Monitoring global land cover fractions and change from dense satellite time series

Dainius Masiliūnas

Propositions

1. Advancements in global land cover change monitoring come primarily from improved regression- and postprocessing algorithms.
(this thesis)

2. The key constraint in global land cover analysis is the limited availability of infrastructure for processing big data.
(this thesis)

3. Ease of use determines method popularity.

4. Effective deep learning requires feature engineering.

5. Formalising teaching duties can solve PhD salary inequality.

6. All public bodies should use, teach and produce only free and open-source software.

7. Females are better programmers than males.

Propositions belonging to the thesis, entitled

Monitoring global land cover fractions and change from dense satellite time series

Dainius Masiliūnas
Wageningen, 29 May 2024

Monitoring global land cover fractions and change from dense satellite time series

Dainius Masiliūnas

Thesis committee

Promotor:
Prof. Dr M. Herold
Professor of Geo-information Science and Remote Sensing
Wageningen University & Research

Co-promotors:
Dr J. Verbesselt
Associate Professor, Laboratory of Geo-information Science and Remote Sensing
Wageningen University & Research

Dr N.-E. Tsendbazar
Assistant Professor, Laboratory of Geo-information Science and Remote Sensing
Wageningen University & Research

Other members:
Prof. Dr J. Vila-Guerau de Arellano, Wageningen University & Research
Prof. Dr E. Pebesma, University of Münster, Germany
Prof. Dr B. Demir, Technische Universität Berlin, Germany
Dr L. Parente, OpenGeoHub Foundation, Wageningen

This research was conducted under the auspices of the C.T. de Wit Graduate School of Production Ecology & Resource Conservation (PE&RC)

Monitoring global land cover fractions and change from dense satellite time series

Dainius Masiliūnas

Thesis
submitted in fulfilment of the requirements for the degree of doctor at
Wageningen University
by the authority of the Rector Magnificus
Prof. Dr C. Kroeze,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on 29 May 2024
at 1:30 p.m. in the Omnia Auditorium.

Dainius Masiliūnas
Monitoring global land cover fractions and change from dense satellite time series
182 pages.

PhD thesis, Wageningen University, Wageningen, the Netherlands (2024)
With references, with summary in English

DOI: 10.18174/649013

Summary

Land cover change is an essential process that results from a complex interaction of social, economic, policy and environmental factors and that defines how our planet is changing. Monitoring land cover change allows humans to know what happened in the past, how policies affected land, what the trends are, and to act in order to prevent undesired change from taking place. Due to its complexity, accurate monitoring of land cover change is challenging, especially at the global scale. The differences between two land cover maps are more often indicative of the differences between models and their input data, rather than actual land cover change having taken place between the two dates. Another challenge in land cover mapping is to provide a realistic and flexible representation of the variety of land cover that appears across the globe. That includes not only challenges in defining precise classes applicable across the world, but also allowing users to define classes fitting their specific needs. This could be achieved by mapping land cover fractions, indicating the proportion of area covered by each land cover class. Therefore, the goal of this thesis is to improve the methods of predicting land cover fractions and monitoring their change at a global scale, in order to aid the creation of frequently updated, comprehensive and temporally consistent global land cover products.

Chapter 1 is an introduction to the thesis, introducing the concept of land cover fraction mapping at a global scale and the challenges involved in making the maps accurate and consistent in time. This chapter covers the objective and research questions of the thesis and how they are connected to each other.

Chapter 2 focuses on machine learning methods to improve global land cover fraction map accuracy. In this chapter, a number of machine learning algorithms are compared to determine their suitability for performing land cover fraction mapping, and the map errors are analysed. In addition, a multi-step framework is proposed to deal with the zero inflation issue that is specific to the field of land cover fraction mapping, where most land cover fractions are zero in any given location. The results showed that having a separate model to determine whether a