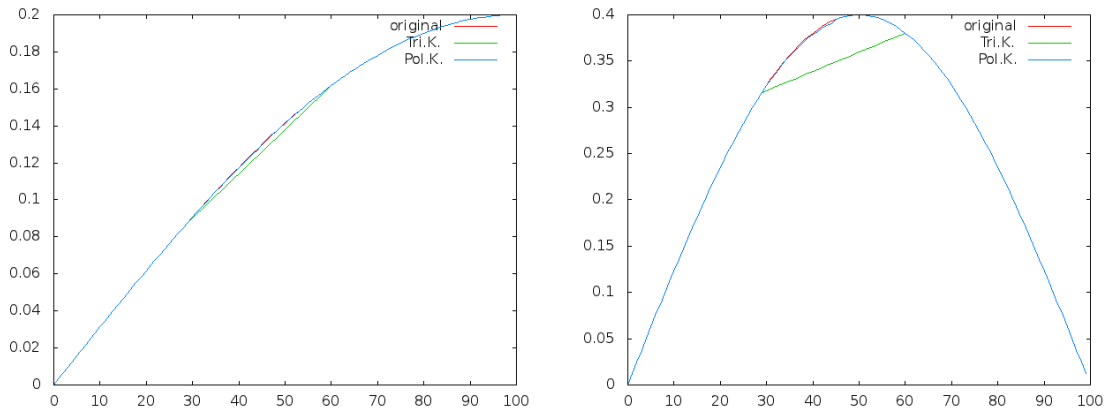


Figure 4.8: Examples of original curves (in red), and Tri.K and Pol.K. interpolations (in green and blue, respectively).

4.4.3 Impact on classification

The feasibility of the use of the two selected variants for interpolation for missing information processing has been evaluated by analysing its impact on classification results in two different ways: over single Landsat tile areas (with two datasets of $12100km^2$) and over an extended Landsat area comprising several tiles (with a total area of $84000km^2$) requiring temporal sampling uniformisation in order to allow joint processing. The following subsections analyse classification results for different approaches based on these different real data datasets.

Firstly, an SVM model learning algorithm allowing subsequent classification is run on unprocessed images, and on unprocessed images together with invalid information masks generated as described in Chap.2. Although it is a common approach apparently providing fairly good classification results, it does not allow for generalisation and it is not applicable for large area land cover generation for two reasons: (a) learning is based on the available dates for a given area, and thus it cannot in any way be applied on other regions with different acquisition dates; (b) learning is based not only on valid information but also on the specific error of the source time series, i.e. missing or corrupted information, and thus the model cannot be applied on other regions with different cloud, cloud shadow and missing information patterns.

Secondly, it is shown how the proposed approach allows for generalisation and overcomes the enumerated problems. The two selected variants for interpolation implemented and compared, both with and without the application of regular resampling, demonstrating that the approach is feasible for large area land cover generation. Finally, variables affecting classification results are analysed, and automatic classification over a very large and heterogeneous area is undertaken.

Two factors need to be taken into account for result interpretation. On the one hand, the available ground truth constitutes an imbalanced dataset with great differences in the number of samples for different classes. In such cases, the application of random sampling for sample extraction and of SVMs for classification results in classification biases towards the classes with larger training sizes due to equal penalties of misclassification for training samples. Thus, in order to undertake learning with SVMs and avoid bias, two strategies are possible, either balancing the training data so that classes are equally represented in the training set by either undersampling or oversampling depending on the class, or applying a weighted SVM so that margin errors are weighted differently. The first of these approaches has been applied here.

On the other hand, regarding the nomenclature, there is some variation in the number of classes available for the different tests here summarized. Based on the nomenclature of 22 classes described in Sec.2.9, different configurations are used in the results presented in this section, with different numbers of classes, depending on the test area, studied variables, and comparable result sets. This is due to the fact that the number of classes that are actually available differ from region to region, and also due to the fact that, for a given number of samples, it is necessary that enough pixels are available not only for sampling but also for validation with the remaining pixels of each class, as non-overlapping sets are used for training and validation, as explained in Sec.4.4.1. For example, if 500 samples are selected per class in order to build the training set, and for a given class only 300 pixels are available, then not enough pixels are available, and the class is not considered for validation. Another example is the use of a refined dataset to evaluate the impact of ground truth accuracy

on classification results, in comparison to a classification using a non-refined ground truth. Then, not all classes are available with enough pixels for validation over the same area. Two solutions are possible, either reducing the number of pixels per class, or eliminating the concerned classes (approach that has been applied here).

4.4.3.1 Common approach

Two classifications not applying missing data processing have been generated as a reference for test tile L930463; the first of them using extracted indexes without missing data treatment, and the second one applying extracted indexes with mask information as part of the feature vector. A reduction of the original 22 classes to 14 classes has been necessary in order to allow comparability with further tests with a more limited ground truth (presented in Sec.4.4.3.2). Results are summarized in Tab.4.4.

Inds	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	51.171	82.497	65.243	75.223	81.748	55.021	63.971	88.112	90.695
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	44.307	43.879	67.196	79.577	88.076	73.009	0.556	0.662	
Inds+Msks	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	50.855	81.792	68.559	76.067	82.591	60.999	62.636	88.799	91.658
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	51.372	46.807	70.107	79.176	92.911	74.209	0.637	0.675	

Table 4.4: Percentages of well-classified pixels per class for approaches not applying interpolation, without temporal virtual point resampling and with a non-refined reference dataset (tile L930463). Average results for 20 independent classifications of 3665x3665 pixels are given. Inds: approach using only indexes. Inds+Msks: approach using indexes and masks. Av: average accuracy, stdev: standard deviation, k: kappa statistic. Nomenclature with 14 classes. See Tab.2.10.

The first of the approaches in Tab.4.4 uses all extracted indexes for a number of acquisitions, without any mask use, which means that the classifier learns with and classifies samples that are contaminated by noise, either saturated by clouds or darkened by cloud shadows, and with missing information due to regions outside the satellite coverage area. The second one applies also missing information masks in order to build the feature vectors for training and classification, that is to say, for a contaminated time series, the corresponding masks indicating the presence of missing or noised data are also given to the classifier.

Average overall accuracy is slightly higher for the second approach (74.209% in comparison to 73.009%), as well as the κ metric (0.675 in comparison to 0.662), but showing a higher standard deviation in the results (0.637 instead of 0.556). For both approaches, the worst accuracies correspond to mixed land cover classes ("other summer crops" and "other winter crops") and a related winter crop ("wheat"), which can be explained by the heterogeneous nature of the first two of these targeted classes.

Results seem to indicate that the common approach of learning and classifying on noise-contaminated time series is valid, as it provides apparently fairly good classification results. However, in reality the approach fails in two respects. Firstly, learning is based not only on valid

information but also on the specific missing and corrupted information of the source time series, thus not being applicable to other regions with different noise patterns. Secondly, learning is based on the specific training dates available for a region, and cannot be applied on other regions with different acquisition dates.

In the next subsections, it is shown how the proposed approach can overcome these problems and permit generalisation: on the one hand, local temporal interpolation minimises the presence of missing data affecting the learnt model; on the other, the application of interpolation allows in turn temporal resampling in a regular grid and application of the model to heterogeneous time series corresponding to different areas.

Local temporal interpolation to eliminate missing data
Results for classifications based on the two selected implemented approaches for missing information processing described in Sec.4.3 are given in Tab.4.5.

Tri.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	53.103	83.418	66.982	75.765	84.098	53.705	64.287	87.441	83.206
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	46.540	41.184	62.802	69.975	77.277	69.643	0.655	0.624	
Pol.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	49.745	80.823	65.689	75.797	82.589	51.237	62.267	85.836	82.593
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	46.799	41.434	62.107	68.682	78.113	68.137	0.801	0.606	

Table 4.5: Percentages of well-classified pixels per class for each of the compared approaches, without temporal virtual point resampling and with a non-refined reference dataset (tile L930463). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 14 classes. See Tab.2.10.

In general, a decrease in average overall and class accuracy can be observed with the triangular kernel and polynomial kernel approaches (with 69.643% and 68.137%, respectively, κ being 0.624 and 0.606). The fact that overall accuracy remains higher with the reference cases seems to indicate possible overfitting in the classifier, which can occur when there is correlation between artifacts and classes (for example, the occurrence of clouds over a forest area); in other words, the constructed model seems to describe noise as well as underlying relationships, which in turn produces better results when the learnt model is applied on the same area.

Furthermore, this problem can generally be reduced with a proper tuning of the regularization parameter C for linear SVMs and, for non-linear SVMs, the kernel and the kernel parameters. Here, a linear SVM has been applied, and no optimization of the parameters, as the objective of testing is not to yield the best possible classification, but in fact to compare results in equal conditions of testing in order to estimate the quality of the implemented missing data processing approaches. The triangular kernel method obtains a slightly higher average overall accuracy (69.643%) with a smaller standard deviation (0.655) than the polynomial kernel solution (68.137% and 0.801). Both methods prove to be feasible approaches to perform signal reconstruction and to reduce missing data in satellite time series, allowing SVM classification based on the interpolated series with good performances in the substantial standard of strength of κ. Also, they produce successive analytic

expressions for complete time series, which in turn is applied for temporal resampling.

Temporal resampling in a unified grid

The selected methods after interpolation assessment discussed in Sec. 4.4.2 have been tested for comparison after applying temporal resampling in a unified grid. The possibility of obtaining an analytical formulation for temporal gaps around non-corrupted data, allows for the selection and estimation of new virtual temporal points according to the temporal reconstruction achieved by each method. In this way, the constructed vector of features is not based on the input acquisition dates, but on a set of regularly-spaced virtual dates. This will allow in turn the application of the classification algorithm over large areas, where the acquisition dates differ from region to region, as a regularised set of virtual dates can be used throughout. Tab.4.6 summarizes classification outcomes by providing average results of all performed tests on a regularised version with 12 time points of the input feature vectors used in Tab.4.5 for the same region area.

Tri.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	50.866	83.815	62.559	77.019	84.437	48.949	65.300	87.872	83.316
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	45.942	40.902	62.404	70.841	69.927	69.509	0.640	0.622	
Pol.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	49.363	82.725	62.919	76.581	83.068	47.0437	63.514	85.938	81.159
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	45.495	39.32	59.463	69.691	71.208	67.615	0.791	0.601	

Table 4.6: Percentages of well-classified pixels, per class, for each of the compared approaches, with temporal virtual point resampling and a non-refined reference dataset (tile L930463). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 14 classes. See Tab.2.10.

Both implemented methods safely obtain overall accuracies just slightly below the non-regularised case presented in Tab.4.5 (69.509% and 67.615% (with κ of 0.640 and 0.791), for triangular and polynomial kernel interpolation, respectively). Both methods are based on local interpolation approaches admitting a varying radius, which is advantageous in relation to approaches using a fixed radius, for which calculation of virtual points is restricted to cases where sufficient amounts of non-contaminated data is present in the set temporal range, and also with respect to approaches using average or weighted average methods, such as the Gaussian kernel interpolation, as shown in Sec.4.4.2, which are heavily affected by gap size. Indeed, presented results are based on the use of not only cloud and cloud masks, but also of snow masks, in order to to process pixels affected by snow as missing information. As snow presence can span for several months, the adaptability of the approach for the processing of large gaps of missing information is shown.

Thus, results demonstrate the feasibility of temporal resampling with virtual points for the targeted satellite time series datasets, and also that of the application of triangular and polynomial kernel interpolation approaches for classification over large areas where satellite acquisition dates differ from region to region and missing information gaps can be of a significant size.

4.4.3.2 Evaluation and implementation aspects

This section analyses the following topics relative to the implementation of the proposed approach: firstly, the impact of ground truth accuracy on results, which needs to be taken into account for the interpretation of results; secondly, the sensitivity of the method to the number of chosen virtual points, which can serve as a guide for implementation; thirdly, the effects of the number of training samples, for result analysis; and finally, the impact of using non-spectral information to enhance classification, as a further implementation aspect.

Ground truth accuracy

In order to evaluate the impact of the accuracy of ground truth on results, classifications presented above have been regenerated using a refined ground truth dataset. Procedures to obtain this improved reference include erosion of available polygons in order to minimise misregistration error, and operator selection of samples having an expected temporal profile. Results with and without temporal resampling in a uniform grid are given in Tab.4.7 and Tab.4.8, respectively.

Tri.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	67.658	93.631	89.380	93.114	88.284	71.054	86.346	67.659	92.385
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	64.056	71.542	76.074	78.385	86.356	80.457	1.749	0.747	
Pol.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	66.173	92.430	88.062	94.213	96.738	69.653	84.827	97.805	92.272
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	66.181	68.423	75.776	78.829	86.687	81.482	0.608	0.760	

Table 4.7: Percentages of well-classified pixels per class for each of the compared approaches, without temporal virtual point resampling and with a refined reference dataset (tile L930463). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 14 classes. See Tab.2.10.

Tri.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	66.130	93.851	85.659	93.675	97.612	68.358	86.685	98.295	91.143
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	61.977	67.126	72.491	80.328	77.831	80.943	0.534	0.755	
Pol.K.	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	HMO
	65.071	92.679	86.473	93.718	96.608	63.977	86.199	97.798	89.828
	OWC	OSC	DBL	PNL	MFO	av	stdev	κ	
	62.865	65.117	70.919	79.276	79.217	79.590	0.443	0.738	

Table 4.8: Percentages of well-classified pixels per class for each of the compared approaches, with temporal virtual point resampling and a refined reference dataset (tile L930463). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 14 classes. See Tab.2.10.

Average overall accuracy is greatly improved for the two selected approaches: the linear interpolation case ameliorates from 69.643% to 80.457% without temporal resampling, and from 69.509% to 80.943% in the regularised case; for the polynomial kernel interpolation case, average amelioration includes 68.137% to 81.482% and 67.615% to 79.590%, respectively. This indicates that a considerable amount of error is embedded in the ground truth used for this work. Also, the

refined reference set, with a reduced in number of pixels, prevents the inclusion of some classes, as they do not have enough pixels for sampling and validation. This fact has determined, as explained, the reduction of the original 22 classes to 14 classes, which has also been extended to the non-refined dataset, in order to maintain comparability.

In conclusion, the triangular and polynomial kernel interpolations methods perform similarly, although with temporal resampling the linear interpolation algorithm performs slightly better (80.943% against 79.590%), but with a larger standard deviation (0.534 instead of 0.443). The triangular kernel has thus been retained for tests in the following subsections.

Number of virtual time points

The impact on classification results of the number of virtual time points used by the temporal resampling task is analysed. With a time series spanning a year, the following approaches have been tested: bimonthly (6 time points), monthly (12) and biweekly (24) sampling. First and last dates considered remain the same as in the rest of tests, fixed as 01/10/2009 and 30/09/2010. This corresponds to the use of a crop calendar year, as crops can rotate from one crop year to another. Results are given in Tab.4.9. Here as well, average overall accuracy increases with the number of virtual time points, and standard deviation decreases. The best results are thus obtained with 24 time points, with an average overall accuracy of 61.360% and a standard deviation of 0.760.

#	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
6	48.444	63.809	49.161	71.936	73.150	80.725	48.692	48.630	76.570	84.539	92.960
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	81.504	64.346	50.702	27.089	65.028	56.762	60.124	56.653	56.971	0.887	0.500
#	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
12	51.331	70.793	57.415	76.396	77.845	85.127	53.006	50.997	83.804	85.327	93.771
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	81.114	74.953	52.824	35.020	71.939	59.797	62.042	59.063	60.142	0.849	0.534
#	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
24	52.782	73.528	60.033	77.195	78.842	87.283	57.050	52.764	88.000	86.235	93.731
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	82.186	76.846	53.263	40.255	75.997	61.083	62.718	60.879	61.360	0.760	0.547

Table 4.9: Sensitivity to the number of time points used for temporal virtual point resampling. Percentages of well-classified pixels, per class, for each of the compared approaches: 6, 12, 24 and 48 pixels per class, with 500 training samples per class, proposed approach with virtual point resampling, with a non-refined reference dataset (tile L930563). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 19 classes. See Tab.2.10.

Training samples

The sensitivity of the approach to the number of training samples has been tested for Landsat tile L930563 as an aspect additional to the choice of interpolation approach, but with a relevance on results. The observable impact of sample numbers on classification results provides insight on the interpretation of results presented in this work, based on a selection of 500 samples per class, where relative improvements achieved by the described approaches are to be considered. Tests have been undertaken here for 16 classes, in order to assure the presence of sufficient validation pixels for all classes throughout the comparison for 200, 500, 1000 and 1500 pixels per class. Results are given

in Tab.4.10.

#	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	IMP	HMO
200	49.017	70.364	53.440	73.812	77.132	53.217	51.316	83.090	92.860	79.411
	SMO	OWC	OSC	DBL	PNL	MFO	Av	Stdv	κ	
	73.354	51.801	34.002	54.403	58.791	56.730	57.256	0.803	0.504	
#	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	IMP	HMO
500	52.169	72.569	58.154	77.074	79.123	57.816	53.736	85.740	94.526	81.255
	SMO	OWC	OSC	DBL	PNL	MFO	Av	Stdv	κ	
	76.293	52.846	38.302	60.213	61.730	59.249	60.718	0.895	0.540	
#	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	IMP	HMO
1000	53.785	73.326	60.848	78.066	80.004	59.863	54.502	86.953	95.328	82.202
	SMO	OWC	OSC	DBL	PNL	MFO	Av	Stdv	κ	
	78.644	53.875	41.185	62.768	63.495	60.399	62.398	0.547	0.556	
#	WHE	COR	BAR	RAP	SFL	PHC	TMF	WAT	IMP	HMO
1500	53.515	73.772	62.051	78.762	80.384	60.465	55.724	87.073	95.921	82.807
	SMO	OWC	OSC	DBL	PNL	MFO	Av	Stdv	κ	
	78.949	54.660	42.230	63.443	64.518	60.981	63.006	0.430	0.562	

Table 4.10: Sensitivity to the number of pixel samples used for training. Percentages of well-classified pixels, per class, for each of the compared approaches: 200, 500, 1000 and 1500 pixels per class, proposed approach with virtual point resampling, with a non-refined reference dataset (tile L930563). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 16 classes. See Tab.2.10.

It can be observed that average overall accuracy increases with the number of pixels per class used for training, the best accuracy being obtained with the sampling of 1500 samples per class (63.006%), as well as the smaller standard deviation (0.430).

Non-spectral information

An aspect also additional to the choice of interpolation approach, the impact of the use of non-spectral information for classification has also been tested. The effects of using non-spectral information to enhance classification are examined with the incorporation of height and height-derived information. Using Landsat tile L930563, the following approaches are used: multispectral information (MX) + height information (H); MX+aspect information (A); MX+ slope information (S); and finally a combination MX+H+A+S. Results are given in Tab.4.11. Here as well average overall accuracy increases with each of the incorporations, attaining the highest accuracy with the last of the approaches (61.639%) and also the lowest standard deviation (0.723).

4.4.3.3 Extended area dataset

The impact on classification results with the compared approaches is evaluated on a second dataset comprising the whole study area of $84000km^2$. In this case, different regions present disparate acquisition dates. In the non-resampled version of the proposed approaches, the feature vector used for classification is made up of the values of primitives at the different temporal points determined by the acquisitions. With the different regions configuring the targeted area having dissimilar acquisition dates, the application of a model learnt on one region for classification on the rest of regions is simply not possible. Alternatives here include either (a) training and classifying each of the regions separately, or (b) applying the proposed resampling into a unified temporal grid so that

#	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
MX	51.331	70.793	57.415	76.396	77.845	85.127	53.006	50.997	83.804	85.327	93.771
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	81.114	74.953	52.824	35.020	71.939	59.797	62.042	59.063	60.142	0.849	0.534
App.	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
MX+H	51.693	71.551	58.251	76.804	77.945	86.645	61.359	51.866	88.598	85.553	93.868
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	85.425	79.667	52.884	38.520	76.177	62.153	61.566	70.399	61.462	0.768	0.548
App.	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
MX+A	51.117	70.803	57.283	76.384	77.668	85.034	53.128	51.104	84.467	85.274	93.665
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	81.146	75.739	52.844	34.967	71.773	59.578	62.249	59.065	60.050	0.828	0.533
App.	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
MX+S	51.384	71.126	57.598	76.480	77.885	85.579	54.415	53.043	85.813	85.755	93.798
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	81.740	77.870	52.808	35.184	75.568	61.353	61.695	62.509	60.874	0.841	0.542
App.	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP
MX+HAS	51.643	72.090	58.155	76.812	77.852	86.840	61.998	53.206	88.804	85.776	93.790
	HMO	SMO	OWC	OSC	FNO	DBL	PNL	MFO	Av	Stdv	κ
	85.572	80.890	52.892	38.685	78.074	62.367	61.872	70.933	61.639	0.723	0.550

Table 4.11: Enhancement of separability between classes by additional non-spectral information addition. Percentages of well-classified pixels, per class, for each of the compared approaches (App.): MX (multispectral), H (height), A (aspect), S (slope). Proposed approach with virtual point resampling, with a non-refined reference dataset (tile L930563). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 19 classes. See Tab.2.10.

training can be attained using information available on the whole area, and then classification can be undertaken for each of the regions. Option (b), the application of resampling, has been favoured for a variety of reasons:

1. The strategy allows for the reduction of the algorithm complexity that would be required to apply different learning tasks and classification tasks in different regions.
2. The use of an region-by-region approach would be constrained by the available ground truth, as some classes are not represented in particular regions. This would prevent their appearance in these particular regions in the final land cover map.
3. The spectral variability of the targeted classes across regions would not be captured with option (a), but only the local properties of classes for the considered region.
4. Different classifiers would classify path border areas differently, for the stated reasons, and no uniformity would be achieved in the output land cover map.

Tab.4.12 shows results achieved for the whole targeted area using the proposed approach with a local linear interpolation algorithm, and with the polynomial kernel algorithm. All 22 classes in the nomenclature of Sec.2.3 are present, and a sampling of 1500 pixels per class has been applied, determined by the available ground truth numbers.

Two factors are to be taken into account when analysing results in comparison with those presented previously: more classes are present, which allows more room for confusion, and less

Tri.K.	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP	HMO
	51.888	80.267	59.754	76.633	78.378	86.505	33.309	51.9	77.302	82.991	92.962	71.303
	SMO	OWC	OSC	FNO	DBL	PBL	DNL	PNL	MFO	RMO	av	κ
	58.849	52.621	44.774	56.761	43.158	82.421	66.915	59.014	50.2	86.293	0.560	0.502
Pol.K.	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	VIN	WAT	IMP	HMO
	48.659	79.041	56.15	74.743	77.701	84.429	31.046	49.543	75.345	80.478	90.881	69.069
	SMO	OWC	OSC	FNO	DBL	PBL	DNL	PNL	MFO	RMO	av	κ
	56.312	52.22	37.79	52.922	38.238	81.034	66.915	57.982	51.058	82.416	0.532	0.474

Table 4.12: Percentages of well-classified pixels, per class, for the linear interpolation approach without snow interpolation, with temporal virtual point resampling and a non-refined reference dataset (tiles L930363, L930463, L930563, L930663, L930362, L930462, L930562 and L930662). Nomenclature with 22 classes. See Tab.2.10.

pixels per class are selected for training (1500 pixels per class are used for eight regions, while 500 pixels per class were used before for one region), an undersampling that has an impact on classification results, as previously explained.

Results obtained by applying triangular kernel interpolation are superior to those using the polynomial kernel algorithm also with the second dataset, both for overall accuracies (0.56 in comparison to 0.53) and for most land cover classes when considered individually. The polynomial kernel algorithm seems to closer mimic phenological curves with synthetic data than its linear counterpart, as it applies least squares fitting with a polynomial of degree two. However, this is a fine gain in relation to the interpolation obtained with a local linear method. On the other hand, the algorithm fails faster when clouds, cloud shadows or snow are undetected and with the presence of different types of noise, due to the fact that calculations are based on a higher polynomial degree. For the given datasets, the linear approach performs slightly better as a local interpolation method for further temporal resampling. For this reason, this method is used in the rest of this work.

In general, these results should not be considered as the best attainable results but, more in the opposite sense, as an indicator of the potential of the approach, due to the many constraints that have conditioned the testing. Firstly, a high number of classes tends to yield lower accuracy classification results than a smaller number of classes. Many approaches in the literature are based on only a few classes, and thus accuracy results are not comparable. Secondly, the number of training samples is extremely low (under 0.07%). In this respect, much higher ratios tend to be used, of the order of 20 to 30%. With higher numbers of training pixels, higher accuracies can be obtained, as previously shown, although with longer computation times. In addition to this, the choice of a linear SVM kernel without optimisation determines faster computation, with the penalty of lower accuracies; higher accuracies can be obtained with for example an RBF kernel and optimisation. With the characteristics of the Sentinel-2 series, higher resolutions and the corresponding decrease of spatial mixing may probably translate, as well, into higher classification accuracies with the proposed method.

As explained, the method remains flexible for usage with different nomenclatures and time series, allowing for the simple integration of other indexes depending on target land cover. In addition to this, this flexibility is also relative to diverse sorts of known noise or missing information, as it is applicable to different problems such as snow presence (as it has been demonstrated) and other types of missing valid information. For example, airplane trails visible in optical imagery

may be detected using temporal information and line detection or another approach. The derived masks may be easily incorporated to the method in the same way clouds, cloud shadows, snow and border information are applied here.

Finally, Figs.4.9 to 4.14 show the resulting average temporal profiles for each of the targeted land cover classes after the application of the proposed approach with the triangular kernel interpolation method on the example NDVI series discussed in Chap.3. The corresponding profiles based solely on available valid information are also shown for comparison.

4.5 Conclusions

The problem of missing information for large-area land cover mapping has been analysed, and an overview on the topic of cloud and cloud shadow detection and missing information processing in the literature has been given.

A multi-temporal method to produce temporally and spatially continuous HR time series for large-scale processing has been presented. It has been devised for the automatic generation of land cover maps over large areas with great data volumes such as those to be provided by the twin-satellite mission Sentinel-2. It overcomes the difficulties posed by the common problem of missing information, generally due to the presence of clouds, cloud shadows, snow or satellite passage borders, as well as that of temporal irregularities due to the existence of different acquisition dates for different regions over large areas.

The methodology is based on the application of local temporal interpolation, which in turn allows for analytical expression formulation and thus for regular temporal sampling for feature extraction over regions with different acquisition dates. Based on the characteristics of the targeted remote sensing time series, three missing data processing algorithms using variable radiuses have been implemented and compared: triangular, Gaussian, and polynomial kernel local temporal interpolation. Best results have been obtained with the first of these methods. Also, it has been shown how local interpolation approaches allow for the application of temporal resampling strategies in order to obtain temporally and spatially continuous time series that can be processed by standard classification methods over large regions.

The application of local temporal interpolation for missing information processing has been analysed in terms of the impact on classification results generated over two large areas of $12100km^2$ presenting temporal irregularities and both missing and cloud-related noisy information, and over an extended area of $45100km^2$, characterised also by a markedly larger temporal irregularity of acquisitions. In this way, the potential of missing data processing with local interpolation methods has been demonstrated, and in particular their applicability for the operational generation of land cover maps over large areas.

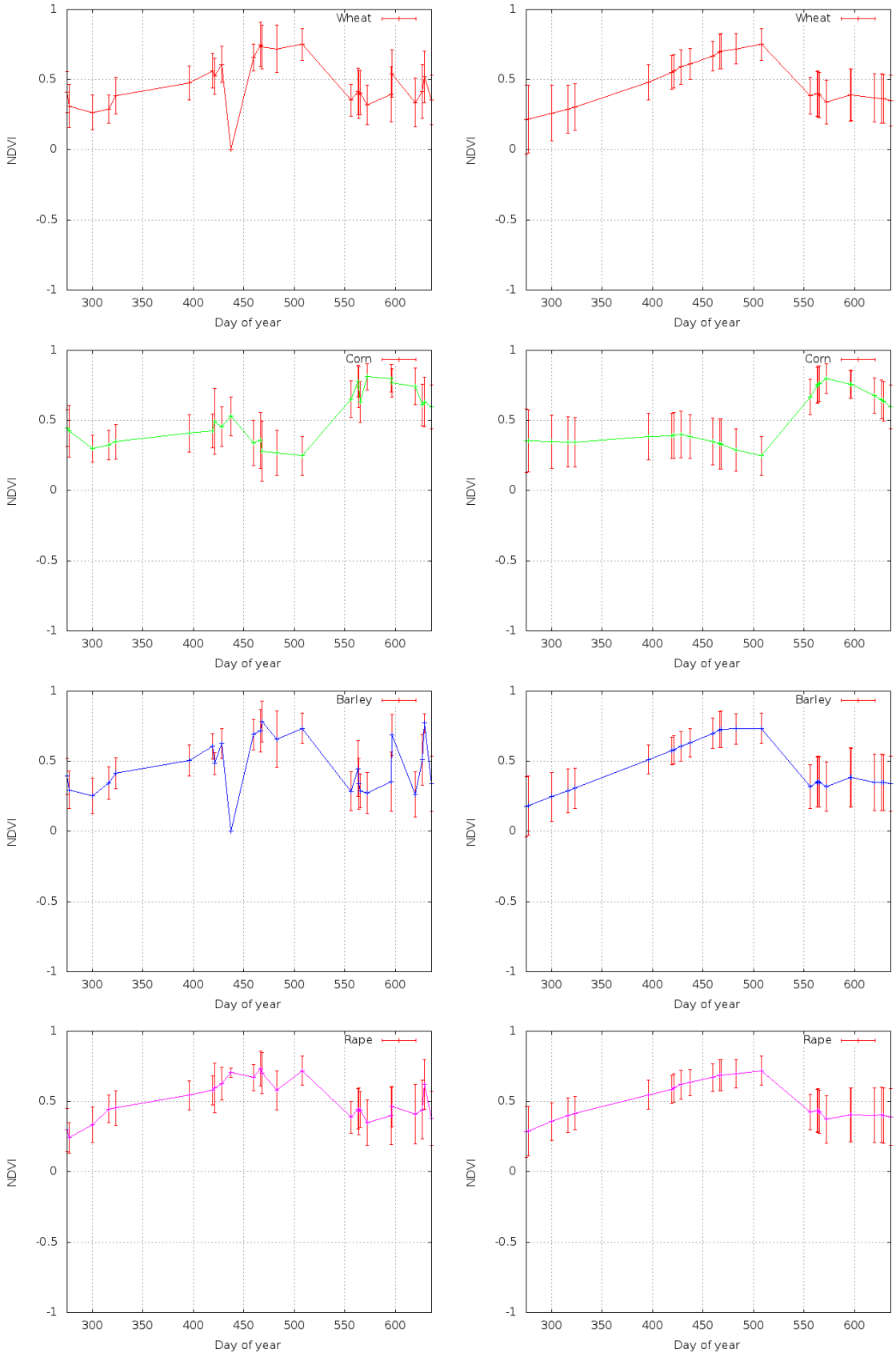


Figure 4.9: Time series spectrum-derived information for the targeted land cover classes as shown in Sec.3.3.4 (left column) and after application of the proposed approach with the triangular kernel interpolation method (right column). NDVI average values and variances for each class correspond to the period Oct 2009 - Sep 2010 (day of year is referenced to Jan 2009).

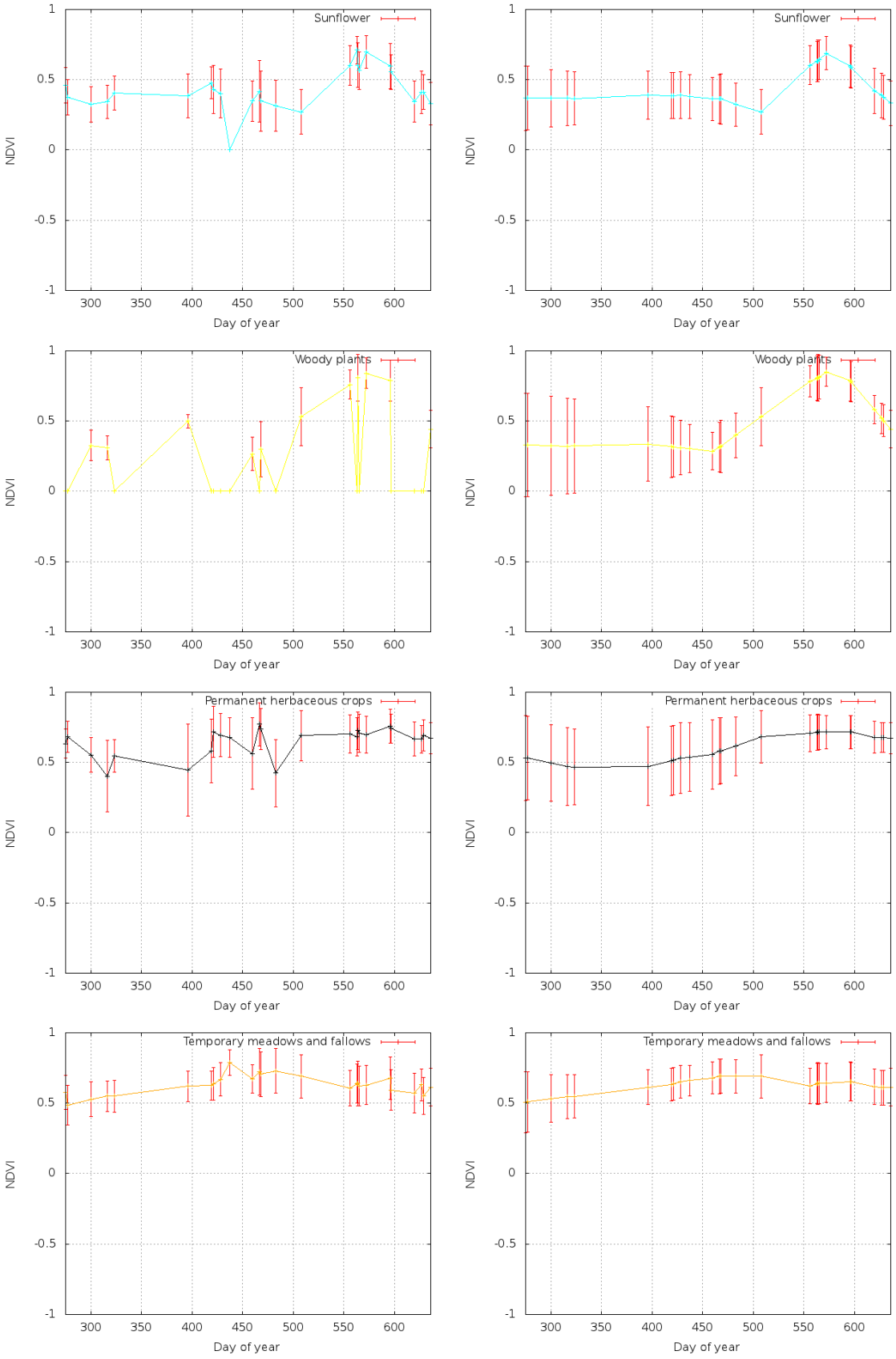


Figure 4.10: As in Fig.4.9.

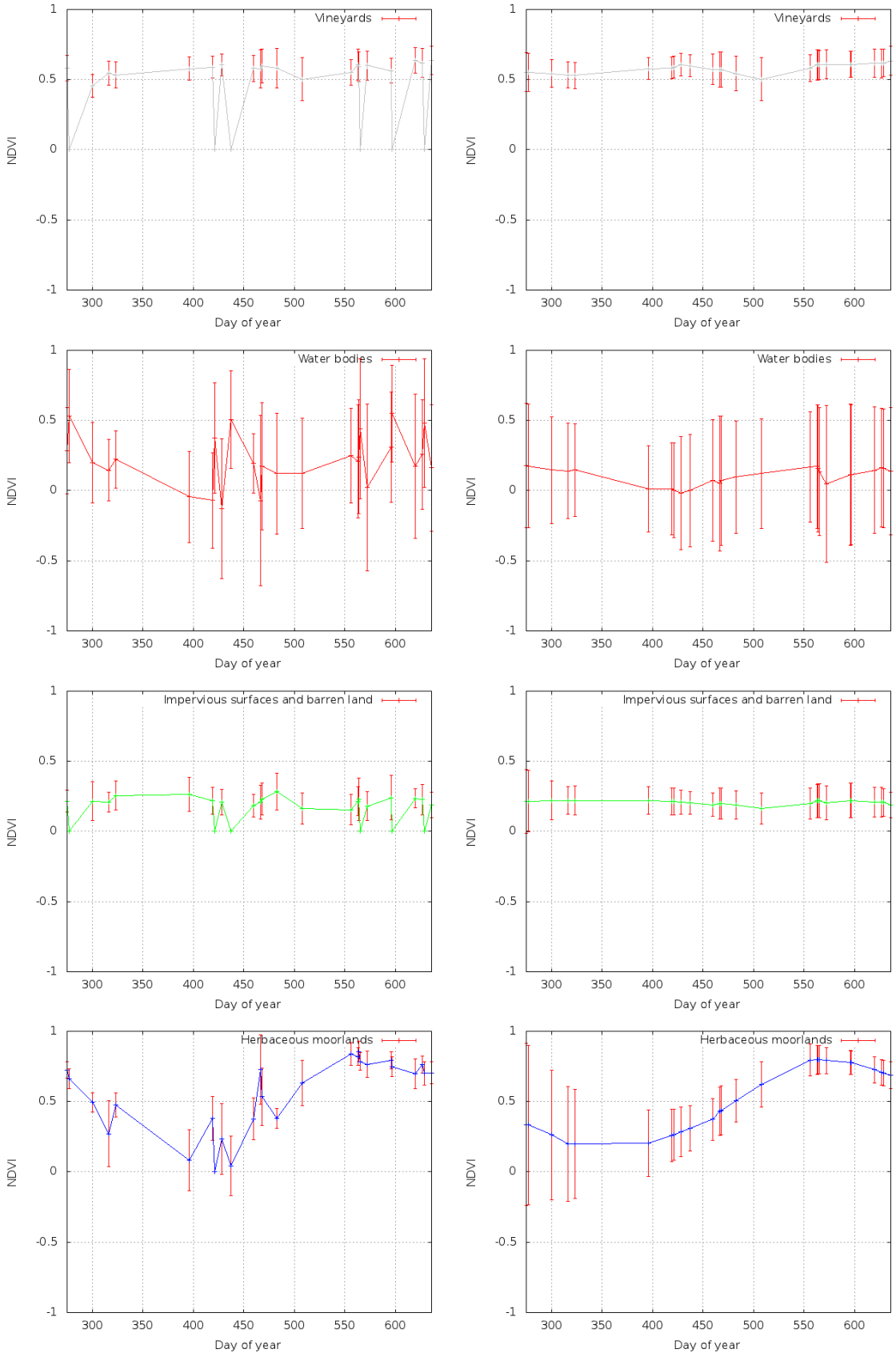


Figure 4.11: As in Fig.4.9.

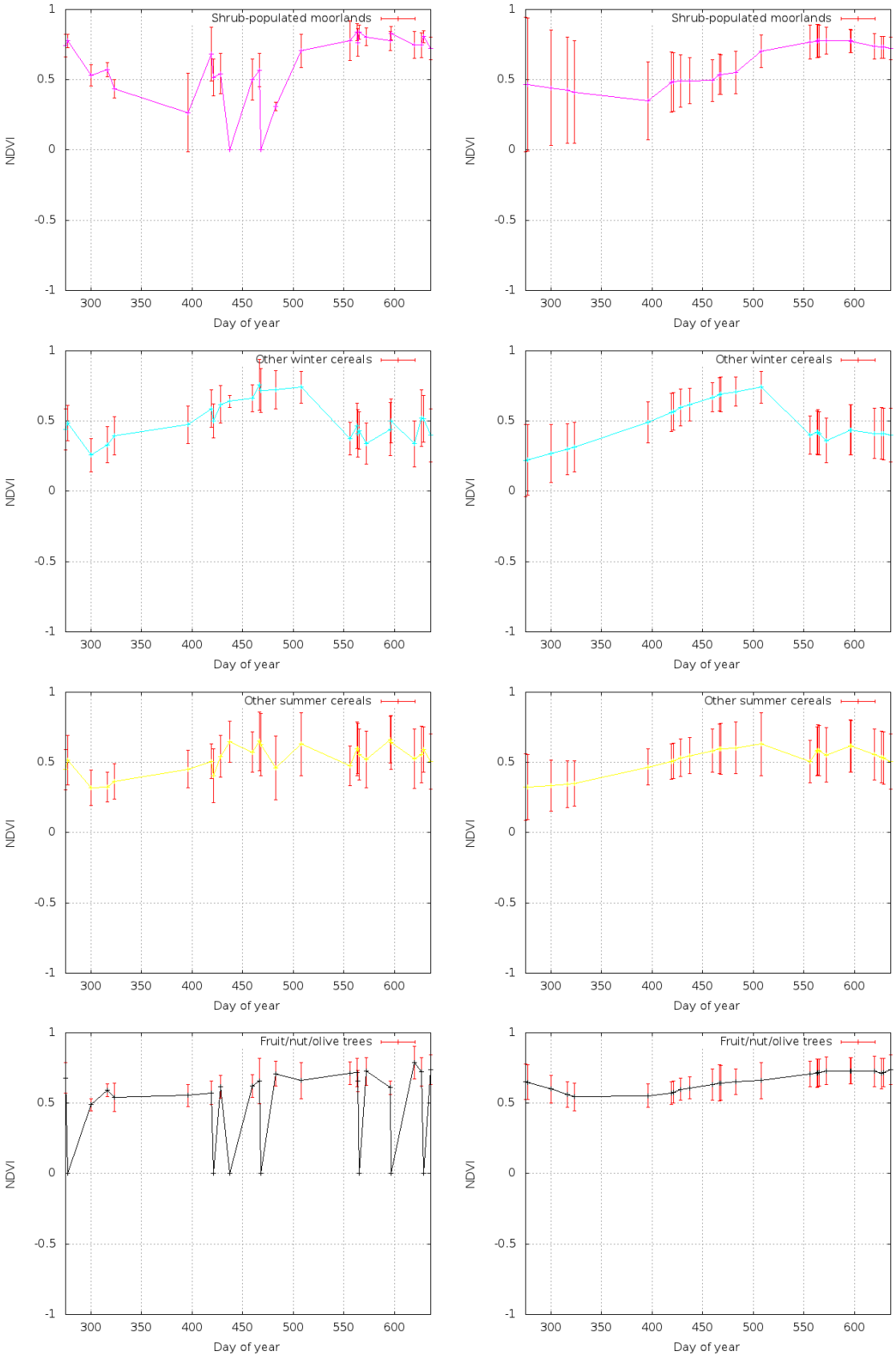


Figure 4.12: As in Fig.4.9.

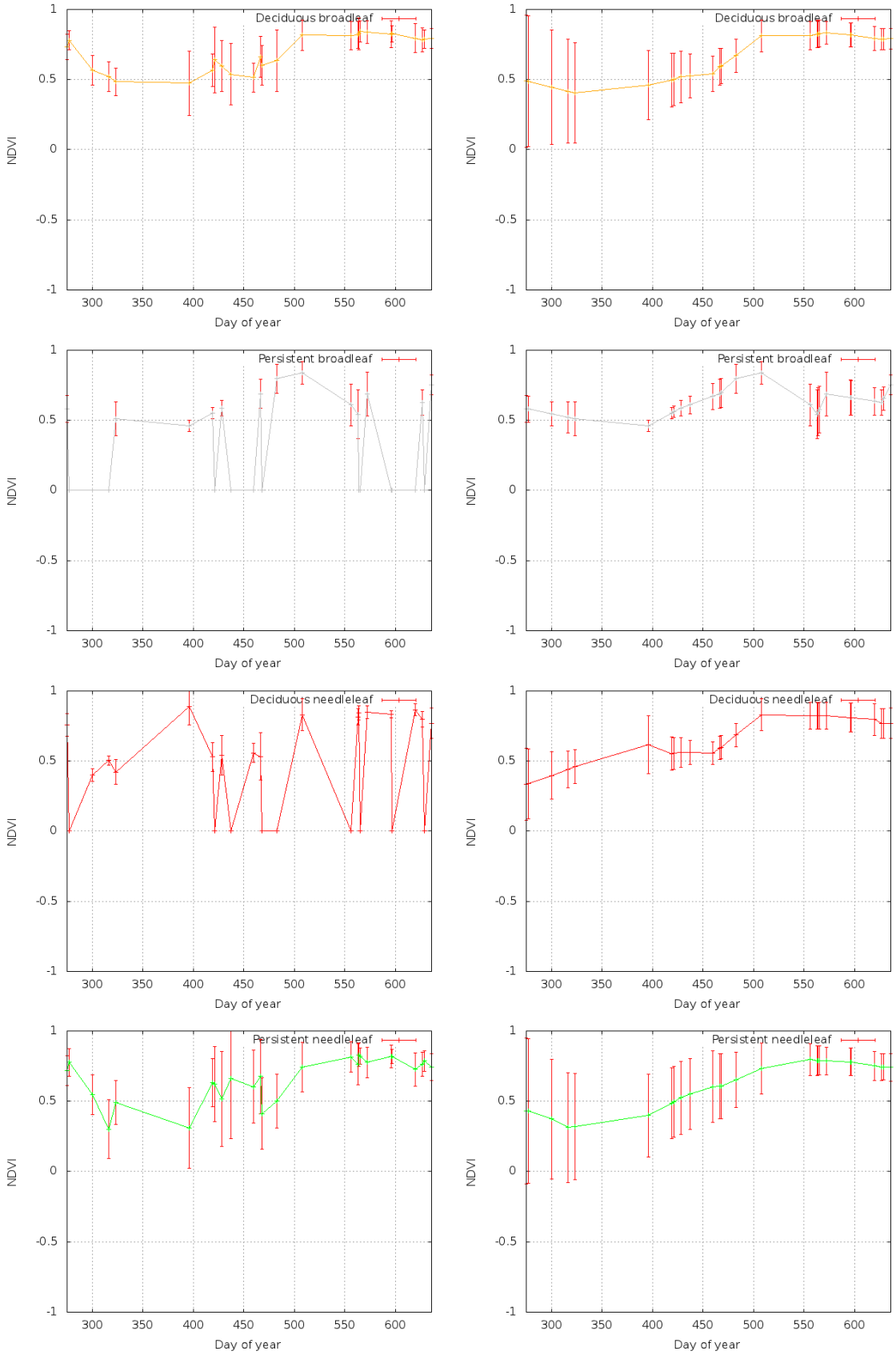


Figure 4.13: As in Fig.4.9.

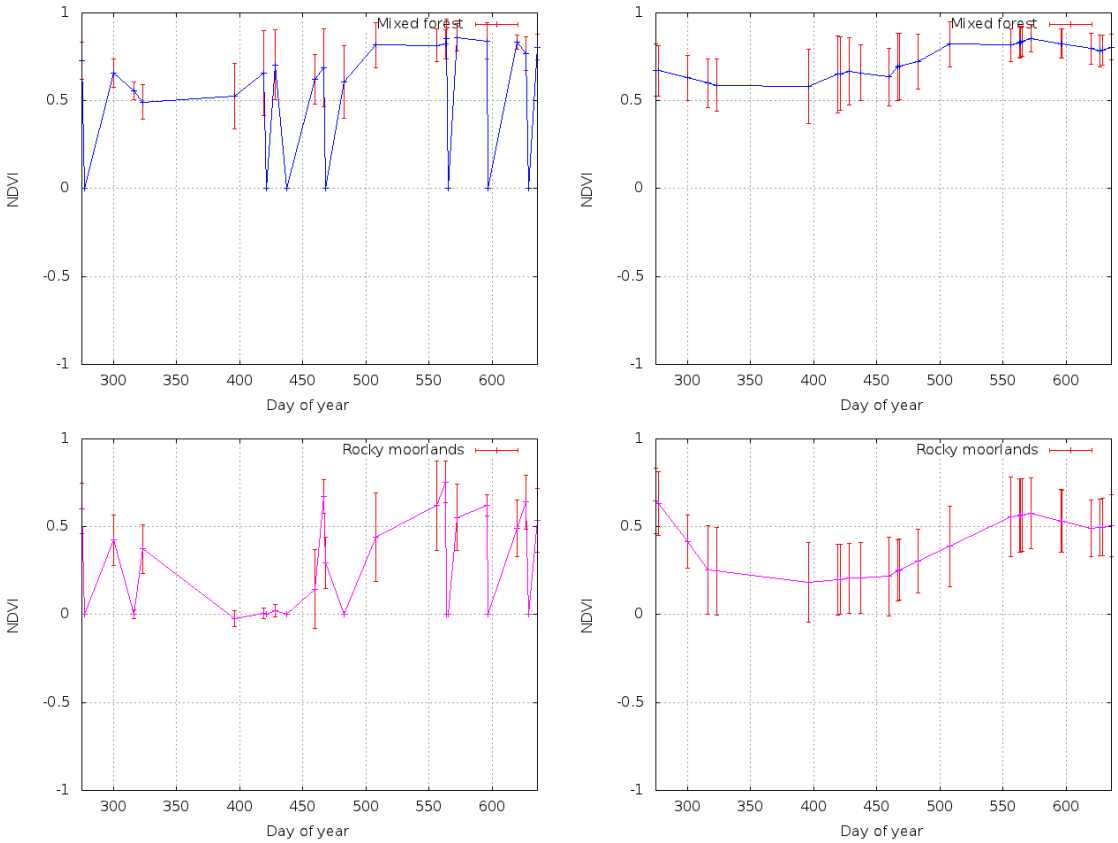


Figure 4.14: As in Fig.4.9.

Chapter 5

Spatial sampling strategies for large area land cover classification

Contents

This chapter introduces the problem of data variability and volume, which are commonly encountered for land cover mapping with satellite data time series when large areas are targeted. A review on the state of the art for the processing of these issues is offered. Also, an approach targeting these issues is presented, contributing towards the generation of large-area land cover maps with operational times. It is based on the application of automatic spatial sampling strategies on available reference data for the production of low volume and highly representative training sets, allowing for fast machine learning tasks and enhancing classification. The approach is applied in combination with the temporal sampling strategies described in the previous chapter, offering a methodology that operates in this way in the spatial space for dataset selection, and in the temporal space for missing data processing and virtual timepoint resampling. Different approaches for spatial sampling are compared for the chosen study area, and in particular different stratification approaches are examined, including the incorporation of strata derived from topographic end ecoclimatic information. These approaches are tested and evaluated based on classification results over large regions.

5.1 The problem of learning with great data variability

Supervised classification has been long applied in remote sensing for the production of thematic maps representing land cover, and much research has focused on the applicability and comparison of a number of classification techniques. Another factor which can have a significant impact on classification accuracy is that of the training stage, determining the model which is learned and that is subsequently applied for classification. Indeed, training procedures can have a greater impact on classification accuracy than the selected classification technique [Bartlett et al., 1980].

In broad terms, the training set selection for machine learning aims at providing the data that will most accurately define a model of the classes. This is achieved by providing a representative description of the overall population, fully describing the classes [Chen and Stow, 2002] in order to derive meaningful features that can be exploited for classification. This typically requires, however, complete training data volumes, large enough to account for examples of classes in their full range, and well spread over the study region in order to sufficiently depict spectral response variability [Campbell, 2006] [Foody and Mathur, 2004b].

When dealing with remote sensing data over large areas, encountering class data spectral and temporal variability is unavoidable. This is true for most land cover target classes, specially significant with vegetation classes, and mainly stemming from two factors of different nature: on the one hand, the impact of soil and canopy characteristics variation across territories; on the other, the effect of differing vegetation phenological cycles due to topographic and climatic characteristics.

Regarding the first factor, the variability of soil differences in spectral and temporal response is related to the soil composition itself, which varies in space; its texture and configuration; its conditions of dryness or wetness; and finally the degree to which this soil reflectance is mixed with vegetation reflectance, depending on degrees of vegetation presence and on the type of soil itself, which affects vegetation growth.

Regarding the second factor, the variation of phenological cycles refers to the biological, chemical and physical stages that vegetation types experience periodically, and which typically encompass different phases, such as emergence, rapid growth, a period of maximum leaf area, senescence or harvesting, and dormancy. These cycles are greatly impacted by topographic and climatic characteristics, altering the timing of the phenological cycle of a class and thus the feasibility of identification with remote sensing time series. Many studies for different vegetation species have demonstrated phenological cycle changes depending on air temperature, among other factors. For example, and assuming a decrease of temperature of 0.65°C per 100m of altitude, [Rötzer and Chmielewski, 2001] demonstrated and advance of leave unfolding of Fagus sylvatica (European beech) by 3.2 days per °C, and of the flowering of Prunus avium (wild cherry) by 4.5 days.

These differences in soil characteristics and sparseness of vegetation, and in particular the varying growth responses of vegetation classes depending on geographic characteristics affecting temperature, such as latitude or longitude, are most impactive when large areas are considered. For example, the spread of the new growth in spring in North America was documented by [Dethier et al., 1973], a phenomenon called the “green wave” and observable from south to north in

North America. For different species, the emergence phase started earlier in the south than in the north, appearing progressively in time. In this context, species modelled with the southern samples would not correctly represent the north occurrences. The same effect would be observable with the European beech and wild cherry cases mentioned above: characterization would require samples at different altitudes in order to be representative. Thus, the bigger the area considered, and the bigger the topographic differences, the most impactive these variables can be on intraclass variability and, consequently, on interclass separability when the former is not well represented in the training stage.

As a result, traditional approaches with satellite time series to training data acquisition for supervised classification can significantly fail when applied to very large areas. Training data selection of a common occurrence of a targeted class, based on a common habitat for the class or even as a result of random sample selection, can completely neglect other occurrences determined by different topographic, soil, or climatic conditions. Thus, an intelligent data acquisition for training with a standard learning algorithm such as SVMs may be useful in order to make samples more pertinent and to obtain a more robust learning.

In addition to this, the new spatial missions will provide high resolution imagery (spectral, temporal and spatial data) and thus unprecedented large data volumes, which imposes also the need for an operational methodology allowing for short learning and classification times, in conjunction with the need for an optimised training data acquisition. Thus, the intelligent selection of training samples can give a solution not only to the described variability capture problem, but also to the problem of processing such large volumes of data. It can allow the use of small training sets without loss of discriminatory power, reducing the classifier costs in terms of speeding model learning and classification times.

In summary, the problem of learning with great data variability defines some main necessities. Firstly, a more robust learning is needed, in order to capture the variability that is inherent to large areas and that determines classification feasibilities. Secondly, the intelligent selection of the most pertinent samples for training becomes also necessary in order to manage this variability. Thirdly, the maximisation of the representativity of the selected samples can facilitate the selection of a small number of samples to obtain short computational times without loss of accuracy. Fourthly, the exploitation of external knowledge on physical processes becomes important in order to guide the selection of samples. Finally, automation is also required for the methodology to be operational with the coming satellite series.

Given these necessities, an automatic spatial sample selection method allowing for an intelligent training data acquisition for large area land cover classification with a supervised approach is proposed. It allows to capture the variability found in large areas by providing a representative description of the overall populations. It is based on the inclusion of additional information in order to guide the characterization process and permits a fully automatic selection of training data based on available reference data and topographic or bioclimatic information.

5.2 Background

The theory of sampling has been developed during the last decades and it is now in its maturity. The statistical community has developed a multiplicity of approaches for the sampling or selection of a subset of a population with the purpose of deriving estimates about the characteristics of this population. A variety of methods exist, some being general-purpose, and others being tailored to specific domains. The random sampling technique (further developed in Sec.5.3.2), for example, has been used in a wide variety of fields, from analytical chemistry and biotechnology to Earth sciences and the social and behavioral sciences, to name a few. The diverse nature of sampling methods encompasses probabilistic and non-probabilistic approaches, one-stage or multi-stage techniques, or non-adaptive or adaptive schemes where characteristics such as sampling size are not fixed prior to sampling. The objective of sampling design is the selection of the most appropriate sampling method given the objectives and constraints of the targeted sampling sets.

More particularly, the applicability of sampling methods has been widely investigated for variety of tasks, such as association rule mining, classification or clustering. Notably, attention has been paid to the use of sampling as a solution to the problems posed by large datasets. Indeed, voluminous sets with either a large number of instances, a large number of features, or both, can pose significant obstacles to data mining and machine learning. These imply an increased cost for model fitting or even complications in relation to physical memory limits. In this context, the use of sampling methods can alleviate the problems associated to voluminous data while offering comparable or improved results. Reduced cost and greater speed, scope and accuracy have long been considered advantages of sampling as compared with complete enumerations of large datasets [Cochran, 1977].

When considering classification for the extraction of thematic information from remote sensing imagery, most research has focused on the techniques used for classification rather than on other aspects, such as the nature of the training stage sample selection. However, training data selection may be as or even more important than the selected classification algorithm [Campbell, 2006]. In [Scholz et al., 1979], for example, several classification schemes are evaluated for crop identification, and the importance of the training data selection is established. In addition to this, and when indeed the design of the training sample extraction has been questioned, a statistical view has been traditionally adopted. Focus has been typically placed on the accurate and full spectral definition of classes, targeting statistical representativeness for the whole population [Kuo and Landgrebe, 2002] [Muchoney and Strahler, 2002].

Thus, much of the literature appears to be based on the classical statistical view of the classification process using conventional probabilistic classifiers. However, the different characteristics of classifiers determine different training needs [Foody and Mathur, 2004b]. Also, attention has been placed on training set size estimation with much work on the effects of training size on the accuracy of different classifiers [Cochran, 1977]. Generally, accuracy is typically positively related to training set size [Arora and Foody, 1997] [Foody and Mathur, 2004a], with factors affecting size requirements such as classifier complexity and dataset dimensionality [Foody and Mathur, 2004b].

Training dataset data reduction has been studied from a variety of angles, both focusing on the

training set number of samples and the reduction of the dimensionality of the data. For example, the application of feature selection techniques for dimensionality reduction [Kuo and Landgrebe, 2002], or the use of data spatial dependence information for training dataset acquisition [Atkinson, 1991]. Also, refinement strategies have been commonly applied in order to emphasize the average response of classes, by removing or down-weighting atypical samples. These strategies, focused on class centroids, are useful with some classifiers but not others, and can actually contribute to the reduction of very informative samples for classifiers like SVMs [Foody and Mathur, 2004b].

An alternative point of view is given by [Foody and Mathur, 2004b], based on the recognition that non-parametric classifiers such as SVMs, neural networks or decision trees are often more accurate than conventional statistical classifiers [Arora and Foody, 1997] [Foody and Mathur, 2004a] and on the fact that, not being based on an assumed parametric model, the full description of features with training samples may not be necessary when using them. Additionally, it is not only with non-probabilistic and non-parametric classifiers that the capture of the necessary diversity in the sampling stage is important. Also, the needs of neural networks and Bayesian classifiers are similar, but different to those of SVMs.

With SVMs, for example, a key concern for class discrimination is the identification of spectral responses lying near the decision boundaries partitioning the feature space. To some degree atypical of their own class, these samples are more useful than those in the central core of the class, and thus a reduction of samples taken from this core can help reduce training costs without reducing classification accuracy [Tambouratzis, 2000]. Thus, [Foody and Mathur, 2004b] suggest the use of ancillary information in order to determine the location of regions from which informative training samples can be obtained, so that a small training set can be used for accurate image classification. They test on the separability of 2 classes, assessing the accuracy of each classification on 40 testing pixels per class and conclude that a small training set can be selected based on information value without loss of discriminatory power.

5.3 Proposed method

In the context of the state of the art presented in the previous section, the proposed method allows for the automatic construction of intelligently selected training sets with representative samples in order to enhance model learning and subsequent classification and to manage the problem of data volume in order to ultimately contribute to the fast and automated generation of land cover maps over large areas.

5.3.1 Description

The main idea is that the targeted selection of training samples can allow for the use of smaller training sets at least without degrading classification accuracy, which has the advantage of permitting classifier training cost reduction. Also, the directed selection of samples for training dataset construction has the potential of enhancing not only this efficiency but also the accuracy of classifications.

In this sense, a method is presented inspired in the line of research of [Foody and Mathur, 2004b],

where two main aspects of non-parametric classifiers are highlighted. Firstly, training samples are considered as not equally informative and having varying degrees of contribution to the classification analysis. This is particularly true with SVM classifiers, where it is the samples located close to the hyperplane locations, rather than those in the central core of the class, which are truly relevant. Secondly, the focus is placed on the exploitation of knowledge based on classifier behaviour, emphasizing the extraction of statistics that enable a classifier to discriminate among classes, rather than on the achievement of a full description of classes in the feature space.

As explained in the previous section, in [Foody and Mathur, 2004b], efforts are directed to obtain atypical training samples, with other samples with a typical behaviour beyond the border region being considered as presenting a lower contribution to the analysis. Conditioned by its definition variables, the application of stratification increases the probability of selection, when compared to other approaches, of samples corresponding to relatively small strata regions having particular characteristics impacting phenological response. Hypothetically, then, the selection of class subtypes, including those to be found in underrepresented contexts, will provide more useful samples than those typically selected by traditional approaches (such as random sampling and periodic sampling), which are more typical of the class being more numerous in the reference dataset.

The issue of imbalanced training sets with SVM classifiers is also addressed here, due to the acknowledged fact that it can degrade classification accuracy. An SVM learnt on an imbalanced training set is prone to generate a classifier with a high estimation bias towards the majority class [Tang et al., 2009]. Imbalanced reference datasets occur frequently due to the nature of commonly available data on land cover, often collected from different sources. As a result, this composite sets are commonly limited regarding particular classes or regions, with some categories presenting abundant samples and on the contrary others being just minimally represented.

Recursively, this imbalance applies to subclasses as well, with different spectral responses in time for a same vegetation class present in different environments. While standard occurrences in the most common environment in which a class occurs can be available in high numbers of samples, other regions corresponding to more specific vegetation growth conditions can be underrepresented. These situations can be counteracted with an appropriate spatial sampling design. In this way, knowledge on the behaviour of the classifier with imbalanced training sets is applied in order to improve sample selection.

Also, regarding the issue of training samples not being equally informative and having varying degrees of contribution to classification, emphasis is placed on the extraction of statistics that enable the classifier to discriminate among classes by targeting data training extraction towards areas susceptible of containing informative samples, based on knowledge on variables affecting the vegetation classes spectral response in time.

As environmental factors have an impact on location of classes in the feature space, we use additional information on variables influencing the spectral response of classes in order to select samples for training more intelligently. In this way it is possible to capture particular behaviours of classes (due to topographic position or substrate conditions differing from the most common occurrences, for example) which are far from the cluster center. Ancillary datasets can thus help

guide the training set selection by application of stratification, that is to say, the partitioning of the data into different sets for sampling, based on topographic or environmental data variables. In order to target discrimination between classes, this approach is used with the objective of minimizing variability within target types and maximizing variability between target types, as it is further explained in Sec.5.3.3.

A stratified spatial sampling method is proposed here. It is tested in three different variants based on the joint application of stratification and random sampling without repetition. Shortly, these variants are the following: stratified sampling per class, stratified sampling per class and strata, and stratified sampling and classification per strata. These stratification approaches are further described in Sec.5.3.3. Two further spatial sampling approaches have been implemented for comparison, using the following traditional techniques: random sampling, and periodic sampling. These approaches are also described in Sec.5.3.3.

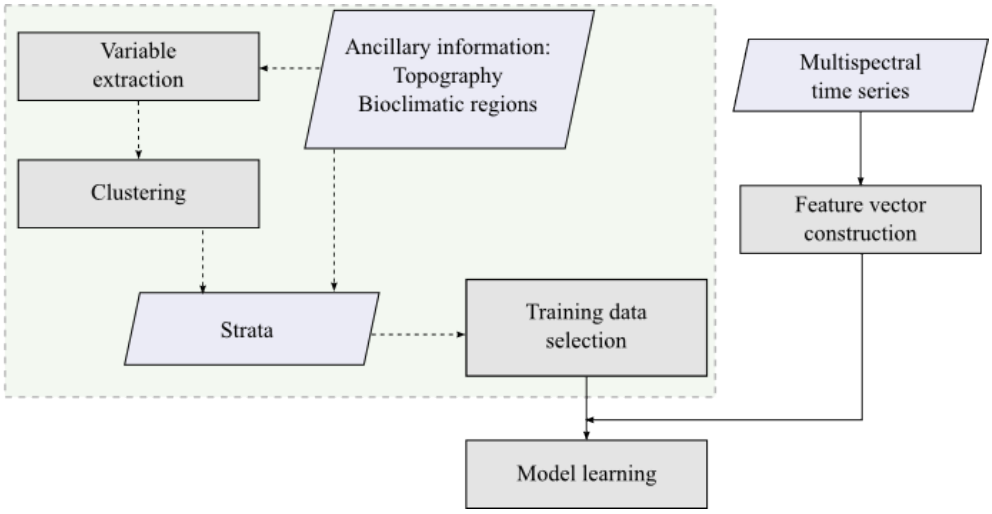


Figure 5.1: Proposed automatic spatial sampling approach for intelligent supervised classification model learning. See Fig.3.1 for integration in the proposed general framework

.

Fig. 5.1 illustrates broadly the proposed stratification method to automatically obtain an enhanced training set for supervised classification model learning. It integrates in the proposed general framework for automatic large area land cover map generation given in Fig.3.1. Given ancillary information, strata can be obtained in order to perform spatial sampling in two different ways. If a map of the homogeneous bioclimatic regions within the zone to process is available, defined regions or merged regions can be used to stratify the samples. Other information such as topography data can also be used after extraction of relevant variables, such as aspect or height, and the application of automatic non-supervised clustering. Stratification techniques are further described in Sec.5.3.3. Then, training data selection is performed, based on strata for the stratified approaches, or applying other spatial sampling methods. Spatial sampling approaches are further described in 5.3.2. This spatial sampling methodology is used in combination with the temporal sampling strategies described in the preceding chapter 4. Finally, the training set and the feature vectors derived from the available multispectral series are jointly applied for model learning for classification.

5.3.2 Spatial sampling approaches

Three different variants of the proposed stratified spatial sampling method have been implemented and tested, and are described here: stratified sampling per class, stratified sampling per class and strata, and stratified sampling and classification per strata. Two other common and simpler sampling approaches have been implemented as references for comparison: random sampling, and systematic sampling. These approaches are also described in this section.

5.3.2.1 Spatial systematic sampling

Given a specific population of N units, with the systematic sampling (SS) approach, n units are selected at regular intervals of size k from a randomly chosen startpoint. Every kth element is chosen, where $k = N/n$.

Advantages of SS include a facility to locate sample positions, a uniform coverage of the population. It is considered to give a spatially balanced sample, and it is often adopted for large scale national inventories [Köhl et al., 2006]. As a disadvantage, an irregular distribution of the population can give subsets of unequal size. Also, systematic sampling is vulnerable to periodicity: in this case, and depending on the used intervals, the sampling can be highly unrepresentative of the population.

5.3.2.2 Spatial random sampling

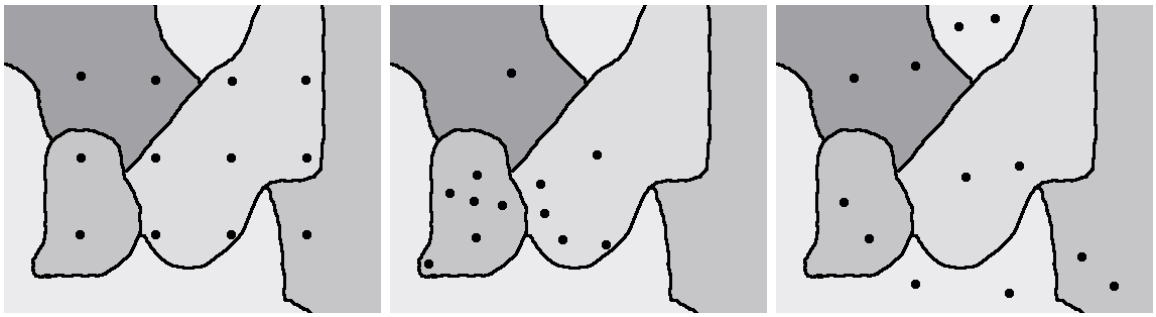
Given a specific population of N units, with the random sampling (RS) approach n units are selected at random, with all possible samples having the same selection probability at each draw. Two types of RS exist: RS with replacement and RS without replacement. With the first of the approaches elements are at all draws given the same probability of being in the sample. In this case, the probability of a single unit being in the sample is always n/N, and thus the same element can be drawn more than once. With the second of the approaches each drawn element is removed from the set before the next draw, so that the probability of being drawn at the $(k+1)th$ draw is $(n-k)/(N-k)$ for $k = 0, ..., n-1$. In this case, selecting n units out of N, $1/\binom{N}{n} = n!(N-n)!/N!$ distinct samples are possible. In traditional statistical sampling, allowing repetition of the same element is justified in some cases, as some variance estimators can only be derived with this approach [Brewer and Hanif, 1983]. As sampling is undertaken here for classification with an SVM, and thus additional copies of an element do not provide new information, only RS without repetition is applied here.

RS is widely applied, due to a variety of advantages such as its simplicity, which eases data interpretation and algorithmic implementation, and the fact that it requires no knowledge on the population of samples. A main disadvantage of RS, however, is that random draws can make up a sample not being representative of the total population.

5.3.2.3 Stratified spatial random sampling approaches

Fig.5.2 offers a simplified comparison of the extraction of samples with RS, SS, and the proposed stratified sampling approach in its most simple form, stratified sampling per class. Several classes are depicted with different grey levels, and possible sample spatial distributions with the three

different techniques are shown with dots.



(a) (b) (c)

Figure 5.2: Selected dataset examples with systematic sampling, random sampling, and stratified random sampling per class. The rectangular region represents a target land cover type, each grey level corresponds to a subtype, and dots mark selected samples. Some subclasses are lost with the first two approaches.

Possible drawbacks of SS and RS are illustrated here. With SS, classes can be completely excluded from the sampling due to the spatial distribution of the data: in Fig.5.2(a) two of the classes are not included in the selection. With RS, a random sample can result in a selection not representing the composition of the initial population: in Fig.5.2(b) three of the classes are represented in the sample. The stratified technique here proposed and represented in Fig.5.2(c) aims at overcoming these problems by using additional information about the population in order to obtain more representative samples.

Thus, stratified random sampling (StrRS) is based on the use of auxiliary information for the stratification or partitioning of the population into mutually disjoint strata or datasets. It targets the reduction of within-stratum variation by applying strata where units belonging to the same class present higher homogenization. StrRS is then applied, which helps ensure a more representative sample, particularly in cases where data are skewed, that is to say, when the percentages of occurrence of the classes in the dataset differ significantly. The representation of classes with limited number of elements is in this way enhanced, as these subgroups may be lost in a more generalised RS.

Stratified sampling can be most effective when variability within strata is minimized, variability between strata is maximised, and the stratification variables are correlated with the targeted classes. As advantages, it focuses on populations of interest and allows equal number sampling from strata widely varying in size. As disadvantages, the effectivity of the method is based on the use of relevant stratification variables, which can be difficult to determine, particularly when multiple criteria are used. Also, the cost and complexity of the sampling can be increased.

Its generally acknowledged that SVMs are more accurate than other classifiers when learnt on moderately imbalanced data [Tang et al., 2009] due to the fact that majority samples far from the decision boundary can generally be eliminated without loss of accuracy [Akbani et al., 2004]. However, cases of high class imbalance can generate a classifier with a strong estimation bias towards the majority class [Tang et al., 2009] and significantly affect classification performance. This fact has determined decisions on sample size allocation for each of the stratification designs described

in the following subsections.

Stratified sampling per class The first of the tested approaches applies stratification per class, that is to say, each of the targeted classes is considered as a strata without applying any additional information. Then, random sampling is applied on each of them for a balanced extraction of samples per class.

In this way, the sample size allocation is equal for all targeted classes (x pixels per class), independently of their respective availability in the reference ground truth. Tests undertaken in this chapter have applied, as explained above, x=500 pixels per class. Classes with fewer available pixels have not been taken into account.

Stratified sampling per class and strata The second of the approaches applies stratification per class and strata derived from additional datasets. That is to say, each of the targeted classes in each strata is considered as target subclass, in order to apply then random sampling on each of them for a sample extraction per subclass as balanced as possible.

In this way, the sample size allocation is equal for all targeted classes (x pixels per class) while ensuring extraction from all possible strata where the class is present. Thus, the size allocation for each subclass, as defined by the strata, is equal for each subclass in the class ($x/number\_of\_class\_subclasses$). When not enough pixels are available for a subclass, they are randomly taken from other subclasses. In this way, an equal representation of each subclass is targeted, while still providing stratification per class. Indeed, the approach constitutes a variant of the stratified sampling approach per class, only ensuring representation of all subclasses possibly having different feature characterization due to varying spectral response in time as a consequence of topographic or bioclimatic characteristics.

Stratified sampling and classification per strata Alternatively, strata can also be applied by applying both sampling and classification per strata. From this perspective, the third of the approaches applies stratified sampling per class and classification independently in each of the strata obtained from additional datasets, so that one classifier is applied on each area.

Sample size allocation is based on the percentage of the reference truth covered by each of the strata ($x * number\_of\_classes * area\_percentage\_of\_strata$), that is to say, the total extracted number of samples is exactly the same as in the previous approaches, but samples are proportionally distributed among the strata, reflecting the percentage of the total area covered by each stratum. In this way, for the number of pixels allocated in each strata, equal representation of the (varying) present classes is ensured. When not enough pixels are available for a selected class in a strata, the class is necessarily eliminated.

5.3.3 Techniques for the generation of strata

Strata have been obtained in two manners, described in the following subsections.

5.3.3.1 Clustering on topography-derived variables

The first method consists in the application of a clustering algorithm on topography-derived variables (see Chap.3) computed on elevation model information: altitude, slope, aspect, and their combination. Thus, the following steps have been undertaken:

1. Slope and aspect information have been extracted based on the available height information, and aspect has been normalised with respect to a North-South orientation, as explained in Sec.3.3.2. These three variables have been applied for strata extraction, with the hypothesis that they constitute factors that affect the development of vegetation and that may thus explain different phenological behaviours and their corresponding spectral characteristics in time. The impact of each of these variables on phenological responses has also been given in Sec.3.3.2.

2. Lloyd's algorithm (k-means) [MacQueen, 1967] has been applied on the available and extracted variables, both in isolation and in combination, in order to create stratification variables. Briefly, given a set of d-dimensional vectors $(\vec{x_1}, \vec{x_2}, ..., \vec{x_n})$, the k-means clustering algorithm aims at partitioning the n observations into $k \leq n$ sets $S = S_1, S_2, ..., S_k$ that minimise the within-cluster sum of squares, as in Eq.5.1, where $\vec{\mu_i}$ is the mean vector centroid of points in S_i:

$$\arg\min_{S} = \sum_{i=1}^{k} \sum_{\vec{x} \in S_i} ||\vec{x} - \vec{\mu_i}||^2 \tag{5.1}$$

The algorithm proceeds as follows. Given an initial set of k means $(\vec{m_1}^{(1)}, ..., \vec{m_k}^{(1)})$ which can be randomly obtained, the algorithm proceeds by alternating between the following steps until the assignment step no longer changes, which means that the algorithm has converged to a local optimum:

a. Assignment: assignation of observations to the cluster whose mean provides the least within-cluster sum of squares.

$$S_i^{(t)} = \vec{x_p} : ||\vec{x_p} - \vec{m_i}^{(t)}||^2 \leq ||\vec{x_p} - \vec{m_j}^{(t)}||^2 \forall j, 1 \leq j \leq k \tag{5.2}$$

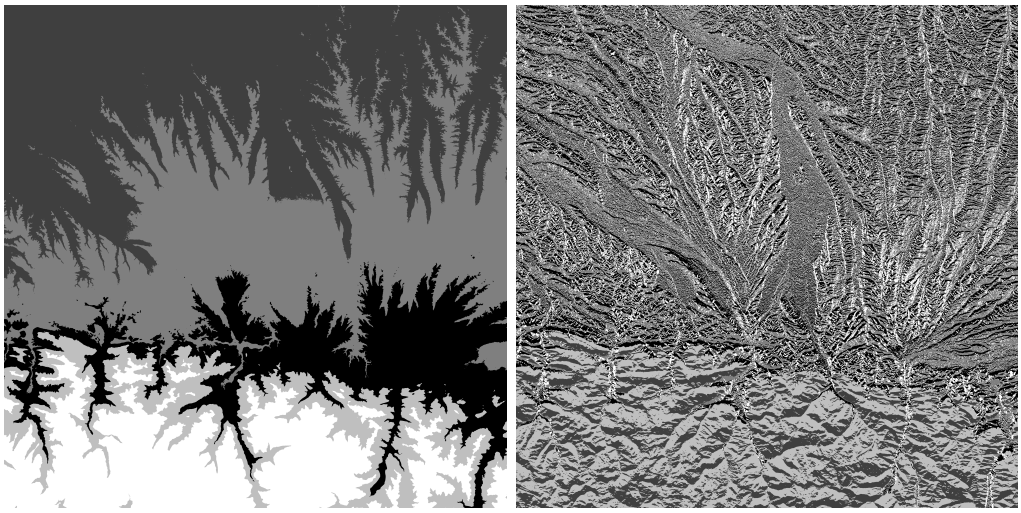
b. Update: new mean calculation to estimate the centroids of the new clusters.

$$\vec{m_i}^{(t+1)} = \frac{1}{|S_i^{(t)}|} \sum_{\vec{x_j} \in S_i^{(t)}} \vec{x_j} \tag{5.3}$$

The application of this clustering-based segmentation permits to identify similar geographical regions, presenting similar patterns for slope, height, orientation or their combined occurrence, so that a higher similarity is present between the elements within each cluster than between patterns belonging to different clusters. The algorithm is a data-driven segmentation technique, and cluster-oriented, where the partitioning of one-dimensional or multi-dimensional information into clusters yields as a result a number of non-overlapping

regions corresponding to a predefined number of target clusters. Initialisation is based on random seed extraction. In this way, machine learning is applied to obtain stratification regions based on natural surface or terrain parameters, which are important controls on a number of vegetation processes.

Fig.5.3 gives two illustrations of stratification maps (tile L930463) obtained by clustering into 5 clusters of height or height, aspect and slope data considered in combination.



(a) (b)

Figure 5.3: Five-cluster stratification maps (tile L930463) based on topographic information and obtained by clustering of (a) height, and (b) height, aspect and slope data. Each strata is shown in a different grey level.

Alg.1 gives a simplified summary of the procedures applied in order to obtain stratified maps for spatial sampling based on topographic information. It is worth noting that stratified maps are dependent on available information and can be of very different nature, not only regarding spatialisation resolution, but also in relation to their coverage, projection type and type of information, which may in turn limit their application. The procedures described here are automatic for the described available information, desired feature combinations being set as input variables.

Alg.1 can be further explained as follows. In a pixel-wise manner, a digital elevation model is used to compute slope information by gradient calculation. Then, the arctangent of the gradient is computed on all quadrants (atan2), and a north-south normalised orientation map of range [-1,1] is obtained by computing the cosine of the arctangent. Slope, aspect and height information are all normalised before application of the k-means clustering algorithm. Feature combinations depending on user input are then produced, such as for example height and aspect or height and aspect and slope, and stored into matrices containing vectors for each of the area pixels. Finally, clustering with the k-means algorithm is applied, as briefly outlined in Alg.1: following random initialisation and assignation of vectors to the initial seeds, recomputation of the centroids of the new clusters follows by obtaining the least within-cluster sum of squares. This procedure is repeated until recomputation generates a negligible change. The yield of this procedure is a strata map with

Algorithm 1: Stratification maps extraction based on topographic information

```
Input  : digital elevation data
Output: stratification map based on topographic information
READ digital elevation data;
while not end of image do
    COMPUTE slope:
        gradient with directional derivatives;
    COMPUTE aspect:
        atan2 of slope;
    COMPUTE north-south normalised orientation:
        rad to deg conversion;
        cos of arctan;
COMPUTE normalized matrices:
GENERATE feature matrices:
COMPUTE clustering into n number of strata:
    random initialisation of clusters;
    repeat
        assignation of vectors;
        recomputation of clusters;
    until no change occurred
GENERATE stratification map
```

the user-requested number of clusters. For testing, 3 and 5 clusters have been applied, in order to be allow the comparison of all the proposed approaches based on the same strata (stratified sampling and classification per strata requires a lower number of clusters in order to allow sufficient sampling); also, 5 clusters have been tested in order to verify the contribution of an increasing number of clusters.

5.3.3.2 Ecoclimatic areas

The second approach consists in the application of knowledge on climatic areas for the definition of strata. The applied climatic typology of the French continental territory was undertaken by [Joly et al., 2010] and corresponds to a geographical limit delineation based on measurements by Météo-France recording stations over a period of 30 years (1971-2000). Briefly, the spatial distribution of French climatic areas provided by the authors is based on the following steps:

1. Constitution of a climatic variable dataset. Two climatic parameters are studied, precipitation and temperature, and represented by a set of 8 and 6 variables, respectively. Other parameters such as cloud cover, wind and sunshine are discarded by the authors: they have a lower impact in terms of territorial structuring, and their measurement stations are limited in number. Thus, temperature variables include yearly temperature average, number of days with a temperature inferior to -5°C, number of days with a temperature superior to 30°C, annual amplitude, interannual variability in January and interannual variability in July. Precipitation variables, on the other hand, include annual cumulative precipitation, disparity with January values, disparity with July values, number of days of precipitation in January, number of days

of precipitation in July, interannual variability in January, interannual variability in July, and relation between Autumn (September and October) and July precipitations.

2. Spatial reconstruction of climatic variables. A spatial modelisation approach for climatic variables developed by the authors [Joly et al., 2009] is applied to obtain, for each geographical point for which the environmental characteristics are known, closer climatic variable values. These environmental values include 6 factors related to topography and 3 factors related to land cover: altitude, slope, orientation, topographic roughness and relative heightening on the one hand; and global theoretical radiance, vegetation index, distance to the forest and distance to the ocean/sea, on the other hand. With this procedure, and for each of the 14 climatic variables, a map is obtained better reflecting spatial variation at fine scales.

3. Typology extraction. The synthesis of the previously obtained maps is undertaken by coupling a factorial analysis of correspondences with a hierarchical ascendant classification [Joly et al., 2010]. Fig.5.4 shows the climatic type probability distributions obtained and the final typology provided by the authors (in the centre).

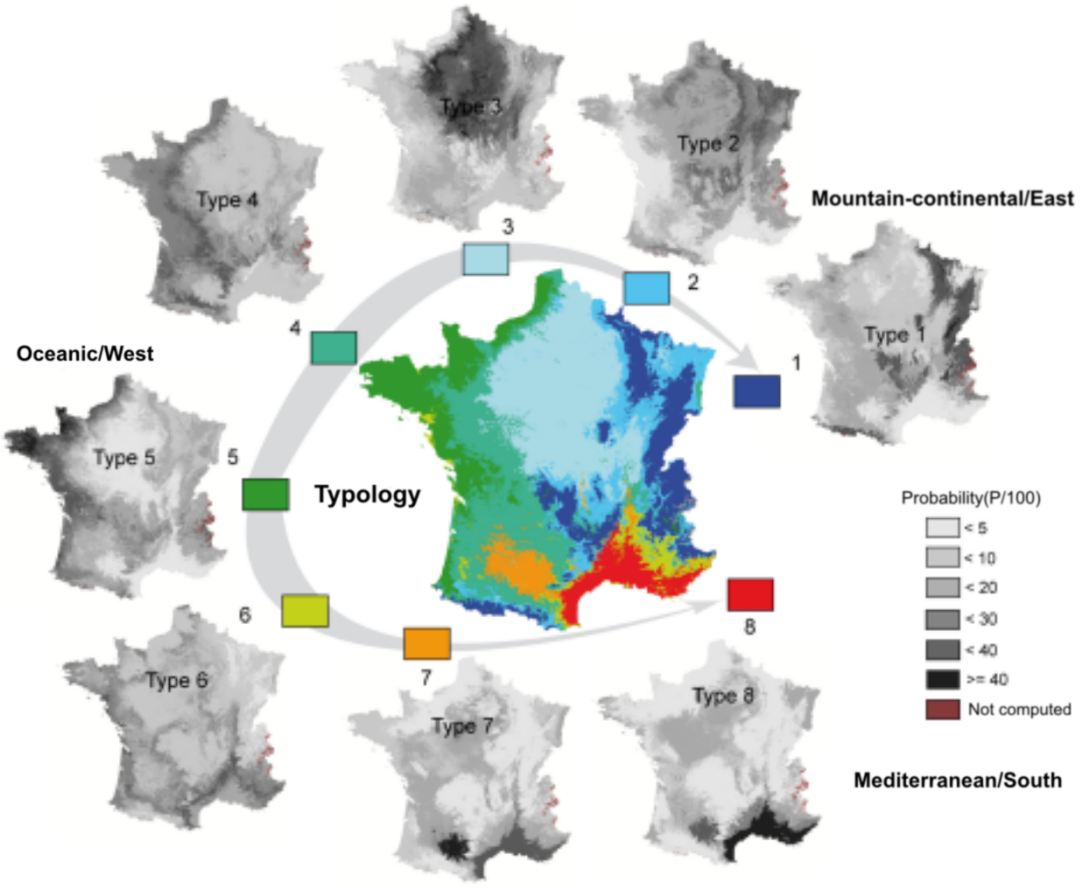


Figure 5.4: Probability distribution space of climatic types. From [Joly et al., 2010].

.

In the derived typological map, four factors appear as greatly structuring the space: the Mediterranean facade (type 8), the opposition between the oceanic type (5) and the semi-

continental type (2), and the impact of altitude (1). In summary, the typology highlights the following 8 types: mountain regions (type 1), semi-continental and mountain margins (type 2), degraded oceanic (type 3), altered oceanic (type 4), oceanic (type 5), altered Mediterranean (type 6), south-west basin (type 7), and Mediterranean (type 8). Please see [Joly et al., 2010] for a detailed description.

In this way, climatic types are applied as stratification regions, as they correspond to similar geographical regions in terms of temperature and precipitation characteristics. This application is based on the idea that these variables are main forces driving vegetation processes, greatly determining observable spatial variations in phenological responses. Fig.5.5 gives an illustration of the obtained stratification map for tile L930463 based on the ecoclimatic areas defined by [Joly et al., 2009].

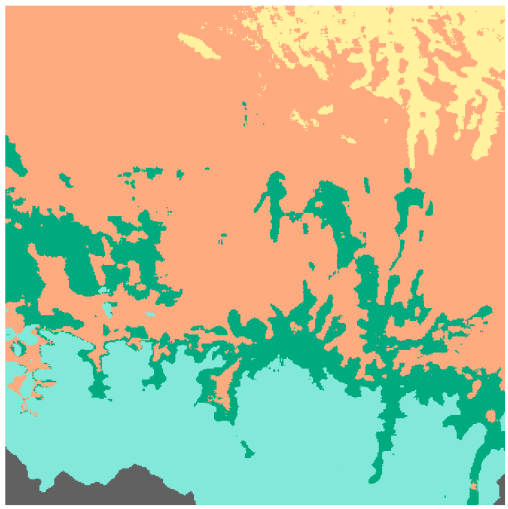


Figure 5.5: Strata based on ecoclimatic areas (tile L930463). Each colour corresponds to a bioclimatic region as defined by [Joly et al., 2009].

.

Alg.2 offers a simplified description of the procedures applied in order to obtain stratified maps for spatial sampling based on ecoclimatic information.

Algorithm 2: Stratification maps extraction based on ecoclimatic information

Input : ecoclimatic and target Landsat area data
Output: stratification map based on ecoclimatic regions
READ ecoclimatic region map, target Landsat area data;
while *not end of image* **do**
    COMPUTE reprojection of into target geometry:
        nearest neighbour interpolation;
GET fusion into n number of strata;
GENERATE stratification map;

The applied procedure can be described as follows. Firstly, reprojection of the available information on ecoclimatic regions into the geometry of the Landsat tiles using available metadata is applied, resampling from the original 250m spatial resolution to the target 30m resolution. Interpolation with the nearest neighbour method is applied, in order to obtain a piecewise-constant interpolant with no modification of the original input values, which reflect categorical information. A binary mask is also derived at this point for further processing, indicating the valid area where stratification information is available. Finally, the desired n number of strata are derived. For testing, 4 strata have been applied, as the distribution of the original bioclimatic areas conditions their appearance in the test region. Also, two further areas were present in the region, but to a very small extent not permitting sufficient sampling, so they were merged into their neighbouring classes.

5.4 Experiments and results

An analysis to identify possible stratification variables was conducted previous to the undertaking of the tests here presented, with the hypothesis that adequate strata minimise within-strata variability, and maximise inter-strata variability. This preliminary study on variance and correlation analysis was based on a limited region for which topographic and pedological data was available. Pedological information was used for the Midi-Pyrénées region with two different levels of nomenclature as described in [CRAMP, 1995]. Refered to as P1 and P2 below, these levels are described based on aspects such as pedogenesis, morphology, composition and drainage characteristics. However, although applied for the preliminary study, pedological data was finally discarded due to lack of coverage over the study region.

Following the preliminary analysis described in the next subsection 5.4.1, the presented spatial sampling methods have been applied for sample selection and SVM classifiers have been trained on the resulting sets. Classification has been undertaken with the learnt classifiers and results have been analysed in order to estimate the impact of each sampling methodology on classification accuracy. Results are discussed in subsection 5.4.2

5.4.1 Variance and correlation analysis

In order to analyse the characteristics determined by different stratification variables, and whether stratification might be advantageous, the population variance of each strata and corresponding substrata of size N has been calculated according to Eq.5.4:

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2, \text{ with population mean being } \mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{5.4}$$

Tab.5.1 shows, for each stratification variable or combination of variables, and the applied clustering centroid number, minimum, maximum and average variances, and thus the average increment of variance of the stratification strategy in comparison to sampling without stratification. All calculations have been performed on NDVI values corresponding to Landsat series data.

Pearson's product-moment non-centered correlation coefficient has been applied to measure correlation between strata corresponding to each of the stratification strategies listed above. A

measure of the correlation between two variables, the sample coefficient is defined as in Eq.5.5, and has been performed by computation of the coefficient for pairs of pixels belonging each one to a different ground truth region or stratification strata:

$$r = \frac{\sum_{i=1}^{n}(X_i)(Y_i)}{\sqrt{\sum_{i=1}^{n}(X_i)^2}\sqrt{\sum_{i=1}^{n}(Y_i)2}} \tag{5.5}$$

This preliminary study has been obtained on a limited region of 110x30km, area covered in the targeted region of study by the available pedological information, with a nomenclature of 15 classes. Tabs. 5.1 and 5.2 summarize, for this region, the obtained variance and correlation statistics, respectively. GT refers to ground truth, and constitutes the reference against which computed statistics are compared.

	A5	H5	P1	P2	S5	HA5	HAS5
Av.	0,030	**0,027**	**0,025**	**0,025**	0,032	0,030	0,030
Std.	0,012	0,013	0,016	0,014	0,016	0,015	0,014
Min.	0,009	0,002	0,004	0,004	0,000	0,010	0,009
Max.	0,059	0,059	0,085	0,085	0,079	0,068	0,063
HS5	GT	A15	H15	S15	HA15	HAS15	HS15
0,030	**0,030**	0,032	**0,025**	0,035	**0,027**	**0,027**	**0,026**
0,014	**0,012**	0,014	0,012	0,018	0,014	0,013	0,013
0,007	**0,010**	0,007	0,001	0,002	0,000	0,000	0,001
0,080	**0,055**	0,104	0,063	0,101	0,068	0,071	0,063

Table 5.1: Variance statistics (columnwise): average, standard deviation, minimum and maximum. Stratification variables: A-aspect, H-height, P-pedology (1-2), S-slope, GT-ground truth. Number of clusters for combined variables: 5 or 15, as indicated

	A5	H5	P1	P2	S5	HA5	HAS5
Av.	0,904	0,900	**0,914**	**0,913**	**0,907**	0,905	0,903
Std.	0,058	0,061	0,056	0,056	0,055	0,056	0,061
Min.	0,625	0,647	0,521	0,521	0,667	0,641	0,591
Max.	0,971	0,989	0,987	0,987	0,983	0,978	0,980
HS5	GT	A15	H15	S15	HA15	HAS15	HS15
0,895	**0,905**	0,903	0,899	0,905	0,900	0,898	0,896
0,063	**0,055**	0,060	0,065	0,058	0,066	0,071	0,073
0,619	**0,683**	0,598	0,577	0,589	0,480	0,379	0,440
0,984	**0,984**	0,979	0,985	0,983	0,988	0,987	0,987

Table 5.2: Summary of correlation statistics per stratification strategy. Nomenclature as in Table 5.1

Analysis of numerical results seems to indicate that a slight reduction of variance in absolute values can be achieved by stratification, with an intra-strata variance decrease when pedologic or slope information are used. Increasing clustering centroid number appears to have a positive effect on height and height-based stratification variables. Inter-strata correlation, on the other hand, seems to diminish with most tested strata. This is coherent with the observation of distinct vegetation evolution patterns in different topographic conditions. However, it must be pointed

out, standard deviations constitute around 50% of the computed averages, while distances between variables are much smaller, so conclusions must be drawn with care.

However, the difficulty of analysis of this particular type of data is illustrated by Tab.5.3. Correlation values have been calculated over validation data using a sampling ratio of 1/3, inter-strata correlation being expected to be higher than correlation with other strata. This is not always the case: with the same dataset, using a nomenclature with 13 classes, "deciduous forest" presents its highest correlation against itself, "evergreen forest" has a higher correlation with "woody permanent crops", and "mixed forest" reaches its maximums against the former two. Many factors may have an impact here, from the sampling for correlation calculation itself, to inaccuracies in the ground truth, and spectral mixing.

	SC	WC	MC	WPC	HPC	M	F	Mo	DF	EF	MF	R	B
DF	0.903	0.886	0.899	0.961	0.957	0.956	0.956	0.847	**0.965**	0.951	0.946	0.953	0.947
EF	0.908	0.852	0.885	**0.945**	0.940	0.933	0.935	0.855	**0.951**	**0.942**	0.938	0.935	0.928
MF	0.905	0.843	0.880	0.933	0.933	0.923	0.925	0.855	**0.946**	**0.938**	**0.934**	0.924	0.916

Table 5.3: Correlations with non-centred Pearson's index. Nomenclature with 13 classes: SC-summer crops, WC-winter crops, MC-mixed crops, WPC-woody permanent crops, HPC-herbaceous permanent crops, M-meadows, F-fallows, Mo-moorlands, DF-deciduous forests, EF-evergreen forests, MF-mixed forests, R-roads, B-buildings.

A great number of classification tests has been undertaken in order to assess the impact of stratification on classification accuracy results. Other datasets have been applied for the purpose, covering larger areas. As no pedological information is available for these greater areas, it has not been further applied for stratification, but remains a open research aspect to be investigated.

5.4.2 Impact on classification

Assessment of the different approaches to stratification is undertaken based on their impact on classification results, following the experimental setting described in Sec.4.4.1. In this way, accuracy is assessed in terms of average global accuracy, average percentages of well-classified pixels per class, average standard deviation of average global accuracy, average kappa statistic, and a final average class percentage accuracy (average of percentages for all classes) for 20 learning sets. A total of 560 classification tests are presented in this section.

This chapter builds on the findings presented in the previous chapter. Firstly, missing information processing is based on the application of linear interpolation, as supported by results in Tab.4.6 and Tab.4.12. Secondly, the sampling number is set to 500 pixels per class (8500 in total for the studied configuration with 17 classes), as it represents a good compromise between accuracy improvement and sampling ratio (Tab.4.10). Thirdly, the enhancement of separability between classes by addition of non-spectral information shown in Tab.4.11 is also applied here. Finally, a temporal sampling of 12 virtual points is used, also giving a good compromise between accuracy increase and processing costs in terms of memory and, as it will be shown later, computation times.

Three stratification approaches have been in this way implemented and tested: stratified sampling per class; stratified sampling per class and strata; and stratified sampling and classification

per strata, as described in Sec.5.3.2. These implementations have been compared against two commonly applied approaches to spatial sampling: random sampling and periodic sampling.

For the rest of the chapter, tables reproduce average results based, each time, on 20 independent sample sets and classifications. For the reference dataset here applied (L930463), the produced number of tests, and the given configuration (where training pixels are not considered for evaluation), each results entry line with classification averages in this section is thus summarizing 20 full area classifications and evaluation, for each of them, against 747523 pixels of ground truth ($756023 - 17 * 500$).

Tab.5.4 summarizes the results for 20 classifications based on random sampling (RS) training sample extraction and 20 classifications based on periodic sampling (PS). As explained in Sec.5.3.2, the former is based on random draws without replacement for the available reference pixels, and the latter on the extraction of pixels in a systematic spatial distribution. As also discussed, these sampling modes do not assure that, for two different sets, the training and classification classes coincide, as small classes can be left out of the sampling. For this reason, and in order to be able to provide comparable average statistics, results in Tab.5.4 correspond to classifications where all classes where represented in the extracted datasets, although this condition cannot be guaranteed with neither of the two approaches.

RS	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	77.152	94.626	10.204	49.677	63.804	39.001	53.956	8.303	79.429	68.257	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	22.919	7.670	5.497	92.018	79.639	0.038	16.684	0.823	0.001	0.761	45.227
PS	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	77.434	94.661	7.777	53.287	61.855	39.669	53.330	12.798	79.932	78.270	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	18.884	6.487	8.131	92.128	79.736	0.0	11.725	0.824	0.001	0.762	45.653

Table 5.4: Percentages of well-classified pixels, per class, for the random (RS) and periodic (PS) sampling approaches, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of a region of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Scores for the RS and PS approaches, summarized in Tab.5.4, show high average overall accuracies (respectively, 0.823 and 0.824) and kappas (0.761 and 0.762). However, judgement based on these metrics is not adequate when reference data is highly imbalanced, as it is the case here. For example, in this particular configuration, out of the total 756023 reference data pixels, the "mixed forest" class counts 1589 sample pixels, while the "deciduous broadleaf forest" class accumulates 279280. The bias effect of this imbalance is made clear in the provided examples: high overall accuracies and kappas are obtained, while classification of some classes is extremely erroneous. For example, both approaches present 8 classes with accuracies under 40%, and some of them reaching as low as (close to) 0% (both RS and PS, "mixed forest"), and 5.497% (RS, "other summer crops") and 6.487% (PS, "other winter crops"). For this reason, an average class percentage is giving clearer indication of the quality of the obtained classification, all classes included, without considering class size. Average class percentages are thus as low as 45.227% and 45.653%, respectively, for the RS and PS example tests.

5.4.2.1 Stratified sampling per class

In this way, Tab.5.5 provides results for the first of the stratified sampling approaches, that is to say, stratified sampling per class (SPC). For each of the targeted land cover classes, sample size allocation is equal independently of their availability in the reference ground truth. This approach has also been used in the preceding chapter.

SPC	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	50.738	82.995	61.784	76.625	83.636	47.784	59.566	81.523	86.288	98.681	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	81.190	45.943	40.580	61.805	68.119	69.307	92.213	0.682	0.008	0.609	69.928

Table 5.5: Percentages of well-classified pixels, per class, for a stratified sampling approach per class, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of a region of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

As shown in Tab.5.5, a great improvement in average class percentage accuracy is obtained with SPC when compared to the previous examples, with an average of 69.928% in comparison with the former 45.227% and 45.653% for RS and PS. With SPC only 3 classes fall below 50%, with a lowest accuracy of 40.580%, and eight classes present over 80% accuracy. Thus, improvement is significant for most of the classes, as all classes are equally represented in the sampling. Indeed, the two most abundant reference set classes, "corn" and "deciduous broadleaf forest" experience a decrease. This worsening is presumably due to the oversampling that the RS and PS approaches permit for highly numerous classes, and also due to the sampling bias mentioned before. Contrarily, classes underrepresented by the previous approaches are now well represented and present a considerably improved classification accuracy.

5.4.2.2 Stratified sampling per class and strata

Tabs.5.6, 5.7 and 5.8 provide results for the second of the stratified sampling approaches, that is to say, stratified sampling per class and strata. This approach is based on the application of strata information derived from additional datasets in order to extract equal final numbers of class samples while ensuring sampling in all strata where the class is present.

Tab.5.6 gives results for stratification per class and strata based on height with 3 (H3) and 5 (H5) clusters, obtained as described in Sec.5.3.3. A considerable improvement with respect to RS and PS is also noticeable here, with scores for average class accuracy reaching 70.318% and 70.362% for 3 and 5 clusters, respectively. A small improvement is also observable in relation to the previous stratification approach, which attained a lower 69.928%.

Tab.5.7 gives results for stratification per class and strata based on aspect and slope with three clusters, and on bioclimatic areas. The latter stratification is derived from the spatial distribution proposed by [Joly et al., 2010] based on presence in the test area and considering

H3	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	50.137	83.565	65.135	76.191	83.258	46.920	60.809	82.445	87.185	98.756	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	82.068	45.200	40.510	62.851	68.637	68.577	93.167	0.689	0.006	0.616	70.318
H5	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	50.145	83.447	63.266	77.106	83.089	48.088	61.188	82.233	86.770	98.905	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	81.892	45.616	40.366	61.883	69.149	69.467	93.537	0.687	0.006	0.614	70.362

Table 5.6: Percentages of well-classified pixels, per class, for stratified sampling approaches based on height with 3 (H3) and 5 (H5) clusters, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of a region of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

the most structuring factors as given by the authors, thus obtaining 4 different strata. Some improvement with respect to RS and PS is also noticeable here, with scores for average class accuracy at 70.291%, 70.361% and 70.260%, respectively.

A3	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	49.619	82.947	64.239	76.656	83.185	47.465	61.383	81.993	86.474	98.557	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	81.993	46.382	40.652	62.320	68.947	68.811	93.315	0.686	0.007	0.613	70.291
S3	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	50.629	83.351	63.338	76.650	83.122	47.794	60.731	82.300	86.873	98.731	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	82.301	45.663	40.751	62.445	68.802	69.467	93.185	0.688	0.007	0.615	70.361
B4	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	49.889	83.998	64.090	77.110	83.107	46.993	61.789	81.966	86.993	98.308	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	81.209	45.765	40.336	62.497	69.040	68.366	92.963	0.690	0.005	0.617	70.260

Table 5.7: Percentages of well-classified pixels, per class, for stratified sampling approaches based on aspect (A3), slope (S3), and bioclimatic areas (B4), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of a region of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Superior results are obtained when a combination of height, aspect and slope is used for stratification when applying 5 clusters. Tab.5.8 summarizes results for stratification per class and strata based on this combination of height, aspect and slope for 3 (HAS3) and 5 clusters (HAS5).

As it can be observed in Tab.5.8, average classification accuracy with the HAS3 approach reaches 70.342%, while the HAS5 approach peaks at 70.366% with the best obtained results. Most classes, 13 out of 17, present accuracies of over 60%, with 7 of them having over 80% results. In relation to the random sampling and the periodic approaches, only 4 classes decrease scores, while 13 classes improve significantly their accuracies with an average of +38.179% and reaching up to +78.009% Tab. 5.9 shows as an example the relative average improvements obtained, in relation to average results of the PS and RS approaches, for 20 tests using the HAS5 approach.

Thus, improvement is significant for all of the stratified tests per class and strata in relation to

HAS3	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	48.707	83.291	64.072	76.991	83.317	47.443	61.416	81.305	87.238	98.855	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	82.011	46.311	40.034	61.784	69.484	70.073	93.481	0.686	0.008	0.613	70.342
HAS5	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	49.272	83.223	64.180	76.859	83.390	48.021	61.287	82.018	86.804	98.657	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	81.872	45.841	41.147	61.396	69.494	69.449	93.315	0.685	0.007	0.612	70.366

Table 5.8: Percentages of well-classified pixels, per class, for stratified sampling approaches based on combined height+aspect+slope with 3 (HAS3) and 5 clusters (HAS5), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of a region of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Improvement	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP
PS/RS-HAS5	-26.555	-11.649	52.794	25.143	20.807	8.449	5.923	70.973	6.608	25.418
	HMO	OWC	OSC	DBL	PNL	MFO	RMO			
	60.289	38.865	33.766	-30.268	-11.568	69.288	78.009			

Table 5.9: Per-class average percentages of improvement for 20 independent classifications using stratified sampling per class based on combined height+aspect+slope with 5 clusters (HAS5), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463), and compared to average results of RS and PS given in Tab.5.4. Nomenclature with 17 classes. See Tab.2.10.

the reference periodic and random sampling approaches. Indeed, grounds can be found, as it was also in the previous approach (stratification per class), in the equal representation of classes in the sampling. Those being underrepresented by the traditional approaches are now well represented and present a considerably improved classification accuracy. Furthermore, tests in this section show also a moderate improvement in relation to the stratified approach per class, as recursively, the improved representation of subclasses seems to have a positive impact on results.

5.4.2.3 Stratified sampling and classification per strata

Tabs.5.10, 5.11 and 5.12 provide results for the third of the stratified sampling approaches, that is to say, stratified sampling and classification per strata. This approach is based on the application of strata information derived from additional datasets in order to extract separate regions on which to perform independent stratified sampling per class and classification, as well as the production of a final classification based on combined results. Strata with 3 clusters (4 for bioclimatic regions) have been applied, in order to allow for a sufficient sampling set in each strata.

Tab.5.10 gives a summary of the undertaken classifications the stratified sampling and classification per strata with bioclimatic areas. As before, the stratification is derived from the spatial distribution proposed by [Joly et al., 2010] and defining 4 different strata in the test region (B4). Stratified sampling per class is applied for each of the strata. Average results for 20 independent classifications are given for each of them (labelled B4A, B4B, B4C and B4D), and it is to be observed that not all classes are present in all of the strata. Final average evaluation is given in B4M (merged results), computed based on the addition of all confusion matrices used for validation into a confusion matrix containing all targeted classes. The final average class accuracy (69.910%

B4A	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	-	-	-	-	71.163	-	-	-	86.918	-	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	75.005	-	-	75.755	72.129	95.338	88.668	0.735	0.008	0.586	80.711
B4B	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	64.680	79.074	73.200	92.439	83.811	46.447	55.970	89.639	86.693	97.883	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	88.250	57.623	50.892	70.431	71.399	-	98.750	0.700	0.016	0.554	75.449
B4C	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	51.986	81.754	64.988	74.789	81.147	51.533	48.078	81.633	88.571	99.375	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	94.404	42.299	37.034	58.044	60.825	65.136	-	0.684	0.007	0.592	67.600
B4D	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	37.783	82.528	79.288	64.471	76.180	56.436	46.593	82.231	73.565	96.666	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	-	43.826	61.715	67.168	64.732	-	-	0.621	0.017	0.574	66.656
B4M	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	48.937	81.44	69.532	73.169	79.861	57.162	48.939	82.594	87.288	97.939	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	78.149	44.209	39.609	65.89	70.382	74.598	88.767	0.698	0.011	0.623	69.910

Table 5.10: Percentages of well-classified pixels, per class, for a stratified sampling approach per strata based on bioclimatic regions, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications. Nomenclature with 17 classes. See Tab.2.10.

) is superior to those provided by the random and periodic sampling approaches, although not reaching results attained by the stratified sampling per class and strata approach.

H3A	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	63.336	77.276	66.487	88.461	90.000	53.827	56.166	-	86.430	-	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	86.713	58.031	45.448	66.919	63.072	85.455	-	0.661	0.021	0.502	70.544
H3B	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	-	-	-	-	-	77.197	-	-	86.597	-	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	73.346	-	-	76.821	71.394	95.652	88.000	0.733	0.019	0.550	81.287
H3C	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	48.576	80.792	65.495	73.547	82.342	49.307	46.145	79.277	89.231	98.725	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	91.323	44.943	41.617	56.629	62.429	74.749	-	0.672	0.008	0.582	67.820
H3M	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	49.664	80.563	65.616	74.72	82.343	58.692	47.845	79.278	88.856	98.725	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	79.193	46.126	42.068	62.672	69.304	81.36	88.0	0.683	0.015	0.608	70.296

Table 5.11: Percentages of well-classified pixels, per class, for a stratified sampling approach per strata based on height information, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications. Nomenclature with 17 classes. See Tab.2.10.

Tab.5.11 gives a summary of the undertaken classifications with the stratified sampling and

classification per strata with height-derived clustering with 3 clusters, obtained as described in Sec.5.3.3. As before, average results for 20 independent classifications are given for each of the strata (H3A, H3B and H3C) and evaluation of merged results is given in H3M. Again, the final average class accuracy (70.296%) does not reach results attained by the stratified sampling per class and strata approach, but it is superior to those provided by the random and periodic sampling approaches.

HAS3A	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	49.652	79.523	63.294	78.346	81.806	48.654	52.046	80.259	85.486	98.471	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	84.844	42.537	41.732	57.297	60.895	68.305	-	0.657	0.011	0.569	63.126
HAS3B	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	-	-	-	-	-	75.437	-	-	86.162	-	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	76.135	-	-	76.371	71.374	95.087	84.531	0.734	0.016	0.572	80.728
HAS3C	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	43.724	81.054	63.467	71.573	80.937	49.605	53.807	79.420	90.448	98.488	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	90.664	46.958	34.139	57.194	61.343	67.243	-	0.651	0.011	0.561	66.879
HAS3M	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	47.188	80.128	63.359	75.337	81.439	56.505	52.76	79.841	86.999	98.48	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	81.51	44.515	38.874	60.73	68.931	75.216	84.531	0.674	0.012	0.598	69.197

Table 5.12: Percentages of well-classified pixels, per class, for a stratified sampling approach per strata based on height+aspect+slope information, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications. Nomenclature with 17 classes. See Tab.2.10.

Finally, Tab.5.12 gives averages for the undertaken classifications using stratified sampling and classification per strata with height+aspect+slope data and clustering with 3 clusters. For each of the strata (HAS3A, HAS3B and HAS3C) and for the final classification (H3M), average classification results are given. Final average class accuracy is slightly lower in this case (69.197%), again not reaching results attained by the stratified sampling per class and strata approach, but it is superior to those provided by the random and periodic sampling approaches.

Two main inferences based on the stratification and classification approach tests are the following. Firstly, some classes seem not to be present in some of the strata, which allows classifications with high accuracies in those strata (for example B4A in Tab.5.10). This indicates that, with a well chosen stratification, some classes can be better classified with the approach. However, the problem of the selection of adequate climatic variables for clustering or bioclimatic areas which are adequate for targeted classes still remains. It must be noted, also, that such cases are also affected by the fact that a reduced number of classes is involved, which in turn has an effect on classification scores.

In conclusion, it can be asserted that a great improvement is obtained with the application of stratification for spatial sampling in comparison to results attainable with the traditional RS and PS approaches. In particular, RS and PS examples have been analysed in terms of average

accuracy per class, yielding, respectively, 45.227% and 45.653%. When no exogenous information is available, stratification per class does provide significantly higher scores at 69.928%. When strata data is available, sampling and classification per strata also provides improved results over the RS and PS examples. Among the tested H3M, HAS3M and B4M approaches, height stratification with three clusters seems to provide the best results (70.296%), just slightly superior to stratification per class. Better results are obtained with stratification per class and strata, and in particular with stratification based on height or a combination of height+aspect+slope. The highest average class accuracies, although only slightly better than those of the stratification per class approach, are achieved with the H5 and HAS5 approaches, reaching 70.366% and 70.362%, respectively, much superior to the RS and PS examples.

The approach obtaining best classification results (stratification based on class and strata, based on 5 clusters and height, aspect and slope information HAS5) has been applied on an expanded study region comprising tiles L930363, L930463, L930563 and L93663. The region does not include tiles L930362, L930462, L930562 and L93662 in order to allow comparison among approaches making use of information only available for the France territory, as it is the case in the following Chap.6. In order to permit comparison, also a stratified sampling approach per class has been applied on the same area. Results for both sets of tests are given in Tab.5.13.

SPC	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	
	50.159	81.652	61.179	77.199	78.329	86.436	37.09	50.416	
	VIN	WAT	IMP	HMO	SMO	OWC	OSC	FNO	DBL
	69.223	81.607	92.224	71.84	62.124	52.221	47.569	57.252	45.47
	PBL	DNL	PNL	MFO	RMO	av	stdev	κ	av.class
	60.731	69.793	59.59	52.091	52.091	0.557	0.006	0.497	63.467
HAS5	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	
	50.76	83.338	62.515	76.301	77.622	86.977	39.966	51.946	
	VIN	WAT	IMP	HMO	SMO	OWC	OSC	FNO	DBL
	69.654	81.774	92.282	74.323	61.791	52.317	46.905	57.81	47.138
	PBL	DNL	PNL	MFO	RMO	av	stdev	κ	av.class
	65.711	71.256	61.368	51.565	49.411	0.576	0.005	0.516	64.215

Table 5.13: Percentages of well-classified pixels, per class, for a stratified sampling approach per class (SPC), and a stratified sampling approach based on combined height+aspect+slope with 5 clusters (HAS5), with temporal resampling with virtual timepoints and a non-refined reference dataset for the study area (tiles L930363, L930463, L930563, L930663). Results for each approach correspond to average results for 20 independent classifications of 13667x3665 pixels. Nomenclature with 22 classes. See Tab.2.10.

Tab.5.13 shows comparable results to those obtained on the previous study region, with an increase in classification performance (average class accuracy and kappa) for the stratification per class and strata approach (64.215% in comparison to the 63.467% of the stratification per class approach). This demonstrates further the advantages on ensuring subclass representation, as it makes it clearer on a larger area how classification can be enhanced. Average accuracies need to be interpreted here in relation to previous tests having into account that although the total sampling ratio has been kept the same (under 0.07%) by applying 500 per class (17 classes) on a 3665x3665 area to 1500 (22 classes) on a 13667x3665 area, the application of a superior number of classes has a necessary impact on results. Consequently, percentages are lower, but the relative improvement is anyway made clear. Most significantly, Tab.5.13 demonstrates how sampling on a very small

percentage of pixels on a very large study area generates acceptable classification accuracies.

A final word needs to be added on learning and classification times. Indeed, the method presented targets not only the problem of data variability over large areas, but also the problem of considerable data volume management. The generation of lower volume training sets with a higher representativity allows for fast machine learning model learning and classification. For SVMs, learning times can greatly vary depending on data size (number of learning samples), dimensionality (number of features) and chosen kernel (an RBF kernel is slower than linear kernel). The advantages of the proposed method, allowing the use of a small percentage of representative samples, are further made clear when classification times are considered. Tab.5.6 shows a summary of average learning times for SVMs based on a linear and an RBF kernel corresponding to 5 tests per approach. Standard deviation is also indicated, together with the average extracted number of support vectors for each approach. Different number of training samples have been tested, and their average extraction times are also given.

	Extraction	Linear learning			RBF learning		
samples	av time	av time	stdev	svs	av time	stdev	svs
200	1.38	15.668	1.5469	1475.2	45.43	4.134	2486.8
500	1.398	112.074	7.9914	3361.6	279.765	18.5925	5625.6
1000	1.422	469.8	9.6377	6333	1475.604	20.8122	10086.8
1500	1.482	1087.104	33.6114	9338	3262.254	229.9138	14219.4

Table 5.14: Processing times for sample set extraction and SVM learning with linear and RBF kernels: av.time-average time (in s), stdev-standard deviation, svs-number of computed support vectors.

Computed processing times in Tab.5.14 reflect both the increase of the learning stage computation times with the number of samples (given in samples per class), and the higher cost of the RBF kernel. As classification run-time complexity scales linearly with the number of support vectors [Geebelen et al., 2012], the increase of computation times for classification, and not only for learning, is made obvious. With the large volumes of data to be yield by the new satellite missions, the interest of the approach is made clear. More visually, processing times for learning are given in Fig.5.6. The exponential tendency of the curves (also observable for classification times, although not depicted here) further clarifies the advantages of an appropriate sampling design for classification learning set extraction.

5.5 Conclusions

The problem of data variability and large volumes of data for land cover mapping with satellite data time series over large-areas has been analysed, and a brief review of the literature has been provided.

A training data extraction method based on a spatial sampling approach has been presented. It targets the intelligent selection of pertinent training samples with two objectives: the reduction of data volume on which to base classifier learning, and the enhancement of the performance of the classifier itself. The method tries to capture the variability inherent to large areas by using additional information to guide the selection of samples. It overcomes the difficulties posed by the common problem of large data volumes as well as of spectral data variability that very often hinder

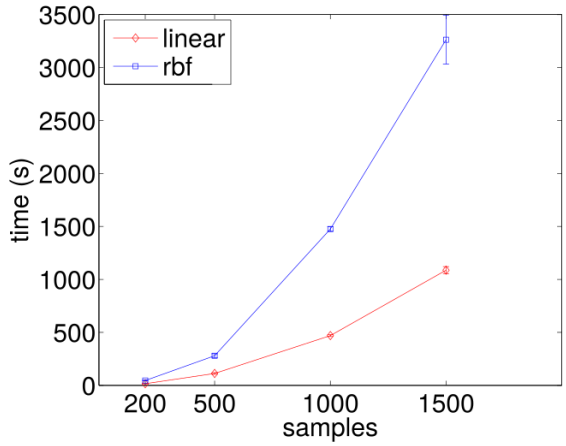


Figure 5.6: SVM learning times with linear and RBF kernels.

map production over large areas.

The methodology is based on the application of stratification for sampling, which allows for the targeted selection of samples based on for example topographical or ecoclimatic information. A spatial sampling technique, it is used in combination with the temporal techniques presented in Chap.4. In this way, a methodology for the automated generation of large area land cover maps is presented, operating in space for dataset selection, and in time for missing data processing. Three different stratification approaches have been implemented and compared against two other traditional and widely applied sampling techniques.

The approaches have been assessed in terms of impact on classification results over two datasets, one comprising an area of $12100km^2$, and the other comprising an extended study region of $45100km^2$. The potentiality of stratification has in this way been demonstrated for training set selection for classification and, in particular, its applicability for the operational generation of land cover maps over large areas.

Chapter 6

Non-linear temperature-based temporal sampling for large area processing

Contents

This chapter introduces the topic of linearity of the temporal axis when dealing with vegetation parameter extraction. It builds on the previously presented method for the generation of large-area land cover maps with operational times, based on the application of interpolation techniques for the processing of missing information, the use of temporal sampling with virtual points, and the application of spatial sampling techniques for training dataset extraction, for the management of temporal irregularities, data variability, and large volumes of data. Non-linear temperature-based temporal sampling is applied on land cover classes, with the hypothesis that phenological evolution and temporal and spectral response of vegetation land cover is dependent on temperature rather than time. An overview of the literature on temperature-related phenological studies and their use in remote sensing literature is given. A novel method for the exploitation of temporal spectral response in the temperature space for classification is given, and two different approaches are developed and tested. Also, correction methods for the refinement of input temperature information are implemented and presented. Finally, linear and non-linear class characterization is evaluated through classification results analysis.

6.1 The problem of temporal linearity with vegetation parameters

Phenology refers to the seasonal changes in terms of vegetative growth and decline that vegetation types exhibit. Indeed, the phenology of a particular plant defines its seasonal pattern of growth, flowering, senescence and dormancy [Campbell, 2006], which can be monitored by remote sensing due to the changes in appearance and structure of the plant. With the onset of senescence, cell walls in the mesophyll tissue deteriorate and cause a decline in infrared reflectance; also chlorophyll declines and loses effectiveness to absorb visible radiation, causing thus a simultaneous increase in visible brightness [Campbell, 2006]. Satellite remote sensing imagery has long been used to record these spectral changes. Indeed, time series data registering these fluctuations are currently commonly exploited for the monitoring and analysis of vegetation types. Typically, approaches for characterisation and classification of land cover are based on the application of phenological evolution along a temporal axis, that is to say, exploiting the characteristics of phenological curves using time as an independent variable.

However, vegetation growth is a dependent variable determined by a variety of other factors, both genetical and environmental, such as for example moisture supply, light exposure, and soil structure and aeration. In particular, a most impactive variable is ambient temperature rather than the passing of time. The reason for this is that temperature directly affects the processes of photosynthesis, transpiration, respiration and absorption of nutrients and water, which in turn determine vegetative growth. Indeed, in the absence of stress conditions such as disease or drought, vegetation develops mostly determined by air temperature and in a cumulative manner. The accumulation of specific quantities of heat determine the passing from one developmental stage to another, such as for example from emergence to maturity.

Research shows that measuring accumulated heat over time rather than time alone provides more accurate physiological estimates of growth stages, more consistently predicting when plant development stages will occur [Miller et al., 2001]. The use of time as an analysis axis can be misleading, as warmer-than-normal days advance plant growth, while cooler-than-normal days slow it. As an illustrative example, two weeks of hotter-than-normal July days can advance lentils from green pod to harvest-ready [Miller et al., 2001].

As a consequence, when studying the phenological phases of vegetation types in relation to time, different profiles can be obtained when temperature conditions are different. When temporal profiles are applied to extract vegetation type features in order to apply machine learning algorithms, this phenomenon can greatly hinder classification tasks. The reason for this is the fact that spectral characterisation of a vegetation class at different timepoints can greatly vary for samples developing in different temperature conditions.

When dealing with remote sensing data over large areas, the diversity of growth environments becomes greatly significant, as generally greater areas encompass greater topographic variation. This affects vegetation in two ways. On the one hand, vegetation types requiring some temperature and environment characteristics appear non-uniformly distributed in the landscape. On the other hand, a vegetation type in different topographic or climatological conditions present different phenological cycles in time. In other words, location determines the appearance or non-appearance

of a given land cover type, and also, for a given vegetation class, phenological evolution in time can be different depending on location.

Thus, [Gruber, 1980b] demonstrated how landscape orientation and thus local conditions determined differences in land cover types presence and particularly variation in altitude distribution for a same species. For the Pyrenees, the author showed how the different stages of vegetation reached higher points on the southern exposures than on the northern ones. The former are commonly hotter and drier because they are exposed to more solar radiation, while the latter are generally cooler and moist. Also, [Gruber, 1980b] noted the appearance of some vegetation types reaching higher elevations on the southern exposures. This variability is captured in the presented method by use of topographical information for feature characterization, and by application of a stratified approach for sample set selection as described in Chap.5.

Similarly, geographical and topological factors determine temperature conditions, and thus plant phenological cycles. This is illustrated at a large scale in [Dethier et al., 1973], where the geographic spread of new vegetative growth in spring, known as the green wave, is observed as it increases in latitude in North America. Participating in this phenomenon, the presence of differences at local levels is necessary, generally sweeping upwards from lower to higher elevations [Campbell, 2006], and depending as said on topographical characteristics.

In this context, the use of satellite time series for the characterisation has commonly been based in the extraction of spectral features at given acquisition dates and the subsequent application of machine learning and classification algorithms. These approaches are based on the notion of learning the phenological characterisation of different land cover classes for the given timepoints, and applying the learnt model on the same time span of the acquisitions. However, although vegetation growth factors such as temperature often average from year to year, there are often cooler or warmer-than-average periods during the season than can advance or slow vegetation growth for that particular year. It is precisely this variation which prevents the applicability of models learnt on time series corresponding to equal periods for different years.

As a result, traditional approaches with satellite time series to model learning and classification can significantly fail when applied to either very large areas or to time series spanning the same period for different years. The common approach to characterisation, based on the extraction of features in the temporal axis, overlooks the physiological nature of spectral evolutions of vegetation types across regions and time spans. Thus, the application of a more consistent axis of analysis is necessary in order to obtain more accurate physiological estimates and a characterisation which is consistent with temperature effects on plant development.

In summary, the problem of linearity with vegetation parameters defines some main necessities. Firstly, a new general framework of phenological characterisation is needed, in order to truly capture phenological change as a physiological process. Secondly, this framework needs to provide a more consistent characterization of plant phenological cycles based on spectral responses, applicable across regions with different growth conditions.

Thirdly, a new reference axis of analysis, different to the classical linear temporal axis and

allowing usage not only across regions but also on different temporal spans, becomes also necessary. Fourthly, this axis needs to be based on an actual independent variable determining plant growth, in order to allow exploitation by machine learning algorithms and thus the enhancement of classification feasibilities. Finally, a fully automated solution is required in oder to obtain an operational methodology capable of managing the coming satellite time series.

It can be asserted that there is no standard approach giving a solution to the stated needs, nor applying the concept of physiological time instead of the traditional concept of time to machine learning algorithms.

Thus, a novel automatic non-linear sampling method allowing for the analysis of time series from a physiological non-linear axis in combination with the traditional temporal linear axis is proposed. It allows to characterise land cover classes based on a variable determining vegetation development. It is based on the inclusion of information on air temperature in order to guide the sampling process, determining extracted features for the subsequent application of machine learning. Finally, it permits an automatic and more consistent characterization of classes that can be applied across large areas, as it is shown by results in Sec.6.4.

6.2 Background

6.2.1 Phenological processes

A widely-accepted definition of phenology is "the study of the timing of recurrent biological events, the causes of their timing (...), and the interrelation among phases of the same or different species" [Lieth, 1974]. Over the last 15 years, the importance of phenology studies in relation to global change has been increasingly acknowledged [Richardson et al., 2013], with ongoing warming being linked to an earlier onset of vegetation growth in spring and an extension of the length of the growing season [Linderholm, 2006], as well as a later occurrence of end-of-season events [Richardson et al., 2013].

Thus, available literature on phenological studies is rapidly increasing, in particular related to climate change. Much research is focused on the impact of warmer temperatures on terrestrial ecosystems [Richardson et al., 2013]. [Julien et al., 2006] reports on changes in land surface temperatures and NDVI values, showing a trend of decrease in NDVI for Southern Europe, and an increase for the rest of the continent. These results evidence that arid and semi-arid areas in Southern Europe have become more arid, while the rest of Europe has experienced an increase in woodland.

Much effort has been placed on the study of the impact of climate change on phenophases of vegetation types (such as emergence, maturity and senescence or abscission). During the last decades, an average earlier onset of plant phases of 3.8 days per 1°C increase has been observed in Europe, with negative shifts for spring and summer stages and positive shifts for autumn stages [Rybski et al., 2011] [Estrella et al., 2009]. Other examples are that of cherry blossoming in Kyoto (Japan), where the negative shift between 1971 and 2009 has been measured in 7 days [Aono and Kazui, 2008], and a similar flowering pattern for other related species also in Japan

[Primack et al., 2009].

The correlation between air temperature and the onset of spring and summer vegetation stages is considered to be significant, while that for autumn stages seem to be weaker [Menzel et al., 2006] [Walther et al., 2002], with other factors becoming more important then, such as moisture and nutrient availability [Estrella et al., 2009]. Plant sensitivity to temperature changes is indeed non-linear (this non-linearity is independent from the application of a non-linear temporal axis resulting of the use of cumulative temperatures) and not uniform [Rybski et al., 2011], and a variety of complex interactions with other variables has also been proposed. For example, [Shrestha et al., 2013] analysed phenological responses of upland rice grown along and altitudinal gradient, with sterility explained very differently with genotype and environment effects at varying altitudes. Similarly, [Shen et al., 2011] analysed the influence of temperature and precipitation before the growing season on spring phenology in China, demonstrating interactions between the two variables.

In [Siebert and Ewert, 2012], a study was undertaken to analyze the patterns of oat crops phenological development in Germany in relation to temperature and day length. Similarly, the timing of tree phenological events is accepted to be tightly correlated to temperature and photoperiod [Doi and Katano, 2008]. In [Vitasse et al., 2009], great differences in phenological sensitivity to temperature for seven woody species in Europe are analysed, with low differences between populations of the same species.

6.2.2 Phenological variation and remote sensing

Satellite remote sensing data has been used to monitor the phenological variation of vegetative cover since several decades. Observing the phenological variation of land cover types throughout typically the course of a year, continental-scale classifications of land cover at coarse resolutions where already presented by [Tucker et al., 1985] and [Townshend et al., 1987]. However, it is only more recently that high resolution multitemporal datasets are available, permitting a much finer exploitation of satellite data. Indeed, with the new generation sensors of the coming satellite missions, an unprecedented potential for the monitoring of phenological development will be available.

Currently, the use of remote sensing data for phenological study is typically based on time series by application of the NDVI index, as it constitutes a biophysical parameter correlating with the photosynthetic activity of vegetation, and thus an indicator of vegetative status [Xie et al., 2008]. Other indexes, such as the perpendicular vegetation index (PVI) have also been used. For example, in [Guyot, 1999] the seasonal variations in the PVI over a five year period were analysed, together with ground-based observations, in order to establish the sufficient capability of medium resolution time series to monitor elevation variations in the phenology of broadleaf forests. Also, the authors noted a reduction in the growing season length with elevation.

Another phenological study based on NDVI dynamics is found in [Alcaraz-Segura et al., 2009], where a characterization of major Iberian vegetation types was attempted, as well as an analysis of the climatic controls of NDVI dynamics. To this aim attributes of the NDVI seasonality were studied (such as annual mean, relative range, maximum and minimum values, and interannual

variabilities) and the relationship of the former two attributes with climatic variables (such as annual precipitation and mean annual temperatures). The authors noted an interplay between precipitation and temperature. Also based on satellite data, [Hwang et al., 2011] identified elevation (following temperature lapse rates) and radiation as main topography-mediated controls on local vegetation, with an additional variable impact of precipitation. This is coherent with the theory that at mid and high latitudes vegetation phenology is controlled by temperature and photoperiod, while in the tropics and semiarid areas the main factor is rainfall [Jolly et al., 2005].

Phenological dynamics has also been applied for classification, by means of applying firstly machine learning on the spectral information of a selected set of acquisitions and then classifying based on the learnt model on the same time span. Some approaches include also non-spectral features. In [Galiano and Chica-Olmo, 2012] extracted spectral primitives from selected timepoints in the time series are combined with additional information such as texture, land surface temperature, and elevation information. Similary, in [Jia et al., 2014] land cover classification is undertaken on Landsat data combined with phenological features extracted from MODIS NDVI data. These features include the beginning and ending dates of the growing season, the length of the growing season, seasonal amplitude, and the maximum NDVI values, extracted with the TIMESAT tool. The authors assert an increase in accuracy for classifications with single Landsat data.

Indeed, the extraction of phenological markers such as start, end, duration and maximum peak of the vegetation season, presents several difficulties. Sensitive factors influencing the extraction include [Homolová et al., 2013]: a) temporal resolution, b) the presence of missing or noisy data, c) the magnitude of the seasonal amplitude, and d) the method for extraction. As for the latter, there is no agreement on a globally appropriate phenological marker extraction method [Schwartz and Hanes, 2010]. A fitting approach is represented by [Zhang et al., 2003], based on the piecewise fitting of logistic functions on vegetation index data. However, this type of approach is devised for the analysis of vegetation phenology rather than constituting an operational solution for the extraction of markers for classification purposes. Similarly, another approach is provided by the well-known TIMESAT software [Jönsson and Eklundh, 2004], which uses three different least-squares methods for the processing of time series, constrained though to the use of at least three consecutive years of data.

The use of cumulative temperature and degree days in combination with remote sensing data has in general been directed towards the identification of optimal accumulated degree day counts for the prediction of phenological stages of specific species and towards the estimation of cumulative temperatures themselves based on remote sensing data. For example, in [Hassan and Rahman, 2012], MODIS-based 8-day surface temperature and reflectance data were used to calculate accumulated growing degree days and used to study the phenological stage of deciduous leaf-out over the deciduous-dominant stands in the Alberta province of Canada. Also, in [Hassan et al., 2007b], the mapping of cumulative temperatures is also based on MODIS composites and a vegetation index, with the assumption that daily mean air temperature is approximated by daily mean surface temperature computed on radiation. The approach is further developed in [Hassan et al., 2007a] with the application of Landsat-7 data for the derivation of EVI estimates.

In conclusion, although much research has analysed the impact of temperature on phenophases

of vegetation species, the use of temperature data in combination with remote sensing imagery is often limited to the establishment of phenological phases for targeted species. Also, the study of phenological markers based on satellite data has been traditionally based on a linear temporal analysis of phenological response and, to the author's knowledge, no previous literature exists introducing temperature data as a virtual axis for characterization and classification of land cover based on satellite time series.

6.3 Proposed method

Reviewed literature in the preceding section using remote sensing data for either phenological studies or land cover classification has something in common: the use of a temporal axis for the characterisation of land cover classes. This applies from the first global mappings and work on phenological indexes to the recent interannual variation and phenological marker extraction studies. In particular, literature making use of phenological characterisation for classification with machine learning algorithms has been to our knowledge generally based on temporal references, for example determined by given acquisition times, or based on linear sampling based on time.

6.3.1 Description

In this context, an essentially different approach is presented here. It is based on the use of non-linear temperature-based interpolation and sampling for the subsequent application of machine learning algorithms. The proposed method allows for a better generalisation in the characterization of phenological responses, as it employs the physiology-related variable, that of temperature, in order to enhance model learning. It ultimately contributes to the automated generation of land cover maps over large areas.

Thus, the approach is based on axis transformation as a key aspect for the application of machine learning algorithms. The main idea is that a change in the characterisation axis or chosen independent variable for the extraction of multi-temporal features can allow for an improved characterisation of land cover types. The hypothesis is that such a representation can better capture the development of vegetation from one stage in their lifecycle to another independently of location, and thus enhance subsequent model learning and classification tasks.

In this sense, the presented method is inspired by the analysis technique used in purely botanical and phenological studies and based on the cumulative view of temperature in relation to phenological evolution that has been described in Sec.6.2. In other words, it applies cumulative thermal times from the phenological science for the extraction of machine learning features. Thus, an ancillary temperature dataset guides interpolation for missing information processing, and resampling is determined by temperature data.

Fig.6.1 broadly illustrates the proposed method to automatically obtain non-linear temperature-based sampling of features for feature vector construction. It integrates in the proposed general framework for automatic large area land cover map generation given in Fig.3.1. Given ancillary temperature listings provided by SAFRAN, as described in Chap.2, extraction is undertaken in order to obtain accumulated temperatures and degree days (another approach for measuring physiological

time) for the available time series data acquisitions. The concepts of accumulated temperature and degree days are further described in Sec.6.3.2. These are spatialised by using available spatial information corresponding to the SAFRAN grid. Then, and given the initial SAFRAN spatial resolution of $8km$, height and radiation correction is applied to the data to improve estimations at the 30m LANDSAT resolution. Corrections are further described in Sec.6.3.4. Then, the approach of linear local interpolation explained in Chap.4 is applied on the temperature axis, and temperature-based sampling is undertaken to obtain new multitemporal data for feature vector construction. Based on the new feature vectors, machine learning and subsequent classification tasks can be undertaken for large area land cover mapping.

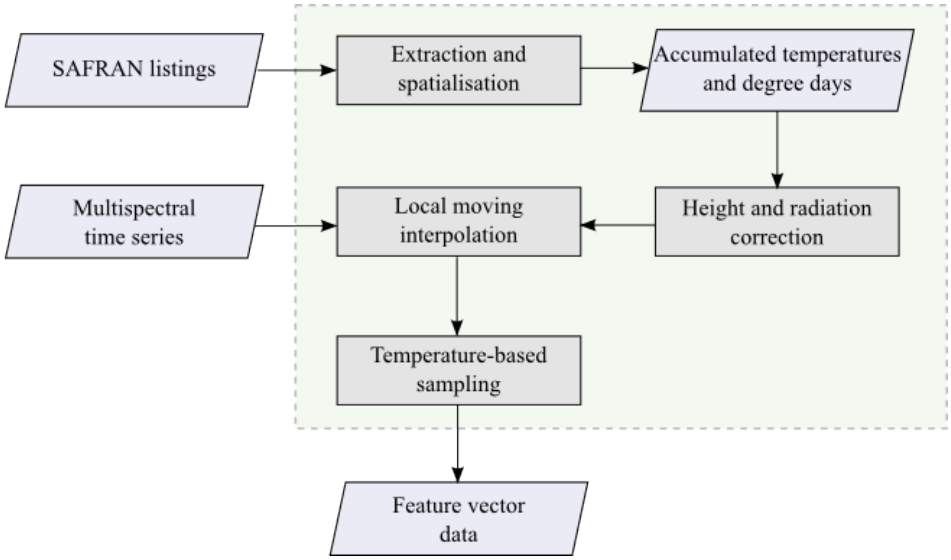


Figure 6.1: Proposed automatic non-linear sampling approach for dealing with spectral variability across large regions. Temperature information is incorporated so that vegetation phenology is more robustly modelled, and sampling is undertaken in the temperature axis. See Fig.3.1 for integration in the proposed general framework.

6.3.2 Sum of temperatures and degree days

Following the works of Lehenbauer and Livingstone at the beginning of the 20th century, numerous research studies have shown that the speed of growth of vegetation grows progressively with temperature until around 30°C and then decreases rapidly towards null until around 40°C, when lethal temperatures are reached [Guyot, 1999]. This phenomenon is illustrated by the author, showing how temperature acts on corn growth, as adapted in Fig.6.2.

Fig.6.2 shows how the curves corresponding to either hourly growth or leaf appearance can be approximated to a line for temperatures between 10 and 30°C. In natural conditions, temperature varies throughout the day and from one day to another, so growth is also variable in time [Guyot, 1999]. In this context, the method of sum of temperatures is based on the approximation that considers development speed linearly varying with temperature. Given T_0 the minimum temperature necessary for development, a the slope of the regression line, which is dependent on

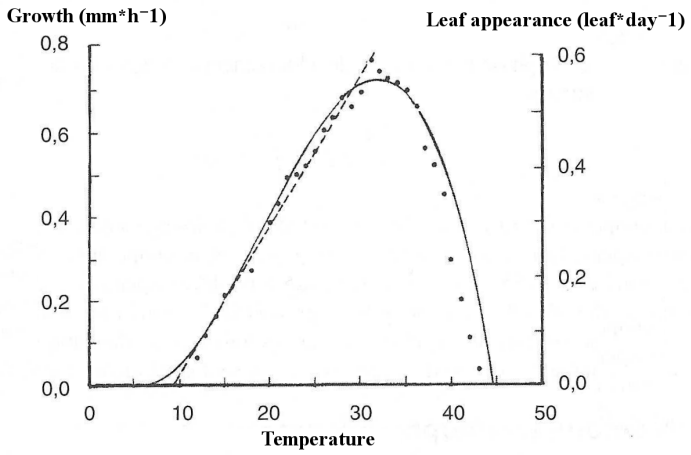


Figure 6.2: Temperature effect on hourly growth (points) and leaf appearance (line) of corn seedlings. From [Guyot, 1999].

vegetation variety, the speed of development can be formulated as in Eq.6.1:

$$V(T) = a * (T - T_0) \tag{6.1}$$

Then, given t_1 and t_2 as the start date for two phenological stages, and ζ an arbitrary growth translating the difference between an initial and a final stage of the vegetative type, the following Eq.6.2 can be formulated:

$$\zeta = a \int_{t_1}^{t_2} (T - T_0) dt \tag{6.2}$$

This concept of cumulative temperature as a driver of plant growth determining passage from one phenological stage to another has been long applied since R. Marsham, who founded modern phenological science by recording first occurrence dates of phenological events. Indeed, many of these accounts have been traditionally locally developed by gardeners, horticulturists and land owners in order to predict plant development and pest susceptibility. Table 6.1 exemplifies characteristic cumulative temperatures for two different varieties of sunflower as given by [Guyot, 1999]:

Variety	Sprouting to blooming	Sprouting to maturity
Cerflor sunflower	660(60)	1500(200)
Mirasol sunflower	740(60)	1600(146)

Table 6.1: Characteristic cumulative temperatures in °C for sunflower varieties (confidence interval between parenthesis).

Schematically, the relation between cumulative temperature, time and phenophases (here exemplified with heading) is illustrated for a variety of corn in Fig.6.3. A reference cumulative temperature average is shown with a continuous line in contrast with examples of cooler-than-average and warmer-than-average evolutions, showing differences in the temporal occurrence of the start of

the chosen phenophase.

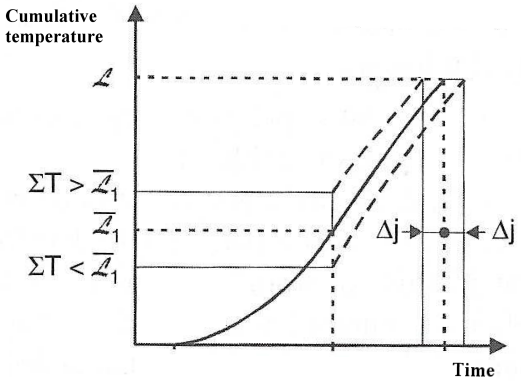


Figure 6.3: Variations in the sum of temperatures for a variety of corn and resulting effects on temporal occurrence of the heading phase. From [Guyot, 1999].

The method of cumulative sums can be further refined by introducing lower and upper thresholds of growth, so that only temperatures favoring growing conditions are taken into consideration when applying a cumulative approach. These optimal temperature sums rely on the correct selection of thresholds, which is dependent on vegetation species and even subspecies, as some are more resistant to extreme temperatures than others. A typical selection of average thresholds when targeted species (or their location) are unknown is 10°C and 30°C. This is based on the assumption that most plants do not grow any faster below or above these reference temperatures; this is the approach that has been applied here.

An alternative to the use of cumulative temperatures is the use of degree-days. Another physiological time measuring approach, one degree-day is defined as one day (24h) with one degree of temperature above the lower development threshold. Considering common daily oscillations of temperature, Fig.6.4 illustrates the ranges of temperatures for a given species (with development range between the illustrated upper and lower thresholds) counting towards cumulative degree-days.

A variety of methods exists for the calculation of degree-days. As explained, accumulations are based on the integral of the area under the temperature curve and between thresholds, so methods generally take into account daily minimum and maximum temperatures and assume either triangle or sinusoidal curves in order to approximate the area. Due to the nature of SAFRAN temperature data listings, which contain hourly temperature information, growing degree-days (GDD) here have been calculated with $n = 24$ as given in Eq.6.3, with T_0 being the base temperature below which no plant growth is expected:

$$GDD = \sum_{1}^{n} (T - T_0)/n \tag{6.3}$$

Thus, the principle of the method here presented is to extract features for machine learning along the curve of temperatures instead of along the traditionally used temporal axis. Given the NDVI development of a pixel based on linear temporal sampling with a number of timepoints, and the

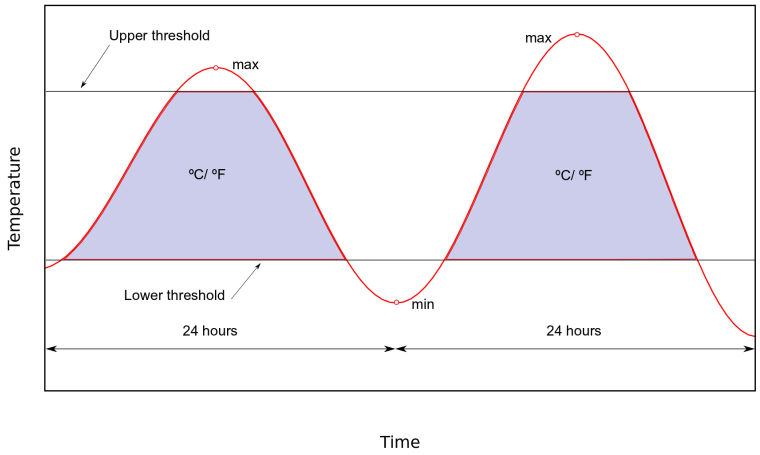


Figure 6.4: Illustration of the concept of degree-day using upper and lower development thresholds.

cumulative temperature counts recorded for the considered crop year, the transformed phenological characterization after temperature-guided non-linear sampling can be computed. A transposition of axes is put in place in this way, in order to use temperature as an analysis variable instead of time. This characterization is intended to more accurately capture phenological evolution, providing greater independence from climatic factors, reduced variance in the characterization, and based on traditional temporal series information.

6.3.3 SAFRAN temperature information extraction and spatialisation

General characteristics of SAFRAN meteorological data are given in Chap.2. Here, details about the extraction and spatialisation of temperature information from the SAFRAN datafiles are provided.

SAFRAN data are available in annual binary files from which listings for each of the points of the SAFRAN network as described in Chap.2 can be extracted. Yearly information comprises hourly data records from 1 August at 0h to 31 July at 23h. The following parameters are available: visible radiation (W/m^2), liquid precipitation ($kg/m^2/sec$), specific humidity at 2m (kg/kg), infrared radiation (W/m^2), solid precipitation ($kg/m^2/sec$), temperature at 2m (K), and horizontal wind speed at 10m (m/sec). Converted to NetCDF format by CERFACS, data for the South of France have been then extracted in CESBIO for the period of interest by V.Rivalland.

The native data projection is conic conforme, Lambert-II extended (Lambert-II Carto Zone Centre), NTF system, and the applied target projection Lambert 93. These have been applied after the transformation of SAFRAN temperature listings into a spatialised matrix (see Alg.3) and are given in Tab.6.2. The characteristics of the SAFRAN network for France have been used for spatialisation (Alg.3), as given in Tab.6.3 [Pagé, 2008]. Finally, network points are defined in an ASCII file as exemplified in Tab.6.4, with the following parameters: original SAFRAN grid number, SAFRAN France grid number, longitude, latitude, Xlamb (m) Lambert-II extended, Ylamb (m) Lambert-II extended, altitude (-999 for sea/ocean), and corresponding Symposium zone (another zone nomenclature).

	Lambert II extended	Lambert 93
Standard parallel 1	45.89891888890	44
Standard parallel 2	47.69601444440	49
Origin latitude	46.79999999999	46.5
Origin longitude	2.33722916670	3
False easting	600000	700000
False northing	2200000	6600000

Table 6.2: Applied map projection specifications

Δx = 8000.0 m	Minimal coordinate X = 60000 m
Δy = 8000.0 m	Maximal coordinate X = 1196000 m
Points in X = 134	Minimal coordinate Y = 1617000 m
Points in Y = 143	Maximal coordinate Y = 2681000 m
Points = 19162	Effective points = 9892

Table 6.3: France SAFRAN network characteristics. "Effective points" stands for points without missing values.

1350	1	2.62204	51.12120	620000	2681000	6	1
1507	2	2.16654	51.04970	588000	2673000	5	1

Table 6.4: SAFRAN point coordinates specification example

Thus, using all specifications given in Tabs.6.2, 6.3 and 6.4, each point can be placed in a three-dimensional matrix representing France in the X and Y, and with each layer in the Z dimension giving a sum of temperatures for each acquisition date. Also, the original Lambert-II extended projection is projected to a Lambert-93 projection matching the used LANDSAT time series. The pseudocode of the extraction algorithm is given in Alg.3:

The procedure of data extraction from SAFRAN point listings is the following: spatialisation is undertaken with cumulative temperatures at the SAFRAN 8km spatial resolution, and then resampled to the target Landsat 30m resolution. Fig.6.5 gives, for given acquisitions between 01/02/2010 and 27/07/2010, the resulting temperature maps for a region in tile L930563. With illustrative purposes, maps are plotted giving relative temperatures, thus showing different temperature evolutions and thus cumulative temperatures for different point locations.

6.3.4 Height and radiation correction

SAFRAN data has an original 8 km spatial resolution. After spatialisation, in order to resample the extracted temperature data to the Landsat 30 m spatial resolution, two factors affecting temperatures for a given site need to be taken into account: those of height and slope orientation. The reason for this is that each original pixel at 8 km resolution averages temperature results for a wide geographic area that can be rugged to different degrees. Mountainous landscapes,

Algorithm 3: Safran temperature data extraction and spatialisation

Input : SAFRAN listings, desired span, reference Landsat data
Output: temperature matrix corresponding to input listing, span and reference area
READ input list;
READ desired span;
GET grid conversion;
while *not end of 24h-day list* **do**
    COMPUTE daily temperature units (average sum of temperatures or degree-day count) threshold observance;
while *not end of dates in desired span* **do**
    while *not end of of point listing* **do**
        STORE cumulative temperature units;
        GET vector location in matrix;
ASSIGN metadata to matrix:
    COMPUTE most extreme SAFRAN map point to N and W;
    COMPUTE SAFRAN image origin given spacing;
    ASSIGN Lambert II extended projection;
while *not end of matrix* **do**
    ASSIGN value vectors;
READ reference Landsat data **while** *not end of target matrix* **do**
    COMPUTE reprojection into target geometry:
        linear interpolation;
        Lambert II extended to Lambert 93 projection;
PRINT temperature matrix corresponding to reference area;

for example, can present very different temperatures in relatively close geographical points of a certain region depending on the location's height and orientation. These different temperatures are averaged at low spatial resolutions such as that of the SAFRAN data, but need to be taken into account when resampling at higher spatial resolutions.

Height impact on temperature is based on a well-known phenomenon by which temperature decreases when altitude increases. This lapse rate has been quantified by a variety of studies ranging from $-0.89^{o}C(100m)^{-1}$ for dry air to about $-0.4^{o}C(100m)^{-1}$ for warm saturated air [Rolland, 2003], depending on a variety of factors. An global mean approximation of $-0.65^{o}C(100m)^{-1}$ is given by [Barry and Chorley, 1987] and is applied in this work.

Slope impact on temperature in the northern hemisphere is observable in increased temperatures for south-facing slopes and decreased temperatures on north-facing slopes in relation to a flat surface at the same elevation. Indeed, in the northern hemisphere, south facing slopes may receive up to six times more solar radiation than their counterparts [Auslander et al., 2003]. Due to this different annual amount of solar radiation falling on a unit area of the surface, north slopes tend to be colder at the same elevation than south slopes. Here, the ratio of slope/flat surface radiation can be used in order to adjust temperatures for slopes with different radiant energy inputs following [Hungerford et al., 1989].

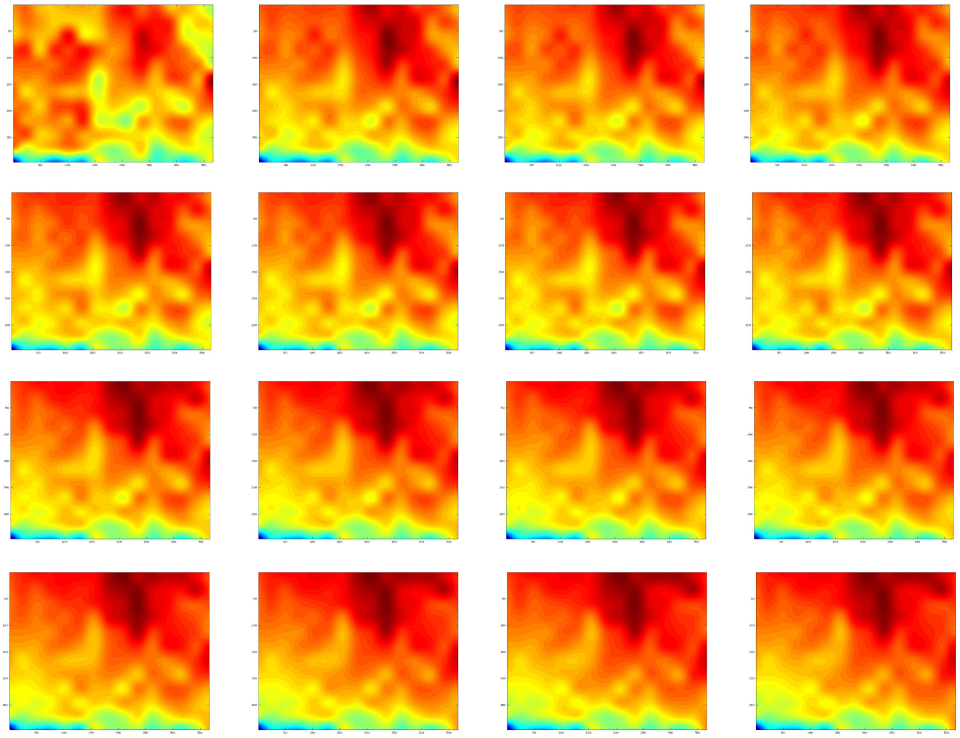


Figure 6.5: Relative cumulative temperature maps for given acquisitions between 01/02/2010 and 27/07/2010 for an example L930563 region (ranges are different for each image, as they correspond to cumulative temperatures; only relative variation is highlighted). Red areas represent hotter areas; blue areas represent colder areas (or missing). Changing variation in the distributions are observable in time, resulting in different cumulative temperatures over time accross regions.

Another aspect to be taken into account is that of canopy differences, as temperature is affected by the characteristics of the energy exchange slope surfaces [Hungerford et al., 1989]. Bare slopes can present slope-related differences of up to 18°C, while closed canopy forests can show no differentials. As a consequence, temperature is also adjusted by applying a vegetation index.

Thus, the following formulas have been applied in order to obtain a more accurate resampling of the SAFRAN data taking into account height, slope orientation and canopy characteristics. These formulas are inspired by calculations applied in the Mountain Microclimate Simulation Model (MTCLIM) by [Hungerford et al., 1989]. The original formulas employed Leaf Area Index values and applied calculations to obtain site-specific results based on close base stations. Here, instead, location-specific results are obtained from averaged wide-area values with Eqs.6.4:

$$\begin{aligned} T_s &= T_a - T_l((E_s - E_b)/1000) + (Rat)(1 - I_s/I_m) \\ T_n &= T_a - T_l((E_s - E_b)/1000) - 1/(Rat)(1 - I_s/I_m) \end{aligned} \tag{6.4}$$

where:
T_s=site temperature (south aspect)

T_n=site temperature (north aspect)
T_a=base station average temperature
T_l=elevational lapse correction (279.65K=6.5°C)
E_s=site elevation
E_b=base elevation
Rat=ratio slope radiation/flat surface radiation
I_s=site NDVI index
I_m=maximum NDVI index

For the height correction, the lapse rate of $-0.65^oC(100m)^{-1}$ is translated into the original SAFRAN units to obtain $279.65(1000m)^{-1}$. For the slope correction, ratios between slope and flat surface radiation are applied as described below. Finally, for canopy correction, NDVI values have been applied.

Calculation of the ratios between slope and flat surface radiation was undertaken at CESBIO by S.Gascoin using the solar radiation toolset of the Spatial Analyst toolbox of the ArcGIS software. Radiation estimations of direct, diffuse and total radiation were generated for all months in the considered period, as illustrated in Fig.6.6.

Figure 6.6: Example radiation estimations generated at CESBIO by S.Gascoin. Images correspond to direct, diffuse and total radiation estimations (in WH/m_2) for a given tile during the month of January. The range for total radiation in the given example is [0-357383]

Total solar insulation is calculated here as the sum of direct and diffuse radiation, disregarding the impact of reflected radiation. Under clear sky, direct radiation is a function of the angle between the Earth surface and the Sun [Pierce et al., 2005]; the rotation of the Earth determines daily changes in the orientation of the Sun, and its tilt and orbit govern yearly changes in solar orientation and period. This geometry of radiation underlies the hemispherical viewshed algorithm developed in [Fu, 2000a] [Fu, 2000b] and implemented in the software used to generate the radiation maps applied here. Fig. 6.7 shows a comparison for SAFRAN resampled data at 30m without correction and with correction for height, radiation and canopy characteristics.

Algs.4, 5 and 6 summarize the processing steps implemented in order to perform height and radiation correction of the extracted SAFRAN temperature information. Alg.4 describes the

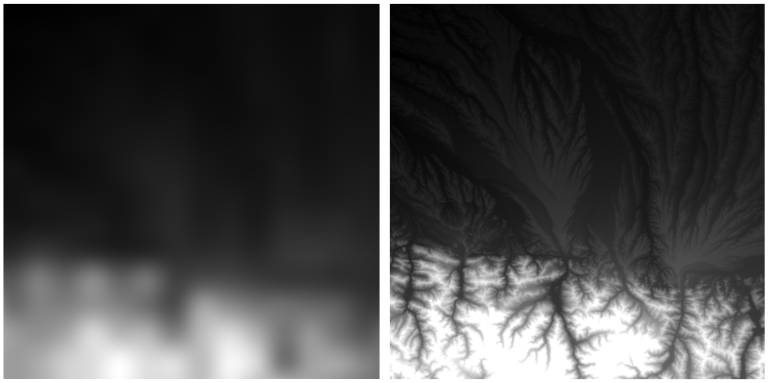


Figure 6.7: Comparison of SAFRAN data resampled to 30m spatial resolution without correction and with correction (tile L930463).

preliminary extraction of the reference elevation map based on SAFRAN data at 8km spatial resolution. Based on the SAFRAN coordinates, the procedure gets elevation data for each of the SAFRAN points and positions them in a matrix. By computation of an extreme location and application of data on spacing, the native SAFRAN Lambert II extended projection can be assigned to the matrix. Then, reprojection into the target geometry of Landsat data is applied, using a Lambert 93 projection and linear interpolation. In this way, a elevation matrix corresponding to targeted Landsat areas is obtained.

Alg.5 describes the implemented approach to correct the cumulative temperature matrix by using target resolution elevation information and based on Eqs.6.4. Dealing with vectors of cumulative temperatures rather than single values, the following computation is applied for each pixel i in the series position $k = 1...n$ following the notation of Eqs.6.4, as given in Eq.6.5:

$$\begin{aligned} T(i)^k_{s/n} &= T(i)^k_a - (E(i)_s - E(i)_b) * \alpha, \quad \alpha = T_l/1000 \\ \sum_{k=1}^{n} T(i)_s/n &= \sum_{k=1}^{n} [T(i)^k_a - (E(i)_s - E(i)_b) * \alpha] \\ \sum_{k=1}^{n} T(i)_s/n &= \sum_{k=1}^{n} T(i)^k_a - n * [(E(i)_s - E(i)_b) * \alpha] \end{aligned} \tag{6.5}$$

As correction is applied using Kelvin degrees, T_i becomes 279.65 in order to apply the formerly explained approximation of $-0.65^oC(100m)^{-1}$ (6.5°C=279.65K). The multiplier n is calculated by univocally identifying each day with an identifier referenced to 01/01/1970. A cumulative temperature matrix corrected for elevation is thus obtained.

Finally, Alg.6 describes the algorithm implemented for radiation correction. Given aspect data normalised north-south, correction is applied differently according to orientation (north or south) on the temperature matrix corrected for elevation. For each vector in the matrix, corresponding NDVI and radiation ratio values are calculated, and correction formulas for radiation thus applied based on orientation. In this way, a final temperature matrix corrected for elevation and temperature is generated.

Algorithm 4: SAFRAN digital model extraction

Input : SAFRAN elevation information (8km resolution), reference Landsat data
Output: SAFRAN derived digital elevation model for Landsat area
READ SAFRAN point coordinates;
GET grid conversion;
while *not end of of point listing* **do**
    GET elevation value;
    GET cell location in matrix;
ASSIGN metadata to matrix:
    as in Alg.3
while *not end of matrix* **do**
    ASSIGN cell values;
READ reference Landsat data **while** *not end of target matrix* **do**
    COMPUTE reprojection into target geometry:
        linear interpolation;
        Lambert II extended to Lambert 93 projection;
PRINT elevation matrix corresponding to Landsat data;

Algorithm 5: Cumulative temperatures elevation correction

Input : temperature matrix, SAFRAN elevation model (8km), target resolution digital elevation model (30m)
Output: temperature matrix corrected for elevation
COMPUTE elevation correction:
while *not end of temperature matrix* **do**
    READ cumulative temperatures vector;
    READ SAFRAN elevation data;
    READ 30m digital elevation data;
    COMPUTE n based on reference date (see Eq.6.5);
    COMPUTE correction vector:
        $T_a - n * (E_s - E_b) * 279.65/1000$ (see Eq.6.4);
            Kelvin multiplier;
    APPLY elevation correction vector;
PRINT cumulative temperature matrix corrected for elevation;

6.4 Experiments and results

Assessment of the different approaches to non-linear temporal sampling is undertaken based on their impact on classification results, following the experimental setting described in Sec.4.4.1. In this way, accuracy is assessed in terms of average global accuracy, average percentages of well-classified pixels per class, standard deviation of average global accuracy, average kappa statistic, and average class percentage accuracy for 20 learning sets. A total of 680 classification tests are presented in this section.

Algorithm 6: Cumulative temperatures radiation correction

Input : aspect data (normalised north-south), elevation-corrected temperature matrix, radiation ratio matrix (monthly data), NDVI matrix

Output: temperature matrix corrected for temperature and radiation

COMPUTE radiation correction:

while *not end of temperature matrix* **do**
  FIND radiation ratio matrix position;
  FIND NDVI matrix position;
  if *south aspect* **then**
    COMPUTE correction vector:
      $Ta + n * (Rat)(1 - I_s/I_m)$ (see Eq.6.4)
  else
    COMPUTE correction vector:
      $Ta - n/(Rat)(1 - I_s/I_m)$ (see Eq.6.4)
  APPLY radiation correction vector;

PRINT corrected cumulative temperature matrix;

This chapter builds on the findings presented in Chap.4 and it is developed following the automatic configuration presented in Chap.3. For comparability purposes, advances in Chap.5 are only incorporated in the final section of the current chapter.

As in Chap.3, incorporated choices include the following. Firstly, missing information processing is based on the application of linear interpolation, as supported by results in Tab.4.6 and Tab.4.12. Secondly, the sampling number is set to 500 pixels per class (8500 in total for the studied configuration), as it represents a good compromise between accuracy improvement and sampling ratio (Tab.4.10). Thirdly, the enhancement of separability between classes by addition of non-spectral information shown in Tab.4.11 is also applied here. Finally, a temporal sampling of 12 virtual points is used, also giving a good compromise between accuracy increase and processing costs in terms of memory and, as it will be shown later, computation times.

Several approaches for non-linear temperature-based temporal sampling have been in this way implemented and tested, making use of the sum of temperatures and degree day approaches described in Sec.6.3.2. In this way, a novel approach to characterization for classification is presented, based on the joint application of linear and non-linear temperature-based temporal sampling. These implementations have been compared against the linear approach to temporal sampling that has been extensively used in the literature to this date for characterisation for classification purposes.

Results are presented in tables summarizing average results based on 20 independent sample sets and classifications. For the reference dataset here applied (L930463), the produced number of tests, and the given configuration (where training pixels are not considered for evaluation), each results entry line with classification averages in this section is thus summarizing 20 full area classifications and evaluation, for each of them, against 747523 pixels of ground truth ($(756023 - 17 * 500)$).

Tab.6.5 summarizes the results for 20 classifications based on a linear time-based sampling approach (TS) based on NDVI and topographic information. The TS approach, currently used in the literature for land cover classification, is determined by the characterisation of land cover spectral response captured in the time series along a temporal axis.

TS	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	32.698	75.542	44.815	58.350	71.475	21.934	50.432	68.428	56.239	92.836	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	67.028	26.29	30.931	52.406	49.737	52.066	81.194	0.559	0.008	0.468	54.847

Table 6.5: Percentages of well-classified pixels, per class, for a linear time-based sampling approach (TS) based on NDVI and topographic information, with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Results in Tab.6.5 show a modest average classification accuracy of 54.847% with high variability in accuracy reaching as low as 21.934%, 26.29% and 30.931% for "permanent herbaceous crops", "other winter crops" and "other summer crops".

Tab.6.6 gives results obtained 20 independent classifications based on non-linear temperature-based sampling on NDVI and topographic information. Two approaches applying sum of temperatures (ST) information with reference to the first day of October and with lower developmental threshold temperatures of 0 (ST^{0}_{oct}) and 10°C (ST^{10}_{oct}), respectively, have been implemented and tested. No linear temporal sampling is applied and only the temperature axis is considered: feature vectors have been obtained only by characterization based on temperature.

ST^{0}_{oct}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	25.825	74.769	46.847	55.666	65.496	22.922	47.878	57.184	56.285	91.592	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	47.811	27.083	31.068	50.696	47.495	34.899	76.111	0.541	0.008	0.446	50.566
ST^{10}_{oct}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	25.825	74.769	46.847	55.666	65.496	22.920	47.878	57.181	56.285	91.592	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	47.815	27.083	31.069	50.686	47.500	34.890	76.111	0.541	0.008	0.446	50.566

Table 6.6: Percentages of well-classified pixels, per class, for a non-linear temperature-based sampling approach based on NDVI and topographic information with sum of temperatures (ST: October, 0°C and 10°C) with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Results in Tab.6.6 show a decrease in average class classification accuracy to 50.566% in comparison to the linear temporal sampling approach average of 54.847% for both minimum developmental thresholds. When analysed class per class, only 3 classes show a non-negligible increase in accuracy, among which two winter classes are found, "barley" and "other winter crops", and also "summer crops", a class closely linked to the former two. This is consistent with a chosen characterisation referenced to the first day of October, the chosen biofix date, indicating the

cumulative temperature starting point. This biofix date varies according to species and is typically chosen depending on planting date.

Differences in Tab.6.6 between the tested approaches with lower developmental threshold temperatures of 0 and 10°C are of a very small order, mostly lost in the rounding for table presentation. In general, it can be said that linear temporal sampling performs thus better than sole temperature axis characterisation with the chosen parameters. Close scores, however, point towards the potentiality of a joint application of both approaches, which is implemented and tested in this chapter with positive results.

Tab.6.7 summarises results for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information. Here, both sum of temperatures (LST) and degree-days (LDD) are jointly applied with the commonly used linear approach for improved results. The LST notation refers to linear (L) and sum of temperatures (ST), and LLD stands for linear and degree-day (DD). The biofix date remains the first date of October and again for both approaches also the two developmental lower thresholds are used, 0°C (LST^{0}_{oct}, LDD^{0}_{oct}) and 10°C (LST^{10}_{oct}, LDD^{10}_{oct}).

LST^{0}_{oct}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	34.184	77.340	46.315	58.793	73.480	30.021	51.815	70.495	60.085	92.463	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	70.326	28.898	33.096	54.664	48.013	59.036	81.824	0.575	0.008	0.488	57.109
LST^{10}_{oct}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	34.184	77.340	46.315	58.793	73.480	30.021	51.815	70.495	60.085	92.463	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	70.326	28.898	33.096	54.664	48.013	59.036	81.824	0.575	0.008	0.488	57.109
LDD^{0}_{oct}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	33.054	75.721	48.513	57.898	71.631	23.515	50.039	68.292	59.943	92.413	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	68.361	29.261	31.333	51.378	49.735	54.114	81.176	0.557	0.006	0.467	55.669
LDD^{10}_{oct}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	34.724	75.760	46.613	58.178	71.332	23.474	49.944	68.632	58.057	92.587	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	66.950	28.419	31.448	52.216	49.801	50.243	81.204	0.561	0.008	0.470	55.269

Table 6.7: Percentages of well-classified pixels, per class, for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information with sum of temperatures (LST: October, 0°C and 10°C) and degree-days (LDD: October, 0°C and 10°C), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Results in Tab.6.7 improve from the previous linear temporal sampling approach score of 54.847% and the exclusive temperature-based characterisation yield of 50.566% to results ranging between 55.269% and 57.109% for the different tested approaches. The sum of temperatures approaches (LST) obtain better results than the degree-day approaches (LDD), indicating a better suitability of the sum of temperatures method, reaching 57.109% (LST^{0}_{oct}, LST^{10}_{oct}). Again, differences between the two tested lower temperature thresholds are mostly lost in rounding for table presentation, although more observable in the degree-day case, with slightly better results for

the 0°C baseline. Best results are attained with the sum of temperatures method (LST), with an average class accuracy of 57.109 and presenting an improved classification score for 15 classes out of 17.

Similarly, Tab. 6.8 summarises results for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information. However, the biofix date is set here to the first date of April. Both sum of temperatures (LST) and degree-days (LDD) are jointly applied with the commonly used linear approach for improved results, and again for both approaches also the two developmental lower thresholds are used, 0°C (LST^{0}_{apr}, LDD^{0}_{apr}) and 10°C (LST^{10}_{apr}, LDD^{10}_{apr}).

LST^{0}_{apr}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	33.747	76.723	50.095	58.216	73.225	31.228	50.802	70.440	58.026	92.910	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	70.489	26.18	31.988	53.469	48.186	58.444	79.269	0.569	0.006	0.480	56.673
LST^{10}_{apr}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	33.747	76.723	50.095	58.216	73.225	31.228	50.802	70.440	58.026	92.910	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	70.489	26.18	31.988	53.469	48.186	58.444	79.269	0.569	0.006	0.480	56.673
LDD^{0}_{apr}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	33.530	76.3196	44.198	58.379	71.912	25.650	50.474	69.566	58.057	93.383	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	66.254	29.045	31.719	51.866	49.172	57.608	81.602	0.561	0.009	0.470	55.808
LDD^{10}_{apr}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	34.031	75.930	46.604	58.431	71.968	23.750	50.500	68.140	58.320	92.363	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	66.266	28.516	30.874	51.694	50.069	54.559	81.074	0.560	0.008	0.469	55.476

Table 6.8: Percentages of well-classified pixels, per class, for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information with sum of temperatures (LST: April, 0°C and 10°C) and degree-days (LDD: April, 0°C and 10°C), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Results in Tab.6.8 also improve from the previous linear temporal sampling approach score of 54.847% and the exclusive temperature-based characterisation yield of 50.566% to results ranging between 55.476% and 56.673% for the different tested approaches. Again, the sum of temperatures approaches (LST) obtain better results than the degree-day approaches (LDD), indicating a better suitability of the sum of temperatures method, reaching 56.673% (LST^{0}_{apr}, LST^{10}_{apr}). Here, differences between the two tested lower temperature thresholds are totally lost in rounding, although more observable in the degree-day case, with slightly better results for the 0°C baseline. Best results are attained with the sum of temperatures method (LST), with an average class accuracy of 56.673 and improving the classification score of 13 classes out of 17. Two winter classes, "rape" and "other winter crops" experience a decrease in accuracy, with is consistent with the selection of a summer crop reference date. Although superior to the reference linear sampling case, results do not reach the improvement attained when applying an October biofix date.

Tab.6.9 presents results for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information and combining

the two previously applied biofix dates, that is to say, the first day of October and the first day of April, in an attempt to better characterise both winter and summer crops, as well as land cover classes with biological developments affected by temperature conditions following the two chosen season timepoints. Both sum of temperatures (LST) and degree-days (LDD) are jointly applied with the commonly used linear approach for improved results, and again for both approaches also the two developmental lower thresholds are used, 0°C ($LST^{0}_{apr-oct}$, $LST^{10}_{apr-oct}$) and 10°C ($LDD^{0}_{apr-oct}$, $LDD^{10}_{apr-oct}$).

$LST^{0}_{apr-oct}$	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	35.201	77.442	47.257	59.042	73.931	33.029	52.21	71.299	60.264	92.587	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	72.400	29.052	33.577	54.836	49.297	60.666	83.445	0.581	0.008	0.494	57.973
$LST^{10}_{apr-oct}$	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	35.201	77.442	47.257	59.042	73.931	33.029	52.210	71.299	60.264	92.587	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	72.396	29.052	33.576	54.842	49.295	60.666	83.445	0.581	0.008	0.494	57.972
$LDD^{0}_{apr-oct}$	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	32.871	75.769	48.860	58.257	71.791	28.202	49.904	69.132	61.417	92.612	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	68.001	30.387	31.421	52.293	49.010	56.905	80.843	0.562	0.007	0.472	56.334
$LDD^{10}_{apr-oct}$	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	34.229	75.863	50.306	58.048	71.452	25.169	49.832	68.868	60.133	92.239	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	66.154	29.332	31.347	51.414	50.253	53.301	81.398	0.560	0.008	0.470	55.843

Table 6.9: Percentages of well-classified pixels, per class, for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information with sum of temperatures (LST: April-October, 0°C and 10°C) and degree-days (LDD: April-October, 0°C and 10°C), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results for each approach correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Results in Tab.6.9 show an improvement with respect to all previous approaches, including the previous linear temporal sampling approach score of 54.847% and ranging between 57.973% and 55.843% for the different tested approaches. The sum of temperatures approaches (LST) obtain better results than the degree-day approaches (LDD), reaching 57.973% ($LST^{0}_{apr-oct}$). Differences between the two tested lower temperature thresholds are mostly lost in rounding, and again more observable in the degree-day case, with better results for the 0°C baseline (56.334%) in comparison to the 10°C threshold (55.843%). Best results are attained with the sum of temperatures method (LST), with an average class accuracy of 57.973% and improving the classification score of 16 classes out of 17.

Tab.6.10 summarizes attained classification accuracies for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information and extending the previous biofix concept to four equidistant points throughout the crop year, that is to say, the first day of October, January, April and July. The sum of temperatures approach is jointly applied with the commonly used linear approach for improved results (LST), and the developmental lower threshold is set to 0°C (LST^{0}_{ojaj}). Classification results are improved, but also the approach requires the use of more features for characterization (12 for the October

biofix, 9 for January, 6 for April and 3 for July, as only one crop year was considered for evaluation, and granting the same number of virtual datepoints per period covered until the end of the year), and consequently an increase in computation times for learning and classification. When targeting classes other than crops subject to rotation (substitution of one crop type by another), time series covering greater time spans could be used, allowing the application of the method over a wider time window for each biofix point. A summary of processing times is given at the end of this section.

LST^0_{ojaj}	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	36.350	77.836	48.067	58.829	74.364	34.407	52.634	71.663	60.250	92.686	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	73.276	30.308	33.877	54.845	50.729	61.722	86.407	0.586	0.008	0.500	58.721

Table 6.10: Percentages of well-classified pixels, per class, for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on NDVI and topographic information with sum of temperatures (LST: January-April-July-October, 0°C), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

Results in Tab.6.10 show again an improvement with respect to all previous approaches with an attained average class classification accuracy of 58.721% (LST^0_{ojaj}). In relation to the reference linear temporal sampling approach score of 54.847%, averages for 20 independent classifications indicate not only a general average class accuracy improvement, but also on all land cover types with the exception of "impervious surfaces and barren land", i.e., on all land cover types affected by temperature.

Tab.6.11 incorporates the previous findings to the proposed method in Sec.3.3. It presents a combined approach applying linear temporal sampling and non-linear temperature-based sampling based and applying sum of temperatures (LST) for NDVI values (nd) for four biofix dates (January, April, July and October) with a lower developmental threshold of 0°C. In other words, spectral indices other than NDVI described in Sec.3.3 are applied in combination with the approach presented in Tab.6.10. Results indicate a significant improvement of the classification, reaching an average class accuracy of 71.478% ($LST^{0,nd}_{ojaj}$ with nd refering to NDVI). It must be noted that improvements reported in Tab.6.11 are not justified by the use of temperature-based sampling alone, but to the application of the methodology as proposed in Sec.3.3. The improvement due to the use of temperature alone can be observed when comparing with results provided in Tab. 4.6, LLI (local linear interpolation), presenting also average results for 20 independent tests on tile L930463 with the proposed approach, and not applying temperature-based sampling. An improvement of results is observable for all statistics: kappa (from 0.609 to 0.622), average accuracy (from 0.682 to 0.693), and average accuracy per class (from 69.92 to 71.478). Compared to a classification based on NDVI alone and linear temporal sampling, a great improvement is attained, with all classes showing improvements.

Finally, Tab.6.12 extends the discovered potential of temperature-guided sampling to other indices in the proposed method in Sec.3.3, that is to say, sampling in the temperature axis is also computed for indices other than NDVI. It presents a combined approach applying linear temporal sampling and non-linear temperature-based sampling based and applying sum of temperatures

$LST^{0,nd}_{ojaj}$	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	49.720	83.218	64.694	77.104	83.813	50.330	62.262	82.737	86.982	98.955	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	83.141	48.811	41.372	62.81	69.996	73.260	95.917	0.693	0.008	0.622	71.478

Table 6.11: Percentages of well-classified pixels, per class, for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on the approach given in Sec.3.3 and applying sum of temperatures for NDVI values (nd) (LST: January-April-July-October, 0°C), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

($LST^{0,all}_{ojaj}$) for all extracted features for four biofix dates (January, April, July and October) with a lower developmental threshold of 0°C. Results indicate a significant improvement of the classification, reaching an average class accuracy of 73.233% in comparison to the application of linear sampling on NDVI values 54.847%. $LST^{0,all}_{ojaj}$ provides thus the best attained results in this chapter, showing the impact of the application of non-linear sampling using the temperature dimension for characterization of land cover classes, and that of the quality of a processing chain design. When comparing with results provided in Tab.4.6, LLI (local linear interpolation), presenting also average results for 20 independent tests on tile L930463 with the proposed approach and not applying temperature-based sampling, an improvement of results is observable for all statistics: kappa (from 0.609 to 0.632), average accuracy (from 0.682 to 0.703), and average accuracy per class (from 69.92 to 73.233).

$LST^{0,all}_{ojaj}$	WHE	COR	BAR	RAP	SFL	PHC	TMF	VIN	WAT	IMP	
	52.814	82.461	70.509	79.616	84.346	51.839	62.598	84.078	88.09	99.055	
	HMO	OWC	OSC	DBL	PNL	MFO	RMO	av	stdev	κ	av.class
	83.800	52.173	43.955	64.148	71.824	77.966	95.685	0.703	0.006	0.632	73.233

Table 6.12: Percentages of well-classified pixels, per class, for a combined approach applying linear temporal sampling with non-linear temperature-based sampling based on the approach given in Sec.3.3 and applying sum of temperatures for all extracted indices (all) (LST: January-April-July-October, 0°C), with temporal resampling with virtual timepoints and a non-refined reference dataset (tile L930463). Results correspond to average results for 20 independent classifications of 3665x3665 pixels. Nomenclature with 17 classes. See Tab.2.10.

The approach presented in Tab.6.11 has been applied to obtain 20 classifications on an expanded study region comprising tiles L930363, L930463, L930563 and L93663, corresponding to areas for which temperature information was available. The same dataset and test region was applied in the previous Chap.5, in order to assure comparability between results for the two chapters.

The applied temperature approach corresponds to the $LST^{0,nd}_{ojaj}$ method presented in Tab.6.11, applying linear temporal sampling and non-linear temperature-based sampling with sum of temperatures (LST) for NDVI values and four biofix dates (January, April, July and October) with a lower developmental threshold of 0°C. The approach is not the one obtaining best classification results over the first analysed dataset (71.478% in comparison to 73.233% for the best results), but it has been chosen instead because it offers a good compromise between classification improvements and computation times due to the use of a reduced number of features in relation to the method obtaining best accuracies, the $LST^{0,all}_{ojaj}$ method presented in Tab.6.12. A summary of computation

times is offered at the end of this section. In order to permit comparison, also reference results for the same area without the application of temperature-guided temporal sampling are provided. Results for both sets of tests, with linear temporal sampling (LS) and with non-linear temperature-guided temporal sampling (LST) are given in Tab.6.13.

LS	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	
	50.159	81.652	61.179	77.199	78.329	86.436	37.09	50.416	
	VIN	WAT	IMP	HMO	SMO	OWC	OSC	FNO	DBL
	69.223	81.607	92.224	71.84	62.124	52.221	47.569	57.252	45.47
	PBL	DNL	PNL	MFO	RMO	av	stdev	κ	av.class
	60.731	69.793	59.59	52.091	52.091	0.557	0.006	0.497	63.467
$LST^{0,nd}_{ojaj}$	WHE	COR	BAR	RAP	SFL	WOO	PHC	TMF	
	51.103	81.908	61.549	77.071	78.226	86.194	40.212	51.899	
	VIN	WAT	IMP	HMO	SMO	OWC	OSC	FNO	DBL
	71.172	81.906	92.187	74.213	65.889	52.884	48.3	62.365	45.67
	PBL	DNL	PNL	MFO	RMO	av	stdev	κ	av.class
	62.91	75.225	60.193	58.422	53.72	0.564	0.004	0.504	65.146

Table 6.13: Percentages of well-classified pixels, per class, for a linear temporal sampling approach (LS), and for the proposed temporal sampling approach applying applying sum of temperatures for NDVI values (LST: January-April-July-October, 0°C) (LST^{0}_{ojaj}) , with temporal resampling with virtual timepoints and a non-refined reference dataset for the study area (tiles L930363, L930463, L930563, L930663). Results for each approach correspond average results for 20 independent classifications of a region of 13667x3665 pixels. Nomenclature with 22 classes. See Tab.2.10.

Tab.6.13 shows comparable results to those obtained on the previous study region, with an increase in classification performance (average class accuracy and kappa) for the LST^{0}_{ojaj} approach (65.146% in comparison to the 63.467% of the linear temporal sampling approach). This demonstrates further the advantages of introducing temperature-based temporal sampling for characterization, as it makes it clearer how classification can be enhanced on a larger area. Average accuracies need to be interpreted here in relation to previous tests having into account that although the total sampling ratio has been kept the same (under 0.07%) by applying 500 per class (17 classes) on a 3665x3665 area to 1500 (22 classes) on a 13667x3665 area, the application of a superior number of classes has a necessary impact on results. Consequently, percentages are lower, but the relative improvement is anyway made clear. Most significantly, and coherently with results presented in Chap.5, Tab.6.13 demonstrates how acceptable classification accuracies can be obtained by sampling with a significantly small percentage of pixels on a very large study area.

Thus, the contribution of non-linear sampling based on temperature has been demonstrated on both tested datasets. As explained before, these results should not be considered as the best attainable results but, more in the opposite sense, as an indicator of the potential of the approach, due to the many constraints that have conditioned the testing. Firstly, the number of classes that have been applied is high, which determines necessarily lower accuracy classification results than with smaller numbers of classes. Many approaches in the literature are based on only a few classes, and thus accuracy results are not comparable. Secondly, the number of training samples is extremely low (under 0.07%). In this respect, approaches in the literature tend to use much higher ratios, of the order of 20 to 30%. With higher numbers of training pixels, higher accuracies can be obtained, as demonstrated in Chap.4, although with longer computation times. Also, in order to find a good compromise between computation times and accuracy, choices have included

the use of temporal resampling with with 12 points (described in Chap.4) and the application of temperature-based sampling with four biofix dates only for NDVI values. When the yielding of higher accuracies is a main target, then a higher number of virtual points can be used, and the temperature method can be extended to other selected primitives. Finally, the choice of a linear SVM kernel without optimisation determines as well faster computation, with the penalty of lower accuracies; higher accuracies can be obtained with for example an RBF kernel and optimisation. With the characteristics of the Sentinel-2 series, higher resolutions and the corresponding decrease of spatial mixing may probably translate, as well, into higher classification accuracies with the proposed method.

Finally, Tab.6.14 gives an overview of required computation times, computed support vectors, and classification times for the most relevant tested approaches, with values averaged over 5 classifications per approach. Processing times with SVM classifiers depend on number of features, as described before, such as data size and dimensionality, chosen kernel, SVM parameters or the use of optimisation. For each of the summarized approaches, the number of training samples remains constant (500 per class, i.e, 8500 in total for the test area), and the classification area corresponds to Landsat tile L930463, with size 3665x3665 pixels. A linear kernel without optimisation has been applied.

Approach	Dimensionality	Learning (s)	#sv	Classification (s)
TS	15	10.718	6818.8	11m58.113s
LST^{0}_{ojaj}	45	25.19	6500.8	25m25.359s
$LST^{0,nd}_{ojaj}$	129	46.052	5160.2	49m00.961s
$LST^{0,all}_{ojaj}$	339	118.6	4916.0	2h47m26.320s

Table 6.14: Average SVM learning and classification times for the most relevant tested approaches with an indication of data dimensionality and computed support vectors.

Tab.6.14 shows a constant decrease in the number of support vectors for the tested approaches, from the TS approach with an average of 6818.8 to the $LST^{0,all}_{ojaj}$ approach with 4916 svms. Thus, the number of support vectors decreases with the number of primitives, which indicates that the primitives that have been added are pertinent and allow for the generation of a simpler model. Also, as shown in Tab.6.14, computation time increases with data dimensionality. This is true also with the use of non-linear kernels, a greater cost parameter, greater data sizes or the use of optimisation. The improvement brought by the presented approach has been demonstrated; the trade-off between computation times and accuracy scores remains a choice depending on the specific necessities of the target land cover map users. The user needs to define their choices depending on a variety of factors, such as purpose, targeted nomenclature, available data, and spatial, spectral and temporal characteristics, as described in Chap.3. In any case, the method remains operational and fully automatic, and intends to be useful not only for remote sensing specialists, but also for other users requiring land cover maps over large areas as input for their research areas. Although designed for the Sentinel-2 satellite constellation soon available time series, it is generalisable to other nomenclatures and input data. Some related research areas remain unexplored, such as the application of the approach for classification accross years; these are further developed in Chap.7.

6.5 Conclusions

The topic of linearity of the temporal axis when dealing with vegetation parameter extraction for land cover mapping over large areas has been analysed, and a brief review on the topic of phenological processes and their application in remote sensing literature has been provided.

A novel non-linear temperature-driven sampling approach aiming at the extraction of robust features for the enhanced characterisation of targeted classes and the subsequent application of machine learning and classification algorithms has been presented. It targets a more robust characterisation of land cover lasses by providing a new perspective of analysis for the extraction of characterisation features. The method tries to reduce the variability of spectral responses in the time series due to climatological environments by applying an axis transposition along cumulative temperature curves. In this way, the method tries to minimise the difficulties posed by the problem of land cover types variability that very often hinder map production over large areas.

The methodology complements the methods presented in the previous chapters, based on the application of local interpolation techniques in the time series for the processing of missing information, the use of temporal sampling with virtual points, and the application of spatial sampling techniques, for the management of missing information, differing acquisition dates, data variability, and large volumes of data. In this way, exploitation is based on both the previously presented spatial and temporal dimensions, with a further virtual dimension determined by temperature. The method is thus part of the general framework presented in this thesis work for the automatic generation of land cover maps with operational times over large areas.

Different temperature-guided sampling algorithms have been implemented and a combined approach including both linear and non-linear sampling techniques has been proposed. The novelty of the approach resides not only in the introduction of a virtual dimension determined by temperature for non-linear sampling aiming at the enhanced characterization of targeted classes, but also on the new application of combined biofix dates for the same aim, thus integrating data processing algorithms and land cover background knowledge on ecological and agricultural issues such as vegetation growth impact variables and practices. Results have been analysed in terms of impact on classification results. These have been generated over two datasets, one comprising a large area of $12100km^2$ (13432225 pixels), and the other one comprising an extended area of $45100km^2$ (50089555 pixels). The potentiality of non-linear temperature-based sampling has in this way been demonstrated for features computation for classification and, in particular, its applicability for the operational generation of land cover maps over large areas.

Chapter 7

Conclusions

This document presents the design and validation of a fully automatic approach that allows the estimation of large-area land cover with high spatial, spectral and temporal resolution Earth observation optical imagery, being applicable with different nomenclatures and satellite time series, and offering operational computation times (with delivery in a matter of hours), in preparation of the coming Sentinel-2 and Venμs missions. The developed system demonstrates the feasibility of fast automatic production of large scale land cover maps, dealing with data volume and variability, limited reference data and missing information, and integrating different data processing approaches.

7.1 Contributions of this thesis

The presented system constitutes an automatic image processing methodology based on the following broad steps: temporal series missing information processing, resampling in the temporal axis, feature vector construction, sampling in the spatial axis, support vector machine model learning, and classification. In practical terms, it allows for the fast and automatic generation of land cover maps at large scales using voluminous high-resolution multispectral time series. It offers a solution to the problems posed by large-scale land cover production with the last satellite generation: coverage by different time series with non-coincident acquisition dates (due to imaging by different satellite passes), extensive presence of missing information due to a variety of reasons (clouds, cloud shadows, passage borders, snow), high processing costs for large volumes of information (time series of high resolution multispectral data), high spectral variability of classes in relation to spatial diversity, and the existence of a limited reference dataset.

In particular, five main contributions have been developed in the chapters of this thesis:

(1) Firstly, the framework for automatic large-area land cover generation with large-volume time series such as those to be provided by the coming satellite mission Sentinel- 2. It gives an answer to the problem of automation in the considered frame. It consists of a pixel-based approach founding the feature vector construction on virtual datepoints after temporal resampling on the input time series and it allows for the production of results in operational times and the managing of great volumes of information. Based on the exploitation of multitemporal and multispectral information, it uses a combination of targeted sample extraction, missing information processing,

characterizing feature extraction, sampling, supervised model learning, and ultimately classification based on the learnt model. This framework allows for the combined application of the novel contributions presented in this thesis in relation to classical classification approaches and for the targeted automation not requiring operator supervision.

(2) Secondly, the comparison between missing information processing approaches. It gives an answer to the problem of missing information in the considered frame. A multi-temporal method to produce temporally and spatially continuous high resolution time series for large-scale processing is presented, based on the application of local interpolation techniques on the temporal series. Different approaches to the proposed method are implemented and compared, and their adequacy is proven, particularly when it is possible to obtain successive analytical expressions in radius-varying windows for the complete time series.

(3) Thirdly, the temporal sampling approach. It gives an answer to the problem of unmatched acquisition datapoints and temporal irregularities with datasets covering large areas in the considered frame. The method takes advantage of the previous missing information processing approach in oder to generate virtual data points for the construction of the feature vector. It allows for large-area learning and classification with one learned model, substituting the need for different learners and classifiers, and avoiding other solutions such as manual area selection (too costly in time terms) and low content image rejection (loss of valid information).

(4) In the fourth place, the stratified spatial sampling approach. It gives an answer to the problem of great data variability in spectral response when large areas are considered in the given frame. It is based on the application of automatic spatial sampling strategies for the production of representative learning sets, thus allowing faster machine learning for subsequent classification. These strategies complement the temporal sampling strategies described above, offering a methodology that operates in this way in the spatial space for dataset selection, and in the temporal space for missing data processing and regularization. Different approaches for spatial sampling are compared for the chosen study area, and in particular stratification is examined by incorporation of environmental science concepts to improve machine learning performance: topographic and ecoclimatic information are applied for the identification of useful stratification variables.

(5) In the fifth place, the non-linear temporal sampling approach based on temperature. It gives an answer to the limitation of linearity when dealing with vegetation parameter extraction. It builds on the previous approaches with the application of interpolation techniques (for the processing of missing information), temporal resampling with virtual points (for the management of irregular acquisition dates and large volumes of data), and stratified spatial sampling (for the handling of data variability and processing costs). Non-linear temperature-based temporal sampling is proposed for vegetation classes, with the hypothesis that phenological evolution and temporal spectral response of vegetation land cover is dependent on temperature rather than time. A method for the analysis of temporal spectral response in the temperature space is given, and different approaches are developed and tested. The novelty of the approach resides not only in the introduction of a virtual dimension determined by temperature for non-linear sampling aiming at the enhanced characterization of targeted classes, but also on the new application of

combined biofix dates for the same aim, thus integrating data processing algorithms and land cover background knowledge on ecological and agricultural issues such as vegetation growth impact variables and practices. Also, correction methods for the refinement of input temperature information are implemented and presented.

The system has been developed as a processing methodology for the coming Sentinel-2 datasets, but establishes a general framework applicable to other satellite data or to multisensor datasets duly corregistered and resampled. The method is flexible to work with any type of raster satellite time series and it is their availability, quality and characteristics (in terms of spatial, temporal and spectral features) that determine feature extraction feasibilities and reachable levels of detail. Also, different target informational classes can be applied, depending on land cover map purposes and available time series, as any number of classes and types of land cover information can be accommodated.

The system is reproducible and robust, as it can be easily replicated and it can cope with errors and input data artifacts. Indeed, reproducibility is linked to the minimisation of the analyst role in the classification, as operator input throughout a procedure results in non-reproducibility [Cihlar, 2000]. It has low requirements in terms of training data, memory needs and processing times. It is totally automated and does not require user interaction or manual work. It restrains the operator input to initial decisions subordinated to available data and desired output, such as for example the targeted land cover types, type of features to be used, or decisions on the parameters of the processes described above, such as number of timepoints used for temporal resampling with virtual points, number of training samples to be used, or study temporal period. In this way, the reproductibility and consistency of the method can be assured, and the impact of operator-related subjectivity and associated costs are minimised.

In conclusion, this thesis tries to provide the remote sensing community, and users who are not specialists of the field but that rely on land cover mapping outputs as inputs for their own pursuit areas, with a methodology to ease the exploitation of information from the new generation sensors, allowing for the production of large area land cover maps based on multitemporal high resolution time series with operational delivery times, and particularly dealing with real datasets, which generally present temporal irregularities, missing information, and great data variability; the suggested strategies try to open new paths for the processing of the new generation satellite datasets.

7.2 Perspectives for future research

A variety of perspectives for future research have been identified and are developed in the following subsections. Firstly, the application of the proposed methodology with the targeted Sentinel-2 data is discussed. In addition to this, other directions stand out as future lines of research in relation with this work: the use of temporal descriptors to improve the characterisation of vegetation classes, and the detection of ground truth outliers to reduce error introduction in the learning process. These two aspects may complement the presented system for the improvement of classification accuracy in two respects: by introduction of distinctive features in the feature vector, and by reduction of probable error in the reference training data. Also, implementation possibilities and applicability

with and in conjunction with other satellite data is discussed and final remarks are provided.

7.2.1 Application with Sentinel-2 data

This work has been tested on combined Landsat 5 and 7 datasets, used to simulate the soon-available Sentinel-2 twin-satellite imagery. Application of the proposed approach on the targeted Sentinel-2 datasets is perhaps the main line of future research to be identified for a new evaluation of the system. The Sentinel-2 mission with two satellites (Sentinel-2A and Sentinel-2B working on opposite sides of their orbit) is estimated to be operational soon, at the end of 2016, with the launch of Sentinel-2B. As explained, six of the thirteen Sentinel-2 spectral channels are considered as heritage of the Landsat system, and thus the corresponding Landsat channels have been exploited here. The additional ones that Sentinel-2 will provide can be introduced in the future in order to obtain better classifications. In particular, the three bands at 60m resolution (B1, B9 and B10) will be helpful for aerosol, water-vapour and cirrus clouds detection; the narrow-band bands at 20m resolution (B5, B6, B7 and B8a) can aid in vegetation discrimination (particularly the B8a vegetation red-edge); finally, the Landsat-related B2, B3, B4 and B8 will have a higher spatial resolution of 10m, which will surely allow for classification improvements (for example by addition of textural features). Also, this application may reveal new difficulties that have not been considered here and which are worth analysing.

7.2.2 Temporal descriptors

The phenological evolution of vegetation and crops has been established by ecological studies to follow identifiable temporal patterns from emergence and rapid growth, through a period of maximum leaf area, and towards senescence and dormancy [Zhang et al., 2003]. Some of the temporal markers to be identified and studied include: duration of greenness, green-up onset, maturity stability, senescence onset or dormancy onset. These descriptors can be interpreted for a selected primitive against a time- or temperature-based variable, and they can be assessed, for the same species, in order to characterise their behaviour according to variables such as soils, latitude, orientation and altitude of growth.

The identification of phenological and more generally temporal markers when working at large scales is however not a trivial pursuit. Vegetation types can exhibit more then one cycle of growth and senescence, or present complex evolutions due to climate, soil or land management factors [Zhang et al., 2003]. Also, several studies point at the fact that the appearance of phenological markers is displaced in time due to differences in altitude or latitude [Rötzer and Chmielewski, 2001]. Multiple efforts have been conducted in the literature for the identification of markers in multi-temporal satellite data [Jönsson and Eklundh, 2002] [Jönsson and Eklundh, 2004] [Zhang et al., 2003]. These temporal descriptors may be applied to further characterise land cover classes, and their impact on classification results can be analysed. In particular, the application of the temperature-based approach presented in this thesis in conjunction with phenological markers may be of interest.

7.2.3 Ground-truth outlier detection

Another line of work is the automatic statistic outlier detection in the reference data, in order to identify and thus minimise the errors introduced in the processing chain and ultimately in the land-cover map generation by erroneous data in the training and reference ground truth.

The quality of the ground truth information used for training supervised classification approaches greatly determines their performance. In remote sensing, reference data are often constituted by a conglomerate of information from different sources, such as cadastres, forest inventories, or agricultural parcel declarations. They generally present a high variability in terms of update frequencies, information collection methods, spatial extent and resolution, and more importantly accuracy and confidence levels. Thus, the presence of outliers in compiled ground truth for land cover classification constitutes a problem than needs to be addressed in order to minimize their impact in subsequent training and classification stages.

Different approaches exist for outlier mining, that is to say, to find the dataset objects which are importantly inconsistent with the rest of the data. It is a non-trivial task, which consists of not only an outlier mining method, but also of the non-obvious task of inconsistence definition to avoid confusion with non-typical class examples or with cases of phenological response differences due to dissimilar geographic, topographic or climatic factors. In very general terms, approaches can be defined as statistical, distance-based, density-based and consisting of deviation rules. Statistical methods focus on an assumed probability model [Barnett and Lewis, 1994]; distance-based methods are based on the number of neighbours based on a distance criteria; density-based local outlier methods focus on non-uniformly distributed data and local outliers relative to the density of their local neighbourhood; finally, methods based on deviation rules identify outliers based on characteristics, by for example applying a dissimilarity function. Also, different research efforts have been conducted in the field of outlier detection for time-series in different domains [Gupta et al., 2014]. These outlier removal techniques may be applied on land cover reference data, and their impact on classification results can be analysed and compared.

7.2.4 Implementation possibilities

Different aspects of the proposed method have been identified as leads for further research. Firstly, as explained, one of its main characteristics is the fact that it can be generalised to different nomenclatures, satellite data and landscapes of any size. The presented framework can be tuned depending on targeted land cover types and on the characteristics of available satellite data. The selection of pertinent features to be extracted needs to be based both on targeted types and on satellite data spatial, temporal and spectral characteristics. For example, the inclusion of texture features in the given framework may be of interest for class separability enhancement, particularly with higher spatial resolution sensor data, and remains as a future investigation aspect.

As other examples of unexplored implementations, particular indexes related to targeted classes can be applied, such as for example the iron oxide ratio in order to enhance altered rocks subjected to oxidation of iron-bearing sulphides [Segal, 1982], when these are of interest. Also, the derivation and application of the tasseled cap transformation indexes (brightness, greenness, wetness) may be further investigated, particularly in conjunction with non-linear temperature-guided sampling.

In addition to this, the methodology is flexible to use any temporal period and does not need to be restricted to crop years depending on targeted land cover types, that is to say, shorter or longer periods (several years, for example) can be used as input. With a different perspective, temperature-based characterization for classification accross different periods may also be further investigated, that is to say, the method may be interesting for comparisons between for example diverse crop years. In addition to this, the applicability of the method on selected land cover classes can be undertaken such as for example only crop types. These ideas remain as future research aspects.

In order to apply the proposed methodology at the regional level, for example to work on territories of the size of Europe, at least two further aspects might need consideration. On the one hand, the lack of temperature information for all considered areas, which may be addressed by fusion of classification results obtained by traditional linear sampling and by temperature-based sampling depending on availability. On the other hand, the widening of longitude and latitude ranges, which may be addressed by incorporating longitude and latitude gradients as two of the clustering features determining strata for stratified sample selection. Also, other difficulties not contemplated here might arise. These topics may be interesting subjects of investigation.

Finally, other aspects outside the scope of this work include the application of data reduction strategies in conjunction with the proposed methodology in order to accelerate the classification process by decreasing data volume. Some techniques in this sense are the selection of most discriminating bands, or the transformation of data into another feature space, such as the principal component analysis method (PCA). In addition to this, another aspect is the combined application of the suggested pixel-based approach with object-based analysis, which allows the introduction of shape features, such as total area, solidity, major and minor axes and orientation, to name a few. An example of integration of pixel and object-based analysis is to be found in [Chen et al., 2015], adding to the two approaches a further knowledge-based verification step.

7.2.5 Other satellite sensors

The proposed methodology offers flexibility for the integration of other satellite sensor data in the optical and other ranges. Indeed, due to the nature of the approach, multisensor information can easily be integrated provided prior resampling and corregistration. In particular, the combined application of microwave and optical sensors may enhance the discrimination of classes, due to their natural complementarity. Indeed, the advantages of each sensor can be exploited in this way, as they cover different spectral ranges and provide thus in combination a more complete description of features to improve classification error [Hegarat-Mascle et al., 2000]. Studies have already been undertaken based on the combination of optical and microwave imaging, particularly in agricultural settings [Hong et al., 2014], taking advantage of the structural information provided by SAR in combination with the reflectance information of optical data. For example, synthetic aperture radar (SAR) and optical imagery are integrated to differentiate grassland and alfalfa in the Prairie Provinces of Canada with significant accuracy improvement in comparison to alternative approaches in [Hong et al., 2014]. The proposed framework allows for a straight-forward combined application of microwave and optical imaging (or other types of sensor data); analysis of the impact of such a strategy for land cover mapping on large areas may be of interest.

7.3 Final remarks

The importance of fast and accurate mapping and monitoring of land cover becomes clear when considering the ever-growing strain on Earth resources: anthropic pressure, climatic warming and global change have an undeniable effect on natural processes; climate changes and their effects on ecosystems are unprecedented over previous decades and are expected to intensify in the future; deforestation, desertification and soil erosion are threatening food security. In these circumstances, research on land cover mapping with satellite remote sensing imagery of the Earth becomes essential, allowing for the extraction of key information for the mapping of past and current ever-changing resources and thus the understanding of the Earth's processes, in order to forecast future scenarios, in this way contributing to the reduction of disaster losses and to sustainable development.

In this context, this work is intended as a contribution to rapid large-area-capable land cover mapping with remote sensing imagery. It constitutes a methodology for the estimation of large-area land cover with high resolution satellite imagery and operational computation times (delivery in a matter of hours). It is specifically conceived to manage large volume datasets of hundreds of millions of pixels (such as those to be provided by the new satellite missions). It has low processing and memory costs (it can be implemented in any standard current-day PC) and it is fully automatic, not applying operator supervision or interaction in order to eliminate subjectivity. It admits variability of land cover classes over space, and manages missing information. Also, one of its main characteristics is generality: it can be tuned according to needs, and generalised to different nomenclatures (depending on the purpose of the land cover map), satellite data time series (taking into account their different spatial, temporal and spectral characteristics) and landscapes of any size. As explained, this work has been intended as a contribution not only to the domain of land cover mapping, but also as a tool for users of other knowledge domains that rely on land cover mapping outputs as an input for their own pursuits.

Chapter 8

Publications and other communications

8.1 International peer-reviewed journal papers

- J.Inglada, A.Vincent, M.Arias, B.Tardy, D.Morin, and **I.Rodes**,"Operational high resolution land cover map production at the country scale using satellite image time series," *Remote Sensing*, 9(1), 95, 2017.

8.2 International conferences

- **I.Rodes**, J.Inglada, O.Hagolle, J.F.Dejoux, and G.Dedieu, "Non-linear time sampling driven by surface temperature for the monitoring of vegetated areas using multi- and hyper-temporal satellite image time series," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Melbourne, Australia, 2013.

- **I.Rodes**, J.Inglada, O.Hagolle, J.F.Dejoux, and G.Dedieu, "Sampling strategies for unsupervised classification of multitemporal high resolution optical images over very large areas," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, Muenchen, Germany, 2012.

8.3 Other communications

- J.Inglada, **I.Rodes**, M.Arias, M.Kadiri, and O.Hagolle, "Assessment of interpolation and approximation algorithms for high temporal and spatial resolution image time series" in *4th International Symposium on Recent Advances in Quantitative Remote Sensing: RAQRS'IV*, Torrent, Spain, 22-26 September 2014.

- S.Valero, J.Inglada, O.Hagolle, **I.Rodes**, J.Osman, J-F.Dejoux, C.Marais-Sicre, and G.Dedieu, "Challenging production of land cover maps for operational agri-environmental applications," in *5th Workshop of the EARSeL Special Interest Group on Land Use and Land Cover: Frontiers in Earth Observation for Land System Science*, Berlin, Germany, 2014.

SAMPLING STRATEGIES FOR UNSUPERVISED CLASSIFICATION OF MULTITEMPORAL HIGH RESOLUTION OPTICAL IMAGES OVER VERY LARGE AREAS

I. Rodes, J. Inglada, O. Hagolle, J.F. Dejoux, G. Dedieu

CESBIO - UMR 5126
18 avenue Edouard Belin
31401 TOULOUSE CEDEX 9 - France

ABSTRACT

Efficient unsupervised production of large-area land cover maps with the volumes of data to be generated by the forthcoming Earth observation missions is challenging in terms of computation costs and data variability. As a solution, introduction of non-spectral knowledge for data reduction and selection is proposed here. Analysis of intra-strata variability and inter-strata correlation for different stratified sampling approaches is presented, and valuable variables for both stratification and classification are identified.

1. INTRODUCTION

The forthcoming Earth observation missions Sentinel-2 (prepared by the European Space Agency, ESA, in the frame of the GMES program) and Venμs, jointly developed by the French Space Agency (Centre National d'Études Spatiales, CNES) and the Israeli Space Agency (ISA), will soon revolutionize present-day environmental monitoring with the yielding of unseen volumes of data in terms of spectral richness, temporal revisit and spatial resolution. Venμs will deliver images in 12 spectral bands from 412 to 910 nm, a repetitivity of 2 days, and a spatial resolution of 10 m; the Sentinel-2 satellites, will provide coverage in 13 spectral bands from 443 to 2200 nn, with a repetitivity of 5 days, and spatial resolutions of 10 to 60m [1].

The efficient production of land cover maps based on the exploitation of such volumes of information for large areas is challenging both in terms of processing costs and data variability. In general, conventional methods make use of supervised approaches, target specialised local models for determined problem areas, or include complex physical models. These approaches, expensive in terms of processing time, are thus inefficient for the exploitation of the new type of data that Sentinel-2 and Venμs will provide. A need arises for the implementation of accurate and fast unsupervised methods that allow for generalisation to large scale areas with high temporal and spectral resolutions.

This work was supported by CNES under the TOSCA program.

2. OBJECTIVE

The objective of achieving unsupervised classification of multitemporal high resolution optical images over very large areas and with near-real time processing poses two main challenges: the minimisation of computation time costs when dealing with high volumes of data, and the treatment of the considerable spectral variability of thematic classes across regions. In the quest for such strategies, a need for sampling approaches is then made obvious [2], the main issues being data reduction, for the managing of time costs, and data selection, for the identification of surfaces in their whole range of spectral behaviours. Among the different sampling strategies, stratified sampling offers the possibility of implementing both data reduction and data selection by means of incorporating additional non-spectral knowledge about the landscape. This work presents data analysis of intra-class variability and inter-class correlation scores for different stratified sampling approaches, with an aim to identify valuable variables for both stratification and classification.

3. METHODOLOGY

3.1. Study area, validation data and dataset preparation

The study area is that of the Pyrénées, a mountain range constituting a natural border between France and Spain. Work has been based on Landsat 5 and Landsat 7 data, provided by USGS and ESA, and treated at CESBIO. These Landsat scenes cover the Pyrénées area during 2010 with a spatial resolution of 30 meters, processed in tiles of 110x110km. Due to their similarities, Landsat (two satellites Landsat 5 and Landsat 7, 30 m resolution, 180 km field of view, 7 bands including SWIR, and a repetitivity of 8 days) is often used as a simulator for Sentinel-2. A processing chain has been devised using the CNES Orfeo Toolbox Library for remote sensing image processing [3] for the elimination of cloud areas from the Landsat time series, based on [1]. A combination of linear and Savitzky-Golay interpolation [4] has been applied, using a weighted version of the smoothing filter, performing local polynomial regression on series of values, for the preparation

of the dataset. Validation data, as shown in Table 1, have been generated based on the Silvapyr hierarchy [5] and the IGN and RPG databases [6].

3.2. Stratified sampling

Sampling consists of the selection of a subset of a population allowing for deriving estimates about characteristics of the whole population. Random, periodic, or stratified sampling approaches generally permit computation cost reduction. The latter, also, proves to be adequate for the treatment of population subgroups presenting different responses of interest for a single thematic label, being advised when attributes are different for different subpopulations [7], as it is the case here. Stratified sampling uses additional information for the partition of the population into M strata, and applies sampling within each strata. The use of prior information for stratification can considerably improve the estimates of population parameters. For precision to increase in relation to other methods, minimisation of intra-stratum variances and maximisation of inter-stratum correlation is needed.

Prior information applied for stratification is derived from two different sources: pedologic and topographic data. The former introduces soil classification, as the different characteristics of soil regions (origin and evolution, composition, degradation) affect the development of their vegetation. The latter refers to relief features such as altitude, slope and aspect. These features also greatly influence the present vegetation, giving way to distinct evolution patterns for the same species. Indeed, a phenological shift is observable for different location altitudes, as well as particular species. Also, slope and aspect determine temperature, wind and irrigation conditions, also affecting the development of vegetation, and thus spectral response. Two pedologic classifications have been used, using two levels of detail in the nomenclature. Clusterings of 5 and 15 cluster centres have been generated with the variables mentioned above, both isolated and after fusion, with the results reported below.

4. EVALUATION

4.1. Data variability

In order to analyse the characteristics determined by different stratification variables, population variance of each strata and corresponding substrata of size N has been calculated as $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$, with population mean being $\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$, using both multitemporal and multispectral dimensions. Table 2 shows, for each stratification variable or combination of variables, and the applied clustering centroid number, minimum, maximum and average variances, and thus the average increment of variance of the stratification strategy in comparison to sampling without stratification. All calculations have been performed on multidimensional data based on NDVI values corresponding to Landsat series data. For comparison purposes, average region values are plotted for each of the thematic classes in Fig.1. The difficulty that spectral identification entails is made clear when the associated variances for these regions, without stratification, are shown in Fig.2.

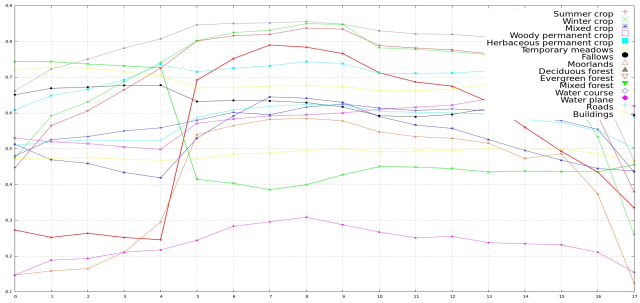


Fig. 1. Landsat time series NDVI average values per thematic class. Pyrénées region (2010, 110x30 km, 18 dates)

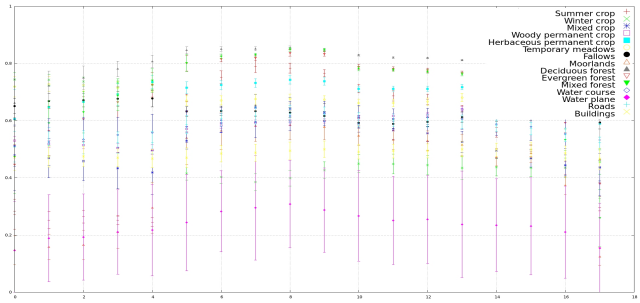


Fig. 2. Overlapping NDVI variance values per thematic class corresponding to Fig.1

4.2. Data correlation

Pearson's product-moment correlation coefficient has been applied to measure correlation between strata corresponding to each of the stratification strategies listed above. A measure of the correlation between two variables within the range of [-1,1] is obtained; geometrically, correlations equal to the extremes correspond to points in a perfect linear relationship. The sample coefficient is defined as $r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i-\bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i-\bar{Y})2}}$.

Calculations have been performed on multidimensional data (multispectral and multitemporal) based on NDVI values. A non-centred version of the sample correlation coefficient has been applied for all couple units $(p_i, p_j) = ((x_i, y_i), (x_j, y_j))$, which can be defined according to their belonging to the reference ground truth regions R_i with $i = 1, 2...I$ or to the stratification strata S_j with $j = 1, 2...J$, which in turn can be contained within the validation data region.

1. Seasonal crops	2. Permanent crops	4.Forests	5. Water areas	7. Temporary meadows
1.1. Summer crop	2.1. Woody permanent crop	4.1. Deciduous forest	5.1. Water plane	8. Moorlands
1.2. Winter crop	2.2. Herbaceous permanent crop	4.2. Evergreen forest	5.2. Water course	9. Fallows
1.3. Mixed crop	3. Roads	4.3. Mixed forest	6. Buildings	

Table 1. Nomenclature for validation

A5	H5	P1	P2	S5	HA5	HAS5	HS5	GT	A15	H15	S15	HA15	HAS15	HS15
0,030	**0,027**	**0,025**	**0,025**	0,032	0,030	0,030	0,030	**0,030**	0,032	**0,025**	0,035	**0,027**	**0,027**	**0,026**
0,012	0,013	0,016	0,014	0,016	0,015	0,014	0,014	**0,012**	0,014	0,012	0,018	0,014	0,013	0,013
0,009	0,002	0,004	0,004	0,000	0,010	0,009	0,007	**0,010**	0,007	0,001	0,002	0,000	0,000	0,001
0,059	0,059	0,085	0,085	0,079	0,068	0,063	0,080	**0,055**	0,104	0,063	0,101	0,068	0,071	0,063

Table 2. Variance statistics (columnwise): average, standard deviation, minimum and maximum. Stratification variables: A-aspect, H-height, P-pedology (1-2), S-slope, GT-ground truth. Number of clusters for (fused) variables: 5 or 15, as indicated

Tables 2 and 3 summarize, for one of the studied regions (110x30 km), the obtained statistics.

5. EVALUATION

Analysis of numerical results indicates that a general reduction of variance in absolute values can be achieved by stratification, with an inter-strata correlation increment only when pedologic maps or slope are used. Increasing clustering centroid number seems to have a positive effect on height and height-fused stratification variables. In particular, the most significant results in terms of variance correspond to pedologic variables and height-derived information. In relation to inter-strata correlation, increases are to be found with slope and soil information. Thus, soil variables $P1$ and $P2$ emerge as appropriate stratification methods for the present purposes. This is coherent with the observation of distinct vegetation evolution patterns in different soil regions.

The difficulty of analysis of this particular type of data is illustrated by Table 4. Correlation values have been calculated over validation data using a sampling ratio of 1/3, inter-strata correlation being expected to be higher than correlation with other strata. This is not always the case: while deciduous forest presents its highest correlation against itself, evergreen forest has a higher correlation with woody permanent crops, and mixed forest reaches its maximums against the former two. Many factors may have an impact here, from the sampling for correlation calculation itself, to inaccuracies in the ground truth, and spectral mixing.

Figs.3 and 4 show the field graph of thematic classes in the elevation/orientation and elevation/slope spaces. These plots are coherent with the undertaken variance calculations of substrata, summarised in Tables 2 and 3. While selection of stratification variables presents difficulties, identification of appropriate classification variables is made possible, as it is observable in the data that different classes are only present in particular elevation, slope and orientation ranges. For example, the distribution pattern of the three considered forest types can be observed, as shown in Figs.3 and 4.

6. CONCLUSIONS AND FUTURE WORK

Characteristics of strata determined by different stratification variables have been analysed in terms of population variance and inter-strata correlation, with the objective of discovering variables improving precision in relation to ground truth values as illustrated in Figs.1 and 2. In this way, sampling is not particular for a given dataset, but admits generalisation. Pedologic variables have proven to be useful stratification variables, and processing costs and data variability have been addressed. Also, not only stratification but classification variables have been identified. The interest of using prior knowledge on terrain and pedologic characteristics to improve sampling for further classification purposes has been shown. This work contributes to the quest for strategies which are able to deal with great volumes of data of high temporal and spectral resolutions and large areas, as required by the forthcoming new Earth observation missions.

Sampling precision in cases where strata variances are minimised and between-strata correlations are maximised is likely to increase in relation to other sampling methods such as random or periodic. With the former, an irregular spatial distribution of sample locations can result in a non-uniform representation of the whole population; with the latter, a periodic arrangement of the population can produce inefficient samples. In future work the adequacy of stratified sampling will be thus tested against these alternative methods in two manners. Firstly, supervised classification with SVMs will be trained on the samples, and the accuracy and kappa index values for evaluation of these classifications will be used as an indication of how well the sampling performed. The analysis of strata here undertaken constitutes a necessary pre-

A5	H5	P1	P2	S5	HA5	HAS5	HS5	GT	A15	H15	S15	HA15	HAS15	HS15
0,904	0,900	**0,914**	**0,913**	**0,907**	0,905	0,903	0,895	**0,905**	0,903	0,899	0,905	0,900	0,898	0,896
0,058	0,061	0,056	0,056	0,055	0,056	0,061	0,063	**0,055**	0,060	0,065	0,058	0,066	0,071	0,073
0,625	0,647	0,521	0,521	0,667	0,641	0,591	0,619	**0,683**	0,598	0,577	0,589	0,480	0,379	0,440
0,971	0,989	0,987	0,987	0,983	0,978	0,980	0,984	**0,984**	0,979	0,985	0,983	0,988	0,987	0,987

Table 3. Summary of correlation statistics per stratification strategy. Nomenclature as in Table 2

	SC	WC	MC	WPC	HPC	M	F	Mo	DF	EF	MF	R	B
Deciduous forests	0.903	0.886	0.899	0.961	0.957	0.956	0.956	0.847	**0.965**	0.951	0.946	0.953	0.947
Evergreen forests	0.908	0.852	0.885	**0.945**	0.940	0.933	0.935	0.855	**0.951**	**0.942**	0.938	0.935	0.928
Mixed forests	0.905	0.843	0.880	0.933	0.933	0.923	0.925	0.855	**0.946**	**0.938**	**0.934**	0.924	0.916

Table 4. Correlations with non-centred Pearson's index. Crops: SC-summer, WC-winter, MC-mixed, WPC-woody permanent, HPC-herbaceous permanent. Forests: DF-deciduous, EF-evergreen, MF-mixed. R-roads, B-buildings, M-meadows, Mo-moorlands, F-fallows

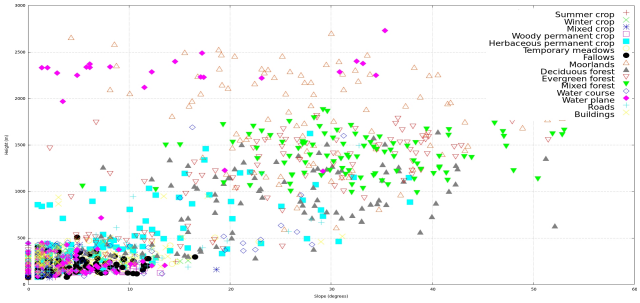


Fig. 4. Field graph of thematic classes in the elevation/slope space. Height is in meters, and slope in degrees

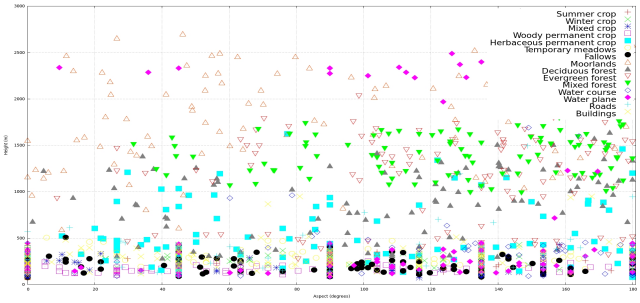


Fig. 3. Field graph of thematic classes in the elevation/orientation space. Height is in meters and orientation in degrees, with 0 for south and 180 for north

vious step, since such validation introduces the error of the SVM classification into the assessment of sampling. Secondly, unsupervised classification with a clustering algorithm will be deployed, with cluster initialisation given by the samples. Again, a kappa index for cluster comparison will be applied for the assessment of the classification. Other work includes the application of decision trees for classification, and also the implementation of a improved interpolation algorithm for cloud elimination during dataset preparation.

7. REFERENCES

[1] D. Villa O. Hagolle, M. Huc and G. Dedieu, "A multi-temporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images," *Remote Sensing of the Environment*, vol. 114, pp. 1747–1755, 2010.

[2] F. J. Gallego, "The efficiency of sampling very high resolution images for area estimation in the European Union," *International Journal of Remote Sensing*, vol. 33, pp. 1868–1880, 2012.

[3] Centre National d'Études Spatiales CNES, "Orfeo toolbox," http://orfeo-toolbox.org/otb, [Online; accessed 31st May 2012].

[4] A. Savitzky and M.J.E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Analytical Chemistry*, vol. 36, pp. 1627–1639, 1964.

[5] Forespir, "Projet Silvapyr," http://www.forespir.com/les-projets-en-cours/silvapyr, 2008, [Online; accessed 31st May 2012].

[6] "Géoportail," http://www.geoportail.fr, [Online; accessed 31st May 2012].

[7] S. V. Stehman, "Sampling designs for accuracy assessment of land cover," *International Journal for Remote Sensing*, vol. 30, pp. 5243–5272, 2009.

NON-LINEAR TIME SAMPLING DRIVEN BY SURFACE TEMPERATURE FOR THE MONITORING OF VEGETATED AREAS USING MULTI- AND HYPER-TEMPORAL SATELLITE IMAGE TIME SERIES

I. Rodes, J. Inglada, O. Hagolle, J.F. Dejoux, G. Dedieu

CESBIO - UMR 5126
18 avenue Edouard Belin
31401 TOULOUSE CEDEX 9 - France

ABSTRACT

This work presents a methodology for the fast exploitation of the large volumes of high temporal and spectral resolution data that will be available with the future Earth Observation missions. A new approach integrating temperature and phenological information for the characterisation of land cover classes is given, as part of a fully automatic system for the generation of large area land cover maps. No selection of cloud-free dates, masking of unsuitable regions, or user interaction is needed. Analysis of its performance is undertaken, and future directions are identified.

1. INTRODUCTION

The new generation optical Earth observation missions at high resolution will soon deeply reconfigure our current understanding of environmental monitoring, yielding unseen volumes of data in terms of spectral richness, temporal revisit and spatial resolution. The production of land cover maps based on the exploitation of such volumes of information for large areas is challenging, especially in terms of excessive processing costs due to the quantities of information to be treated, and also in relation to data variability determined by the diversity to be found in land classes across large regions. Traditional analyses may be made non-viable, not only due to the volumes of the coming high resolution time series, but also to the varying temporal evolutions of land cover elements in the different environments to be found in regional scale contexts. Present-day approaches for land cover mapping with delivery in operational times tend to be tailored to local and specific problems; processing of large territories seems to be too costly in terms of processing time, either due to extended computation with complex models, or to the requirement of intensive manual labour when large areas are considered. These approaches are thus inefficient for the use of the new type of data that missions such as ESA's Sentinel-2 and NASA-USGS Landsat 8 will provide. Thus, a need arises for the implementation of methods that allow for generalisation to large scale areas with high temporal and spectral resolution, in order to provide the remote sensing community with tools to ease the fast and accurate exploitation of information from these new generation sensors.

2. OBJECTIVE

The objective of this work is to provide a methodology facilitating the generation of large area land cover maps based on multi- and hyper-temporal resolution multi-spectral images over very large areas and with feasible processing costs. Application is at regional scales, and the two main challenges to be overcome are the management of the considerable temporal variability of thematic classes across regions and the minimisation of computation time when dealing with such high volumes of data. With the first of these aims, the introduction of climate-related exogenous data has been implemented. In particular, the temporal evolution of land cover has been analysed not as a function of linear time, but as a function of degree-day and accumulated temperature information, due to the fact that phenological processes are strongly linked to the integral of temperature at a given location. In this way, the study of non-linear time sampling driven by surface temperature using multi-temporal satellite image time series for the monitoring of vegetated areas is undertaken, and conclusions can be drawn.

3. DATA

3.1. Study area and datasets

The French part of the Pyrénées mountain range has been chosen as study area. Work has been based on Landsat 5 and Landsat 7 data, provided by USGS and ESA, and previously processed at CESBIO. These Landsat scenes cover the region during 2010 with a spatial resolution of 30 meters, disposed in tiles of 110x110km. Due to their similarities, Landsat (two

This work was supported by CNES under the TOSCA program.

• Seasonal crops	• Forests
∗ Summer crop	∗ Deciduous forest
∗ Winter crop	∗ Evergreen forest
∗ Mixed crop	∗ Mixed forest
• Permanent crops	• Temporary meadows
∗ Woody permanent crop	• Fallows
∗ Herbaceous permanent crop	• Roads
• Moorlands	• Buildings
• Water areas	

Table 1. Nomenclature for validation.

satellites Landsat 5 and Landsat 7, 30 m resolution, 180 km field of view, 7 bands including SWIR, and a repetitivity of 8 days) can be used as a simulator for Sentinel-2, although the resolution and repetitivity of the later will be far better.

3.2. Validation data and nomenclature

Land cover validation data have been generated based on the IFN database for forests, the BD Carthage for hydrology, and the BD Topo dataset for buildings and roads, all by the Institut Géographique National (IGN) [1], and the Registre Parcellaire Graphique (RPG) based on farmer crop declarations [2]. The applied nomenclature is given in Table 1.

3.3. Exogenous data

Temperature data applied in this study have been provided by SAFRAN (*Système d'Analyse Fournissant des Renseignements Atmosphériques à la Neige*), an atmospheric analysis system for surface variables by Météo-France, yielding ground data observations with an hourly time step. These observations include air temperature, wind speed, relative humidity, quantity and phase of precipitation, and radiation. They are based on a method of interpolation integrating the observations of the Météo-France network of stations, information on climatically homogeneous zones and height differences [3], their outputs being provided for a grid of 8 km resolution covering the entire France. Fig.1 illustrates the SAFRAN grid that has been used here.

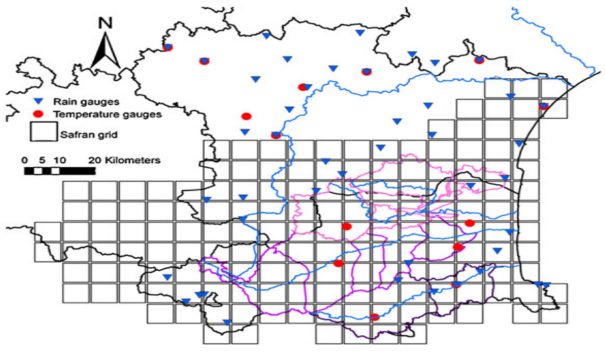


Fig. 1. SAFRAN grid, rain and temperature gauges [4].

4. METHODOLOGY

4.1. Strategy

Phenology is generally understood as the study of timings for recurring biological events such as green-up or senescence for the seasonal stages of vegetation, and their characteristics and causes. A variety of factors that determine the phenological evolution of vegetation have been studied [5] [6]. Some of the highlighted variables include species type, climate, day length, age and substrate [7] [8]. In particular, the impact of photoperiod and temperature patterns has been most established. [9] examines the determinant effects and interactions of both variables on two types of Mediterranean oaks, confirming differing responsiveness of bud-burst to temperature and photoperiod between species; [10] analyses determinant phenology sensitivity to temperature for different tree species in European forests and differences between and within species and along altitudinal and temporal gradients. [11] identifies distinct patterns of crop phenological development in Germany in relation to temperature and day length. These are determined not only by yearly climate conditions, but also greatly by topographic configuration.

Consequently, spectral, temperature and topographic information have been jointly incorporated in the presented approach. In particular, spectral evolution of land cover has been analysed not as a function of time, but as a function of temperature degree-day and accumulated temperature information obtained from the SAFRAN dataset. The rationale behind this change of axis is the fact that measures of accumulated heat can be considered as physiological time for many organisms, giving a common reference for their development and thus, in the case of vegetation, their phenological evolution and consequent phenologically-determined spectral response.

When degree days are used, upper and lower developmental thresholds are applied. These thresholds mark the range of temperatures where development of the organisms happens. For vegetation, they have been set here to 10°C and 30°C, respectively. Fig.2(b) shows the degree-day cumulative curves of different geographical points near and within the Pyrénées, to which spectral responses have been mapped. Fig.2(a) illustrates the ranges concept that has been applied here, and then used to compute the new non-linear time sampling for determining the variety of features to characterise the different land cover classes considered.

4.2. Experimental setup

A fully automatic processing chain has been implemented using the CNES Orfeo Toolbox Library for remote sensing image processing [12] comprising different stages. Firstly, gap filling is undertaken on the Landsat time series. Gaps

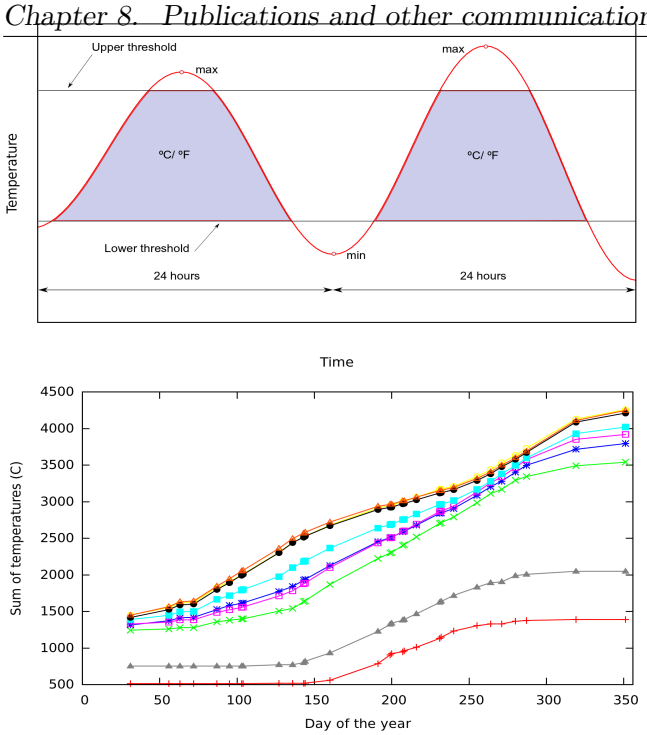


Fig. 2. Degree-day range threshold concept and sum of temperature curves for 8 random example location points for one of the regions considered, with 29 available dates across the year.

are due to clouds and cloud shadows, as detected by [13], and missing data outside the satellite swath. The method combines interpolation and averaging within a time step for the processing of each gap pixel, and yields a product based on non-linear time sampling driven by surface temperature linked to each analysed geographical point. In this way, spectral evolution in the time series is reformulated as a function of temperature as provided by the SAFRAN system.

Secondly, primitives to be applied, such as the Normalised Difference Vegetation Index or the Normalised Difference Water Index, are calculated based on the new temperature-driven available information. Stratified sampling based on topographical information can then be applied to maximally capture the variability of each class with a minimum of samples, as suggested by the study of intra-class variance and inter-class correlation given in [14], in order to minimise learning and classification times of a subsequent supervised classification stage with support vector machines (SVMs).

Temperature and degree-day products are obtained from 8km-resolution hourly SAFRAN listings of temperature data by a transformation of projection coordinates and a resampling process with a combination of bi-cubic and nearest neighbours interpolation for regions bordering areas with no temperature information. A double classifier system has

Approach (%)	a	b	c	d
(a) Linear	70.266	67.236	70.411	40.174
(b) Non-linear (temperatures)	71.393	67.598	68.062	40.134
(c) Non-linear (degree days)	70.227	64.676	69.690	37.073

Approach (%)	e	f	g	h	i
(a) Linear	87.401	64.041	76.817	76.096	76.024
(b) Non-linear (t)	87.024	57.314	74.626	72.106	71.705
(c) Non-linear (dd)	74.618	56.509	66.416	62.247	62.794

Table 2. Automatic classification results. Percentage of well-classified pixels per class. Nomenclature: a-summer crops, b-winter crops, c-woody permanent crops, d-fallows, meadows and herbaceous areas, e-moorlands, f-deciduous forests, g- evergreen forests, h-water areas, i-buildings.

also been implemented, with a targeted treatment of areas specially affected by snow throughout the year. To this aim, MODIS snow cover product data [15] processed at CESBIO has been applied to automatically estimate snow precipitation probabilities in regions and to define their learning and classification based on snow-free periods. Finally, land cover maps have been generated for the whole study area, the complete processing chain described here being completely automatic.

5. RESULTS AND EVALUATION

Evaluation of results is undertaken by performing SVM classification with the generated data for the whole studied area, with the nomenclature specified in section 3.2. The use of non-linear time sampling driven by surface temperature with degree-day and cumulative temperature approaches, as explained above, determines the feature characterisation of the targeted classes. Then, different sampling strategies based on the automatic selection of a fixed number of pixels per class have been investigated, determining the selection of the learning samples itself for the subsequent classification. Training on only a minimal percentage of the pixels is thus made possible (less than 0.1%), generating a model to be applied for the mapping of the entire region. Classification results in terms of accuracy and computation costs with the presented approach are given in Tables 2 and 5.

Different approaches have been implemented for time series manipulation by applying both linear and non-linear time sampling, using the integral of daily temperatures or degree days as a new axis of analysis with a range of configurations. Preliminary results as shown in Tables 2 and 5 show good average percentages of well-classified pixels per class for all tested approaches (validation data contain numerous errors, and greater repetitivity will be available with Sentinel-2), both linear or date-based and non-linear, with an apparent better performance of the sum of temperatures solution in comparison to the degree-day version, and all of them with

Sampling	Type:	Stratified
	Stratification variable:	Altitude (5 clusters)
Classification	Classes:	11
	Samples:	5500 (500 per class)
Data	Input images:	29 (110x110km, 30m)
	Dimensions:	$3665 \times 3665 \times 6$
	Pixels to classify:	13.4×10^6
	% sampled pixels:	0.04094%
Costs	Learning time:	0m 15.66s
	Classification time:	26m 09.23s
	Time per pixel:	1.1×10^{-4}s

Table 3. Automatic classification processing costs.

operational computation times.

No significant improvement is observable by the application of non-linear approaches; however, similar performance results to that of linear temporal sampling indicates the validity of the chosen method for feature characterisation. Issues and future work directions have been identified: firstly, SAFRAN data has 8 km resolution and thus does not provide real temperature information for a given Landsat-resolution 30-m pixel. The impact of this resolution mismatch is not negligible, as it is precisely this information which determines a new axis of temporal interpolation for class characterisation. Secondly, results are also much sensitive to the sampling of virtual temperature points which depend on the obtained SAFRAN temperatures. Finally, north and south expositions are averaged when interpolated on the SAFRAN grid, and thus important differences of temperature between north and south facing slopes are not reflected in our input data. In order to deal with this point, efforts will be directed in the future to improve the available temperature data by use of topographical information. In particular, focus will be given to the introduction of compensation for orientation, slope and height as in [16].

6. CONCLUSIONS

In this work, an approach providing a methodology for the exploitation of the great volumes of information that will soon be available from the future new generation Earth observation missions is presented. It is not specific to a geographic area, and it targets the processing of large regions, for which multi- and hyper-temporal series are available, encompassing a variety of crops, forests and other classes. The novelty of this work is the inclusion of temperature and phenological information in order to characterise and sample a variety of land cover classes. Thus, natural factors and their consequences, such as climatology and its effects on different vegetation evolutions are included in the design of the method itself. In this way, one of the common issues for large-area classification, that is to say, the difficulty of characterisation due to intra-class variability of temporal response, can be tackled by diminishing data variability with a method that is designed for generalisation.

7. REFERENCES

[1] Institut National de l'Information Géographique et Forestière, "IGN," http://www.ign.fr, [Online; accessed 30th December 2012].

[2] "Géoportail," http://www.geoportail.fr, [Online; accessed 30th December 2012].

[3] P. Quintana-Segu, P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, and S. Morel, "Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France," *Journal of Applied Meteorology and Climatology*, vol. 47, pp. 92–107, 2008.

[4] K. Chaouche, L. Neppel, C. Dieulin, N. Pujol, B. Ladouche, E. Martin, D. Salas, and Y. Caballero, "Analyses of precipitation, temperature and evapotranspiration in a French Mediterranean region in the context of climate change," *Comptes Rendus Geoscience*, vol. 342, no. 3, pp. 234–243, 2010.

[5] M. White, P. Thornton, and S. Running, "A continental phenology model for monitoring vegetation responses to interannual climatic variability," *Global Biogeochemical Cycles*, vol. 11, pp. 217–234, 1997.

[6] J. Partanen, V. Koski, and H. Hanninen, "Effects of photoperiod and temperature on the timing of bud burst in Norway spruce (Picea Abies)," *Tree Physiology*, vol. 18, no. 811-816, 1998.

[7] J. Schaber and F. Badeck, "Physiology-based phenology models for forest tree species in Germany," *International Journal of Biometeorology*, vol. 47, pp. 193–201, 2003.

[8] D. Wolfe, M. Schwartz, A. Lakso, Y. Otsuki, R. Pool, and N. Shaulis, "Climate change and shifts in spring phenology of three horticultural woody perennials in northeastern USA," *International Journal of Biometeorology*, vol. 49, pp. 303–309, 2005.

[9] V. Sanz-Perez, P. Castro-Dez, and F. Valladares, "Differential and interactive effects of temperature and photoperiod on budburst and carbon reserves in two co-occurring Mediterranean oaks," *Plant Biology*, vol. 11, no. 2, pp. 142–151, 2009.

[10] Y. Vitasse, S. Delzon, E. Dufrêne, J. Pontailler, J. Louvet, A. Kremer, and R. Michalet, "Leaf phenology sensitivity to temperature in European trees: do within-species populations exhibit similar responses?," *Agricultural and Forest Meteorology*, vol. 149, no. 5, pp. 735–744, 2009.

[11] S. Siebert and F. Ewert, "Spatio-temporal patterns of phenological development in Germany in relation to temperature and day length," *Agricultural and Forest Meteorology*, vol. 152, no. 1, pp. 44–57, 2012.

[12] Centre National d'Études Spatiales CNES, "Orfeo Toolbox," http://orfeo-toolbox.org/otb, [Online; accessed 30th December 2012].

[13] O. Hagolle, M. Huc, D. Villa Pascual, and G. Dedieu, "A multitemporal method for cloud detection, applied to FORMOSAT-2, VENμS, LANDSAT and SENTINEL-2 images," *Remote Sensing of Environment*, vol. 114, pp. 1747–1755, 2010.

[14] I. Rodes, J. Inglada, O. Hagolle, J.F. Dejoux, and G. Dedieu, "Sampling strategies for unsupervised classification of multitemporal high resolution optical images over very large areas," in *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, Muenchen, 2012.

[15] D.K. Hall, V.V. Salomonson, and G.A. Riggs, *MODIS/Terra Snow Cover 8-Day L3 Global 500m Grid (v5). Maximum Snow Extent, January 2009 to January 2012, 4320N, 142W; 415N, 246E*, Boulder, Colorado USA: National Snow and Ice Data Center, 2006.

[16] K.L. Bristow and G.S. Campbell, "On the relationship between incoming solar radiation and daily maximum and minimum temperature," *Agricultural and Forest Meteorology*, vol. 31, pp. 159–166, 1984.

Bibliography

[Aitkenhead and Aalders, 2011] Aitkenhead, M. and Aalders, I. (2011). Automatic land cover mapping of Scotland using expert system and knowledge integration methods. *Remote Sensing of Environment*, 115:1285–1295.

[Akbani et al., 2004] Akbani, R., Kwek, S., and Japkowicz, N. (2004). Applying support vector machines to imbalanced data sets. In *European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases*, pages 39–50.

[Alcaraz-Segura et al., 2009] Alcaraz-Segura, D., Cabello, J., and Paruelo, J. (2009). Baseline characterization of major Iberian vegetation types based on the NDVI dynamics. *Plant Ecology*, 202(1):13–29.

[Aono and Kazui, 2008] Aono, Y. and Kazui, K. (2008). Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. *International Journal of Climatology*, 28(7):905–914.

[Arora and Foody, 1997] Arora, M. and Foody, G. (1997). Log-linear modelling for the evaluation of the variables affecting the accuracy of probabilistic, fuzzy and neural network classifications. *International Journal of Remote Sensing*, (18):785–798.

[Asner, 2001] Asner, G. (2001). Cloud cover in Landsat observations of the Brazilian Amazon. *International Journal of Remote Sensing*, 18(22):3855–3862.

[Asner et al., 2002] Asner, G., Keller, M., Pereira, R., and Zweede, J. (2002). Assessing limitations based on detailed field observations, Landsat ETM+, and textural analysis. *Remote Sensing of Environment*, 80:483–496.

[ASP, 2013] ASP (2013). Registre Parcellaire Graphique (RPG). Available online at: www.asp-public.fr. (Agence de Services et de Paiement).

[Atkeson et al., 1997] Atkeson, C. G., Moore, A. W., and Schaal, S. (1997). Locally weighted learning. *Artificial Intelligence Review*, 11(1-5):11–73.

[Atkinson, 1991] Atkinson, P. (1991). Optimal ground-based sampling for remote-sensing investigations - estimating the regional mean. *International Journal of Remote Sensing*, (12):559–567.

[Auslander et al., 2003] Auslander, M., Nevo, E., and Inbar, M. (2003). The effects of slope orientation on plant growth, developmental instability and susceptibility to herbivores. *Journal of Arid Environments*, 55(3):405–416.

[Ban et al., 2015] Ban, Y., Gong, P., and Giri, C. (2015). Global land cover mapping using earth observation satellite data: Recent progresses and challenges. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:1 – 6. Global Land Cover Mapping and Monitoring.

[Baret et al., 1989] Baret, F., Guyot, G., and Major, D. (1989). TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 1355–1358, Vancouver.

[Barnett and Lewis, 1994] Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. Computers and Geosciences.

[Barry and Chorley, 1987] Barry, R. and Chorley, R. (1987). *Atmosphere, Weather and Climate*. Routledge.

[Bartholomé and Belward, 2005] Bartholomé, E. and Belward, A. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, 26(9):1959–1977.

[Bartlett et al., 1980] Bartlett, J., Kotrlik, J., and Higgins, C. (1980). Evaluation of several schemes for classification of remotely sensed data. *Photogrammetric Engineering and Remote Sensing*, 46:1547–1553.

[Beck et al., 2006] Beck, P., Atzberer, C., Høgda, K., Johansen, B., and Skidmore, A. (2006). Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI. *Remote Sensing of Environment*, 100(321).

[Berberoglu et al., 2007] Berberoglu, S., Curran, P., Lloyd, C., and Atkinson, P. (2007). Texture classification of Mediterranean land cover. *International Journal of Applied Earth Observation and Geoinformation*, 9(3):322 – 334.

[Bignalet-Cazalet et al., 2010] Bignalet-Cazalet, F., S. Baillarin, D. G., and Panem, C. (2010). Automatic and generic mosaicing of satellite images. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 288–290, Honolulu.

[Bojinski and Richter, 2010] Bojinski, S. and Richter, C. (2010). Prioritizing global observations along essential climate variables. *Transactions American Geophysical Union*, 91(51):505–505.

[Bontemps et al., 2015] Bontemps, S., Boettcher, M., Brockmann, C., Kirches, G., Lamarche, C., Radoux, J., Santoro, M., Bogaert, E. V., Wegmüller, U., Herold, M., Achard, F., Ramoino, F., Arino, O., and Defourny, P. (2015). Multi-year global land cover mapping at 300m and characterization for climate modelling: achievements of the land cover component of the ESA Climate Change Initiative. In *36th International Symposium on Remote Sensing of Environment*, pages 323–328, Berlin.

[Bontemps et al., 2010] Bontemps, S., Defourny, P., Bogaert, E. V., Weber, J. L., and Arino, O. (2010). Globcorine - a joint EEA-ESA project for operational land cover and land use mapping at pan-European scale. In *European Space Agency Living Planet Symposium*, Bergen, Norway.

[Borak and Jasinski, 2009] Borak, J. and Jasinski, M. (2009). Effective interpolation of incomplete satellite-derived leaf-area index time series for the continental United States. *Agricultural and Forest Meteorology*, 149(2):320–332.

[Brewer and Hanif, 1983] Brewer, K. and Hanif, M. (1983). Sampling with unequal probability. *Lecture notes in Statistics*, (15):1–164.

[Broge and Leblanc, 2001] Broge, N. and Leblanc, E. (2001). Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sensing of Environment*, 76(2):156–172.

[Burges, 1998] Burges, C. (1998). A tutorial on support vector machines for pattern recognition. data mining and knowledge discovery. *IEEE Transactions on Systems, Man and Cybernetics*, 2(2):121–167.

[Campbell, 2006] Campbell, J. (2006). *Introduction to Remote Sensing*. Taylor and Francis, London.

[Candade and Dixon, 2004] Candade, N. and Dixon, B. (2004). Multispectral classification of Landsat images: a comparison of support vector machine and neural network classifiers. In *ASPRS Annual Conference Proceedings, Denver*.

[Cao et al., 2009] Cao, X., Chen, J., Matsushita, B., Imura, H., and Wang, L. (2009). An automatic method for burn scar mapping using support vector machines. *International Journal of Remote Sensing*, 30(3):577–594.

[Carrão et al., 2008] Carrão, H., Gonçalves, P., and Caetano, M. (2008). Contribution of multispectral and multitemporal information from MODIS images to land cover classification. *Remote Sensing of Environment*, 112(3):986–997.

[Chander et al., 2009] Chander, G., Markham, B. L., and Helder, D. L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM+, and EO-1 ALI sensors. *Remote Sensing of Environment*, 113(5):893–903.

[Chaouche et al., 2010] Chaouche, K., Neppel, L., Dieulin, C., Pujol, N., Ladouche, B., Martin, E., Salas, D., and Caballero, Y. (2010). Analyses of precipitation, temperature and evapotranspiration in a French Mediterranean region in the context of climate change. *Comptes Rendus Geoscience*, 342(3):234–243.

[Cheema and Bastiaanssen, 2010] Cheema, M. and Bastiaanssen, W. (2010). Land use and land cover classification in the irrigated Indus Basin using growth phenology information from satellite data to support water management analysis. *Agricultural Water Management*, 97:1541–1552.

[Chen and Stow, 2002] Chen, D. and Stow, D. (2002). The effect of training strategies on supervised classification at different spatial resolutions. *Photogrammetric Engineering and Remote Sensing*, (68):1155–1161.

[Chen et al., 2015] Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., Zhang, W., Tong, X., and Mills, J. (2015). Global land cover mapping at 30m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, (103):7–27.

[Chen et al., 2004] Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., and Eklundh, L. (2004). A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sensing of Environment*, 91:332–334.

[Cihlar, 2000] Cihlar, J. (2000). Land cover mapping of large areas from satellites: Status and research priorities. *International Journal of Remote Sensing*, 21(6-7):1093–1114.

[Cochran, 1977] Cochran, W. (1977). *Sampling Techniques*. John Wiley and Sons, New York.

[Cong et al., 2012] Cong, N., Piao, S., Chen, A., Wang, X., Lin, X., Chen, S., Han, S., Zhou, G., and Zhang, X. (2012). Spring vegetation green-up date in China inferred from SPOT NDVI data: a multiple model analysis. *Agricultural and Forest Meteorology*, 165:104–113.

[CRAMP, 1995] CRAMP (1995). Les grands ensembles morpho-pédologiques de la région Midi-Pyrénées. Technical report. (Chambre Régionale d'Agriculture de Midi-Pyrénées).

[Crist and Cicone, 1984] Crist, E. P. and Cicone, R. C. (1984). A physically-based transformation of Thematic Mapper data - the TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing*, 22(3):256–263.

[CSAEOS-NRC, 2008] CSAEOS-NRC (2008). *Earth Observations from Space: The First 50 Years of Scientific Achievements*. The National Academies Press. Committee on Scientific Accomplishments of Earth Observations from Space, National Research Council.

[Deheuvels, 1977] Deheuvels, P. (1977). Estimation non-paramétrique de la densité par histogrammes généralisés. *Revue Statistique Appliqué*, (25):5–42.

[Dethier et al., 1973] Dethier, B., Ashley, M., Blair, B., Caprio, J., and Rouse, J. (1973). Phenology satellite experiment (detection of brown wave and green wave in north-south corridors of the United States. Technical report. Ithaca NY: Cornell University.

[Dixon and Candade, 2008] Dixon, B. and Candade, N. (2008). Multispectral land use classification using neural networks and support vector machines: one or the other, or both? *International Journal of Remote Sensing*, 29(4):1185–1206.

[Doi and Katano, 2008] Doi, H. and Katano, I. (2008). Phenological timings of leaf budburst with climate change in Japan. *Agricultural and forest meteorology*, 148(3):512–516.

[Doraiswamy et al., 2007] Doraiswamy, P., Stern, A. J., and Akhmedov, B. (2007). Crop classification in the U. S. corn belt using MODIS imagery. In *Proceedings of the International Geoscience and Remote Sensing Symposium IGARSS*, Barcelona.

[EC, 2014] EC (2014). Copernicus. The European Earth Observation Programme. Available online at: copernicus.eu. (European Commission).

[EEA, 2007] EEA (2007). CLC2006 Technical Guidelines. Technical Report 17, Copenhagen. (European Environment Agency).

[Epanechnikov, 1969] Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, (14):153–158.

[ESA, 2010] ESA (2010). GMES Sentinel-2 Mission Requirements Document. Available online at: www.esa.int. (European Space Agency).

[ESA-UCL, 2011] ESA-UCL (2011). GLOBCOVER 2009. Products description and validation report. Available online at: due.esrin.esa.int/globcover. (European Space Agency and Université Catholique de Louvain).

[Estrella et al., 2009] Estrella, N., Sparks, T., and Menzel, A. (2009). Effects of temperature, phase type and timing, location, and human density on plant phenological responses in Europe. *Climate Research*, 3(39):235–248.

[FAO, 2013] FAO (2013). Forestry. Available online at: www.fao.org/forestry/en. (United Nations Food and Agriculture Organisation).

[Feranec et al., 2007] Feranec, J., Hazeu, G., Christensen, S., and Jaffrain, G. (2007). Corine land cover change detection in Europe (case studies of the Netherlands and Slovakia). *Land Use Policy*, 24:234–247.

[Fletcher, K., 2012] Fletcher, K. (2012). Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. Available online at: sentinel.esa.int.

[Foody and Mathur, 2004a] Foody, G. and Mathur, A. (2004a). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, (42):1335–1343.

[Foody and Mathur, 2004b] Foody, G. M. and Mathur, A. (2004b). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, 93(1-2):107–117.

[Friedl et al., 2010] Friedl, M., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., and Huang, X. (2010). MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114(1):168 – 182.

[Fu, 2000a] Fu, P. (2000a). A geometric solar radiation model with applications in landscape ecology. Technical report. PhD Thesis, Department of Geography, University of Kansas, US.

[Fu, 2000b] Fu, P. (2000b). The solar analyst 1.0 manual. Technical report. Helios Environmental Modeling Institute HEMI, US.

[Galiano and Chica-Olmo, 2012] Galiano, V. R. and Chica-Olmo, M. (2012). Land cover change analysis of a Mediterranean area in Spain using different sources of data: multi-seasonal Landsat images, land surface temperature, digital terrain models and texture. *Applied Geography*, 35(1-2):208–218.

[Gandin, 1965] Gandin, L. (1965). Objective analysis of meteorological fields. *Israel Program for Scientific Translation*.

[Gao, 1996] Gao, B. (1996). NDWI - A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, (58):257–266.

[Gao et al., 2008] Gao, F., Morisette, J., Wolfe, R., Ederer, G., Pedelty, J., Masuoka, E., Myneni, R., Tan, B., and Nightingale, J. (2008). An algorithm to produce temporally and spatially continuous MODIS-LAI time series. *IEEE Geoscience and Remote Sensing Letters*, 5(1):60–64.

[Gascoin, 2013] Gascoin, S. (2013). Gap-filling of MOD10A2 snow cover data over the Pyrenees. Technical report, Centre d'Études Spatiales de la BIOsphère (CESBIO).

[GCOS, 2010] GCOS (2010). Implementation plan for the Global Observing System for Climate in support of the UNFCCC. 2010 update, GCOS-138. Available online at: www.wmo.int/pages/prog/gcos. (Global Observing Systems Information Center).

[Geebelen et al., 2012] Geebelen, D., Suykens, J., and Vandewalle, J. (2012). Reducing the number of support vectors of SVM classifiers using the smoothed separable case approximation. *IEEE Transactions Neural Networks and Learning Systems*, 23(4):682–688.

[GEO, 2005] GEO (2005). The Global Earth Observation System of Systems (GEOSS). 10-Year Implementation Plan Reference Document. Technical report. Available online at: earthobservations.org.

[Giri, 2012] Giri, C. (2012). *Remote sensing of land use and land cover. Principles and applications.* CRC Press.

[Giri et al., 2013] Giri, C., Pengra, B., Long, J., and Loveland, T. (2013). Next generation of global land cover characterization, mapping, and monitoring. *International Journal of Applied Earth Observation and Geoinformation*, 25:30–37.

[Gong et al., 2013] Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., Li, C., Li, X., Fu, W., Liu, C., Xu, Y., Wang, X., Cheng, Q., Hu, L., Yao, W., Zhang, H., Zhu, P., Zhao, Z., Zhang, H., Zheng, Y., Ji, L., Zhang, Y., Chen, H., Yan, A., Guo, J., Yu, L., Wang, L., Liu, X., Shi, T., Zhu, M., Chen, Y., Yang, G., Tang, P., Xu, B., Giri, C., Clinton, N., Zhu, Z., Chen, J., and Chen, J. (2013). Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *International Journal of Remote Sensing*, 34(7):2607–2654.

[Goodwin et al., 2013] Goodwin, N. R., Collett, L. J., Denham, R. J., Flood, N., and Tindall, D. (2013). Cloud and cloud shadow screening across Queensland, Australia: An automated method for Landsat TM/ETM+ time series. *Remote Sensing of Environment*, 134(0):50–65.

[Gorry, 1990] Gorry, P. (1990). General least-squares smoothing and differentiation by the convolution Savitzky-Golay method. *Analytical Chemistry*, 62(6):570–573.

[Gregorio and Jansen, 2005] Gregorio, A. D. and Jansen, L. (2005). Land Cover Classification System (LCCS). Classification concepts and user manual. Software version 2. Technical report. Available online at: www.glcn.org.

[Grekousis et al., 2015] Grekousis, G., Mountrakis, G., and Kavouras, M. (2015). An overview of 21 global and 43 regional land-cover mapping products. *International Journal of Remote Sensing*, 36(21):5309–5335.

[Gruber, 1980a] Gruber, M. (1980a). Dissymétrie climatique et forestière dans les Pyrénées. *Forêt mediterranéenne*, 2(1):135–140.

[Gruber, 1980b] Gruber, M. (1980b). Etages et séries de végétation de la chaîne pyrénéenne. *Ecologia Mediterranea*, (5):147–174.

[Gupta et al., 2014] Gupta, M., Gao, J., Aggarwal, C., and Han, J. (2014). Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 25:1–220.

[Guyot, 1999] Guyot, G. (1999). *Climatologie de l'environnement*. Dunod, Paris.

[Hagolle et al., 2008] Hagolle, O., Dedieu, G., Mougenot, B., Debaecker, V., Duchemin, B., and Meygret, A. (2008). Correction of aerosol effects on multi-temporal images acquired with constant viewing angles: Application to Formosat-2 images. *Remote Sensing of Environment*, 112:1689–1701.

[Hagolle and Huc, 2011] Hagolle, O. and Huc, M. (2011). Manuel de l'utilisateur des données produites au CESBIO. Séries temporelles de produits Landsat de Niveau 2A. Technical report, Centre d'Études Spatiales de la BIOsphère (CESBIO).

[Hagolle et al., 2010] Hagolle, O., Huc, M., Pascual, D. V., and Dedieu, G. (2010). A multi-temporal method for cloud detection, applied to Formosat-2, Venμs, Landsat and Sentinel-2 images. *Remote Sensing of Environment*, 114:1747–1755.

[Hall et al., 2002] Hall, D., Riggs, G., Salomonson, V., DiGirolamo, N., and Bayr, K. (2002). MODIS snow cover products. *Remote Sensing of Environment*, 83(1-2):181–194.

[Hall et al., 2006] Hall, D., Salomonson, V., and Riggs, G. (2006). *MODIS/Terra and MODIS/Aqua Snow Cover 8-Day L3 Global 500m Grid (v5). Maximum Snow Extent, January 2009 to January 2012*. Boulder, Colorado USA: National Snow and Ice Data Center.

[Hansen and Loveland, 2012] Hansen, M. and Loveland, T. (2012). A review of large area monitoring of land cover change using Landsat data. *Remote Sensing of Environment*, 122(0):66–74.

[Haralick et al., 1973] Haralick, R. M., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 6:610–621.

[Hartkamp et al., 1999] Hartkamp, A., Beurs, K. D., Stein, A., and White, J. (1999). Interpolation techniques for climate variables. Technical report, Mexico D.F. Available online at: www.cimmyt.org.

[Hassan et al., 2007a] Hassan, Q., Bourque, C., and Meng, F. (2007a). Application of Landsat-7 ETM+ and MODIS products in mapping seasonal accumulation of growing degree days at an enhanced resolution. *Journal of Applied Remote Sensing*, 1.

[Hassan et al., 2007b] Hassan, Q., Bourque, C., and Meng, F. (2007b). Spatial mapping of growing degree days: an application of MODIS-based surface temperatures and enhanced vegetation index. *Journal of Applied Remote Sensing*, 1.

[Hassan and Rahman, 2012] Hassan, Q. and Rahman, K. (2012). Applicability of remote sensing-based surface temperature regimes in determining deciduous phenology over boreal forest. *Journal of Plant Ecology*, pages 1–8.

[Hegarat-Mascle et al., 2000] Hegarat-Mascle, S. L., Quesney, A., Vidal-Madjar, D., Taconet, O., Normand, M., and Loumagne, C. (2000). Land cover discrimination from multitemporal ERS images and multispectral Landsat images: A study case in an agricultural area in France. *International Journal of Remote Sensing*, 21(3):435–456.

[Hird and McDermid, 2005] Hird, J. and McDermid, G. (2005). Noise reduction of NDVI time series: an empirical comparison of selected techniques. *Remote Sensing of Environment*, 113(1):248–258.

[Homolová et al., 2013] Homolová, L., Malenovský, Z., Clevers, J., García-Santos, G., and Schaepman, M. (2013). Review of optical-based remote sensing for plant trait mapping. *Ecological Complexity*, 15:1–16.

[Hong et al., 2014] Hong, G., Zhang, A., Zhou, F., and Brisco, B. (2014). Integration of optical and synthetic aperture radar (SAR) images to differentiate grassland and alfalfa in Prairie area. *International Journal of Applied Earth Observation and Geoinformation*, 28:12–19.

[Huang et al., 2002a] Huang, C., Davis, L. S., and Townshend, J. R. G. (2002a). An assessment of support vector machines for land cover classication. *International Journal of Remote Sensing*, 23(4):725–749.

[Huang et al., 2008] Huang, C., Song, K., Kim, S., Townshend, J., Davis, P., Masek, J., and Goward, S. (2008). Use of a dark object concept and support vector machines to automate forest cover change analysis. *Remote Sensing of Environment*, 112(3):970–985.

[Huang et al., 2002b] Huang, C., Wylie, B., Yang, L., Homer, C., and Zylstra, G. (2002b). Derivation of a tasseled cap transformation based on Landsat 7 satellite reflectance. *International Journal of Remote Sensing*, 23(8):1741–1748.

[Huete, 1988] Huete, A. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3):295–309.

[Hungerford et al., 1989] Hungerford, R., Nemani, R., Running, S., and Coughlan, J. (1989). MTCLIM: A Mountain Microclimate Simulation Model. *United States Department of Agriculture. Research Paper*.

[Hwang et al., 2011] Hwang, T., Song, C., Vose, J., and Band, L. (2011). Topography-mediated controls on local vegetation phenology estimated from MODIS vegetation index. *Landscape Ecology*, 26(4):541–556.

[Ibarra Benlloch, P., 2002] Ibarra Benlloch, P. (2002). El medio natural de los Pirineos: límites y condicionantes para el desarrollo de actividades económicas. pages 9–42.

[IGN, 2006] IGN (2006). BD Carthage v.3.0. Descriptif de contenu. Available online at: www.ign.fr. (Institut National de l'Information Géographique et Forestière).

[IGN, 2011] IGN (2011). BD ORTHO v.2.0. Descriptif de contenu. Technical report. Available online at: professionnels.ign.fr. (Institut National de l'Information Géographique et Forestière).

[IGN, 2014] IGN (2014). BD TOPO v.2.1. Descriptif de contenu. Technical report. Available online at: professionnels.ign.fr. (Institut National de l'Information Géographique et Forestière).

[IGN, 2014] IGN (2014). Institut National de l'Information Géographique et Forestière. Available online at: www.ign.fr.

[Inglada et al., 2015] Inglada, J., Arias, M., Tardy, B., Hagolle, O., 1, S. V., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., and Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, 7(9):12356–12379.

[IPCC WG, 2013] IPCC WG (2013). Climate Change 2013: The Physical Science Basis. Technical report. (Intergovernmental Panel on Climate Change Working Group).

[Ippoliti et al., 2012] Ippoliti, E., Clementini, E., and Natali, S. (2012). Automatic generation of land use maps from land cover maps. pages 24–27.

[Irish, 2000] Irish, R. (2000). Landsat 7 automatic cloud cover assessment. *SPIE Proceedings Series*, 4049:348–355.

[Itzerott and Kaden, 2006] Itzerott, S. and Kaden, K. (2006). Crop classification based on spectral standard curves. In *Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover*, pages 229–238.

[Jarvis et al., 2008] Jarvis, A., Reuter, H., Nelson, A., and Guevara, E. (2008). Hole-filled seamless SRTM data V4. Available online at: srtm.csi.cgiar.org.

[Jia et al., 2014] Jia, K., Liang, S., Wei, X., Yao, Y., Su, Y., Jiang, B., and Wang, X. (2014). Land cover classification of Landsat data with phenological features extracted from time series MODIS NDVI data. *Remote Sensing*, 6(11):11518–11532.

[Jolly et al., 2005] Jolly, W., Nemani, R., and Running, S. (2005). A generalized, bioclimatic index to predict foliar phenology in response to climate. *Global Change Biology*, 11(4):619–632.

[Joly et al., 2009] Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., and Wavresky., P. (2009). Interpolation par régressions locales: application aux précipitations en France. *L'Espace Géographique*, (2):157–170.

[Joly et al., 2010] Joly, D., Brossard, T., Cardot, H., Cavailhes, J., Hilal, M., and Wavresky., P. (2010). Les types de climats en France, une construction spatiale. *Cybergeo*, (501):1–23.

[Ju and Roy, 2008] Ju, J. and Roy, D. (2008). The availability of cloud-free Landsat ETM+ data over the conterminous United States and globally. *Remote Sensing of Environment*, 112(3):1196–1211.

[Ju et al., 2010] Ju, J., Roy, D. P., Shuai, Y., and Schaaf, C. (2010). Development of an approach for generation of temporally complete daily nadir MODIS reflectance time series. *Remote Sensing of Environment*, 114(1):1–20.

[Julien et al., 2006] Julien, Y., Sobrino, J., and Verhoef, W. (2006). Changes in land surface temperatures and NDVI values over Europe between 1982 and 1999. *Remote Sensing of Environment*, 103(1):43–55.

[Julien et al., 2011] Julien, Y., Sobrino, J. A., and Jiménez-Muñoz, J.-C. (2011). Land use classification from multitemporal Landsat imagery using the Yearly Land Cover Dynamics (YLCD) method. *International Journal of Applied Earth Observation and Geoinformation*, 13(5):711–720.

[Jönsson and Eklundh, 2002] Jönsson, P. and Eklundh, L. (2002). Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, (40):1824–1832.

[Jönsson and Eklundh, 2004] Jönsson, P. and Eklundh, L. (2004). TIMESAT - a program for analysing time-series of satellite sensor data. *Computers and Geosciences*, (30):833–845.

[Kauth and Thomas, 1976] Kauth, R. and Thomas, G. (1976). The Tasseled Cap - A graphic description of the spectral-temporal development of agricultural crops as seen by Landsat. *Proceedings of the LARS Symposia*.

[Kleeschulte and Büttner, 2006] Kleeschulte, S. and Büttner, G. (2006). European land cover mapping – the CORINE experience. *North America Land Cover Summit NALCS*, pages 31–44.

[Knorn et al., 2009] Knorn, J., Rabe, A., Radeloff, V., Kuemmerle, T., Kozak, J., and Hostert, P. (2009). Land cover mapping of large areas using chain classification of neighboring Landsat satellite images. *Remote Sensing of Environment*, 113(5):957–964.

[Kuemmerle et al., 2009] Kuemmerle, T., Chaskovskyy, O., Knorn, J., Radeloff, V., Kruhlov, I., Keeton, W., and Hostert, P. (2009). Forest cover change and illegal logging in the Ukrainian Carpathians in the transition period from 1988 to 2007. *Remote Sensing of Environment*, 113(6):1194–1207.

[Kuo and Landgrebe, 2002] Kuo, B. and Landgrebe, D. (2002). A covariance estimator for small sample size classification problems and its application to feature extraction. *IEEE Transactions on Geoscience and Remote Sensing*, (40):814–819.

[Köhl et al., 2006] Köhl, M., Magnussen, S., and Marchetti, M. (2006). *Sampling Methods, Remote Sensing and GIS Multiresource Forest Inventory*. Springer-Verlag, Berlin.

[Landis and Koch, 1977] Landis, J. R. and Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, pages 159–174.

[Latham et al., 2014] Latham, J., Cumani, R., Rosati, I., and Bloise, M. (2014). Global Land Cover SHARE (GLC-SHARE) database Beta-Release Version 1.0. Rome: FAO.

[Latif et al., 2008] Latif, B., Lecerf, R., Mercier, G., and Hubert-Moy, L. (2008). Preprocessing of low-resolution time series contaminated by clouds and shadows. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2083–2096.

[Lennon et al., 2002] Lennon, M., Mercier, G., and Hubert-Moy, L. (2002). Classification of hyperspectral images with nonlinear filtering and support vector machines. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 1670–1672, Toronto.

[Li et al., 2010] Li, M., Qu, J., and Hao, X. (2010). Monitoring temporal and spatial variations of vegetation phenology from space.

[Lieth, 1974] Lieth, H. (1974). *Phenology and Seasonality Modeling (Ecological Studies Analysis and Synthesis Series)*, volume 8. Springer-Verlag, Berlin.

[Linderholm, 2006] Linderholm, H. (2006). Growing season changes in the last century. *Agricultural and Forest Meteorology*, 137(1–2):1 – 14.

[Loveland and Dwyer, 2012] Loveland, T. and Dwyer, J. (2012). Landsat: Building a strong future. *Remote Sensing of Environment*, 122:22–29.

[Lu and Weng, 2005] Lu, D. and Weng, Q. (2005). Urban classification using full spectral information of Landsat ETM Imagery in Marion County, Indiana. *Photogrammetric Engineering and Remote Sensing*, 71:1275–1284.

[MacQueen, 1967] MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1:281–297.

[Madden, 1978] Madden, H. (1978). Comments on the Savitzky-Golay convolution method for least-squares-fit smoothing and differentiation of digital data. *Analytical Chemistry*, 50(9):1383–1386.

[McFeeters, 1996] McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7):1425–1432.

[Melgani and Bruzzone, 2004] Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing*, 42(8):1778–1790.

[Menzel et al., 2006] Menzel, A., Sparks, T., Estrella, N., Koch, E., Aasa, A., Ahas, R., Alm-Kubler, K., Bissolli, P., Braslavska, O., Briede, A., Chmielewski, F., Crepinsek, Z., Curnel, Y., Dahl, A., Defila, C., Donnelly, A., Filella, Y., Jatcza, K., Mage, F., Mestre, A., Nordli, O., Penuelas, J., Pirinen, P., Remisova, V., Scheifinger, H., Striz, M., Susnik, A., van Vliet, A., Wielgolaski, F., Zach, S., and Zust, A. (2006). European phenological response to climate change matches the warming pattern. *Global Change Biology*, 12(10):1969–1976.

[Mercier and Lennon, 2003] Mercier, G. and Lennon, M. (2003). Support vector machines for hyperspectral image classification with spectral-based kernels. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 288–290, Toulouse.

[Miller et al., 2001] Miller, P., Lanier, W., and Brandt, S. (2001). Using growing degree days to predict plant stages. Technical report, Communications Services, Montana State University, Bozeman.

[Moigne, 2002] Moigne, P. L. (2002). Description de l'analyse des champs de surface sur la France par le système SAFRAN. Technical report, Centre National de Recherches Météorologiques (Météo-France).

[Mountrakis et al., 2011] Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3):247 – 259.

[Mouratidis et al., 2010] Mouratidis, A., Briole, R., and Katsambalos, K. (2010). SRTM 3" DEM (versions 1,2,3,4) validation by means of extensive kinematic GPS measurements: a case study from North Greece. *International Journal of Remote Sensing*, 31(23):6205–6222.

[Muchoney and Strahler, 2002] Muchoney, D. and Strahler, A. (2002). Pixel- and site-based calibration and validation methods for evaluating supervised classification of remotely sensed data. *Remote Sensing of Environment*, (81):290–299.

[NASA, 2014a] NASA (2014a). Aqua Project Science. Available online at: aqua.nasa.gov. (National Aeronautics and Space Administration).

[NASA, 2014b] NASA (2014b). Earth Observatory. Available online at: earthobservatory.nasa.gov. (National Aeronautics and Space Administration).

[NASA, 2014] NASA (2014). Landsat 7 Science Data User's Handbook. Available online at: landsathandbook.gsfc.nasa.gov. (National Aeronautics and Space Administration).

[NASA, 2014] NASA (2014). Visible Earth. Available online at: visibleearth.nasa.gov. (National Aeronautics and Space Administration).

[NGCC, 2010] NGCC (2010). Higher Resolution Global Land Cover Mapping. Available online at: globallandcover.com.

[Nicholas, 2004] Nicholas, A. (2004). *Rendez-vous à risques*. Belin, Paris.

[Nicoloyanni, 1990] Nicoloyanni, E. (1990). Un indice de changement diachronique appliqué à deux scènes Landsat MSS sur Athènes (Grèce). *International Journal of Remote Sensing*, 11(9):1617–1623.

[NRC, 2008] NRC (2008). *Earth Observations from Space: The First 50 Years of Scientific Achievements*. The National Academies Press, Washington DC. National Research Council.

[Ota et al., 2011] Ota, T., Mizoue, N., and Yoshida, S. (2011). Influence of using texture information in remote sensed data on the accuracy of forest type classification at different levels of spatial resolution. *Journal of Forest Research*, 16(6):432–437.

[Pagé, 2008] Pagé, C. (2008). Format des données SAFRAN et scénarios climatiques désagrégés au CERFACS. Technical Report TR/CMGC/08/27. (Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique).

[Pal, 2012] Pal, M. (2012). Advanced algorithms for land use and cover classification. In Yang, X. and Li, J., editors, *Advances in Mapping from Remote Sensor Imagery*, pages 69–90. CRC Press, Boca Raton.

[Parajka and Blöschl, 2008] Parajka, J. and Blöschl, G. (2008). Spatio-temporal combination of MODIS images – potential for snow cover mapping. *Water Resources Research*, 44(3).

[Parkinson et al.(Eds.), 2006] Parkinson et al.(Eds.) (2006). *Earth Science Reference Handbook*. NASA, Washington DC.

[Petitjean, 2012] Petitjean, F. (2012). *Dynamic time warping: apports théoriques pour l'analyse de données temporelles. Application à la classification de séries temporelles d'images satellites*. PhD thesis, Université de Strasbourg.

[Pierce et al., 2005] Pierce, K., Lookingbill, T., and Urban, D. (2005). A simple method for estimating potential relative radiation (PRR) for landscape-scale vegetation analysis. *Landscape Ecology*, 20(2):137–147.

[Poggio et al., 2012] Poggio, L., Gimona, A., and Brown, I. (2012). Spatio-temporal MODIS EVI gap filling under cloud cover: An example in Scotland. *ISPRS Journal of Photogrammetry and Remote Sensing*, 72(0):56–72.

[Pontius and Millones, 2011] Pontius, R. and Millones, M. (2011). Death to kappa: birth of quantity disagreement and allocation disagreement for accuracy assessment. *International Journal of Remote Sensing*, 32(15):4407–4429.

[Primack et al., 2009] Primack, R., Higuchi, H., and Miller-Rushing, A. (2009). The impact of climate change on cherry trees and other species in Japan. *Biological Conservation*, 142(9):1943–1949.

[Pérez-Hoyos et al., 2012] Pérez-Hoyos, A., García-Haro, F., and San-Miguel-Ayanz, J. (2012). A methodology to generate a synergetic land-cover map by fusion of different land-cover products. *International Journal of Applied Earth Observation and Geoinformation*, 19:72–87.

[Qi et al., 1994] Qi, J., Chehbouni, A., Huete, A., Kerr, Y., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, 48(2):119–126.

[Quintana-Seguí et al., 2008] Quintana-Seguí, P., Moigne, P. L., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L., and Morel, S. (2008). Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. *Journal of Applied Meteorology and Climatology*, 47:92–107.

[Reuter et al., 2007] Reuter, H., Nelson, A., and Jarvis, A. (2007). An evaluation of void filling interpolation methods for SRTM data. *International Journal of Geographic Information Science*, 9(21):983–1008.

[Richards, 2006] Richards, J. A. (2006). *Remote Sensing Digital Image Analysis: An Introduction*. Springer-Verlag, New York.

[Richardson et al., 2013] Richardson, A., Keenan, T., Migliavacca, M., Sonnentag, O., Ryu, Y., and Toomey, M. (2013). Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. *Agricultural and Forest Meteorology*, (169):156–173.

[Richardson and Wiegand, 1977] Richardson, A. and Wiegand, C. (1977). Distinguishing vegetation from soil background information. *Photogrammetric Engineering and Remote Sensing*, 43:1541–1552.

[Riggs et al., 1994] Riggs, G., Hall, D., and Salomonson, V. (1994). A snow index for the Landsat Thematic Mapper and Moderate Resolution Imaging Spectroradiometer. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 1942–1944.

[Rolland, 2003] Rolland, C. (2003). Spatial and seasonal variations of air temperature lapse rates in alpine regions. *Journal of Climate*, 16(7):1032–1046.

[Roser, 2015] Roser, M. (2015). World Population Growth. Available online at: ourworldindata.org.

[Ruffin and King, 1999] Ruffin, C. and King, R. (1999). The analysis of hyperspectral data using Savitzky-Golay filtering. Theoretical basis. In *IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 756–758, Hamburg.

[Rulloni et al., 2012] Rulloni, V., Bustos, O., and Flesia, A. (2012). Large gap imputation in remote sensed imagery of the environment. *Computional Statistics and Data Analysis*, 56(8):2388–2403.

[Rybski et al., 2011] Rybski, D., Holsten, A., and Kropp, J. (2011). Towards a unified characterization of phenological phases: Fluctuations and correlations with temperature. *Physica A: Statistical Mechanics and its Applications*, 390(4):680–688.

[Rötzer and Chmielewski, 2001] Rötzer, T. and Chmielewski, F. (2001). Phenological maps of Europe. *Climate Research*, 18:249–257.

[Savitzky and Golay, 1964] Savitzky, A. and Golay, M. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36:1627–1639.

[Scholz et al., 1979] Scholz, D., Fuhs, N., and Hixson, M. (1979). An evaluation of several different classification schemes, their parameters, and performance. In *Thirteenth International Symposium on Remote Sensing of Environment*, pages 1143–1149.

[Schwartz and Hanes, 2010] Schwartz, M. and Hanes, J. (2010). Intercomparing multiple measures of the onset of spring in eastern North America. *International Journal of Climatology*, 30(11):1614–1626.

[Sedano et al., 2011] Sedano, F., Kempeneers, P., Strobl, P., Kucera, J., Vogt, P., Seebach, L., and San-Miguel-Ayanz, J. (2011). A cloud mask methodology for high resolution remote sensing data combining information from high and medium resolution optical sensors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(5):588–596.

[Segal, 1982] Segal, D. (1982). Theoretical basis for differentiation of ferric-iron bearing minerals, using Landsat MSS data. In *Symposium for Remote Sensing of Environment, 2nd Thematic Conference on Remote Sensing for Exploratory Geology*, pages 949–951, Fort Worth.

[Shen et al., 2015] Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., and Zhang, L. (2015). Missing information reconstruction of remote sensing data: a technical review. *IEEE Geoscience and Remote Sensing Magazine*, 3(3):61–85.

[Shen et al., 2011] Shen, M., Tang, Y., Chen, J., Zhu, X., and Zheng, Y. (2011). Influences of temperature and precipitation before the growing season on spring phenology in grasslands of the central and eastern Qinghai-Tibetan Plateau. *Agricultural and Forest Meteorology*, 151(12):1711–1722.

[Shrestha et al., 2013] Shrestha, S., Asch, F., Brueck, H., Giese, M., Dusserre, J., and Ramanantsoanirina, A. (2013). Phenological responses of upland rice grown along an altitudinal gradient. *Environmental and Experimental Botany*, (89):1–10.

[Siebert and Ewert, 2012] Siebert, S. and Ewert, F. (2012). Spatio-temporal patterns of phenological development in Germany in relation to temperature and day length. *Agricultural and Forest Meteorology*, 152(1):44–57.

[Sim and Wright, 2005] Sim, J. and Wright, C. C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3):257–268.

[Srivastava et al., 2012] Srivastava, P., Han, D., Rico-Ramirez, M., Bray, M., and Islam, T. (2012). Selection of classification techniques for land use/land cover change investigation. *Advances in Space Research*, 50(9):1250–1265.

[Steinier et al., 1972] Steinier, J., Termonia, Y., and Deltour, J. (1972). Comments on smoothing and differentiation of data by simplified least square procedure. *Analytical Chemistry*, 44(11):1906–1909.

[Stone, 1977] Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics*, (5):595–645.

[Tambouratzis, 2000] Tambouratzis, T. (2000). Counter-clustering for training pattern selection. *Computer Journal*, (43):177–190.

[Tang et al., 2009] Tang, Y., Zhang, Y., Chawla, N., and Krasser, S. (2009). Svms modeling for highly imbalanced classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 39(1):281–288.

[Tateishi et al., 2014] Tateishi, R., Hoan, N. T., Kobayashi, T., Alsaaideh, B., Tana, G., and Phong, D. X. (2014). Production of global land cover data - GLCNMO2008. *Journal of Geography and Geology*, 6(3):99–122.

[Teague, 1980] Teague, M. (1980). Image analysis via the general theory of moments. *Journal of Optical Society of America*, 70(8):920–930.

[Townshend et al., 1987] Townshend, J., Justice, C., and Kalb, V. (1987). Characterization and classification of South American land cover types using satellite data. *International Journal of Remote Sensing*, (8):1189–1207.

[Tucker, 1996] Tucker, C. (1996). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2):127–150.

[Tucker et al., 1985] Tucker, C., Townshend, J., and Goff, T. (1985). African land-cover classification using satellite data. *Science*, (227):369–375.

[UNPD, 2013] UNPD (2013). World Population Prospects: The 2012 Revision, Key Findings and Advance Tables. Technical report. (United Nations Population Division).

[USGS, 2014] USGS (2014). Landsat Missions. Available online at: landsat.usgs.gov. (United States Geological Survey).

[Vapnik, 1998] Vapnik, V. (1998). *Statistical Learning Theory*. Wiley-Interscience, New York.

[Verbung et al., 2011] Verbung, P., Neumann, K., and Nol, L. (2011). Challenges in using land use and land cover data for global change studies. *Global Change Biology*, 17(2):974–989.

[Viovy et al., 1992] Viovy, N., Arino, O., and Belward, A. (1992). The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time series. *International Journal of Remote Sensing*, 13(8):1585–1590.

[Vitasse et al., 2009] Vitasse, Y., Delzon, S., Dufrêne, E., Pontailler, J., Louvet, J., Kremer, A., and Michalet, R. (2009). Leaf phenology sensitivity to temperature in European trees: do within-species populations exhibit similar responses? *Agricultural and Forest Meteorology*, 149(5):735–744.

[Walther et al., 2002] Walther, G., Post, E., Convey, P., Menzel, A., Parmesan, C., Beebee, T., Fromentin, J., Hoegh-Guldberg, O., and Bairlein, F. (2002). Ecological responses to recent climate change. *Nature*, 416(6879):389–395.

[Wang et al., 2015] Wang, J., Zhao, Y., Li, C., Yu, L., Liu, D., and Gong, P. (2015). Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250 m resolution. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103:38 – 47. Global Land Cover Mapping and Monitoring.

[Wang and He, 1990] Wang, L. and He, D. C. (1990). Texture classification using texture spectrum. *Pattern Recognition*, 23(8):905–910.

[Wang and Tenhunen, 2004] Wang, Q. and Tenhunen, J. (2004). Vegetation mapping with multitemporal NDVI in North Eastern China Transect (NECT). *International Journal of Applied Earth Observation and Geoinformation*, 6(1):17–31.

[Watmough et al., 2011] Watmough, G., Atkinson, P., and Hutton, C. (2011). A combined spectral and object-based approach to transparent cloud removal in an operational setting for Landsat ETM+. *International Journal of Applied Earth Observation and Geoinformation*, 13(2):220–227.

[White et al., 2009] White, M., deBeurs, K., Didan, D., Inouye, A., Richardson, O., Jensen, J., O'Keefe, J., Zhang, G., Nemani, R., Leeuwen, W. V., Brown, J., Wit, A. D., Schaepman, M., Lin, X., Dettinger, M., Bailey, A., Kimball, J., Schwartz, M., Baldocchi, D., Lee, J., and Lauenroth, W. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982-2006. *Global Change Biology*, 15:2335–2359.

[Wiens et al., 2009] Wiens, J., Sutter, R., Anderson, M., Blanchard, J., Barnett, A., Aguilar-Amuchastegui, N., A., C., and Laine, S. (2009). Selecting and conserving lands for biodiversity: The role of remote sensing. *Remote Sensing of Environment*, 113(7):1370–1381.

[Wulder et al., 2012] Wulder, M., Masek, J., Cohen, W., Loveland, T., and Woodcock, C. (2012). Opening the archive: How free data has enabled the science and monitoring promise of Landsat. *Remote Sensing of Environment*, 122:2–10.

[Wulder et al., 2008] Wulder, M., White, J., Goward, S., Masek, J., Irons, J., Herold, M., Cohen, W., Loveland, T., and Woodcock, C. (2008). Landsat continuity: Issues and opportunities for land cover monitoring. *Remote Sensing of Environment*, 112:955–969.

[Wulder et al., 2003] Wulder, M. A., Dechka, J. A., Gillis, M. A., Luther, J. E., Hall, R. J., Beaudoin, A., and Franklin, S. E. (2003). Operational mapping of the land cover of the forested area of Canada with Landsat data: EOSD land cover program. *The Forestry Chronicle*, 79(6):1075–1083.

[Xie et al., 2008] Xie, Y., Sha, Z., and Yu, M. (2008). Remote sensing imagery in vegetation mapping: a review. *Journal of Plant Ecology*, 1(1):9–23.

[Yu et al., 2013] Yu, L., Wang, J., and Gong, P. (2013). Improving 30 m global land-cover map FROM-GLC with time series MODIS and auxiliary data sets: a segmentation-based approach. *International Journal of Remote Sensing*, 34(16):5851–5867.

[Zha et al., 2003] Zha, Y., Gao, J., and Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3):583–594.

[Zhang et al., 2003] Zhang, X., Friedl, M., Schaaf, C., Strahler, A., Hodges, J., Gao, F., Reed, B., and Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84(3):471–475.

[Zhu et al., 2008] Zhu, Y., Chen, S., Wang, Y., and Zhu, Z. (2008). Multi-date satellite images for reducing classification confusion between deciduous forest and evergreen forest. In *Proceedings of the 2008 International Conference on Computer Science and Software Engineering CSSE*, pages 85–88.

[Zhu and Woodcock, 2012] Zhu, Z. and Woodcock, C. E. (2012). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118(0):83–94.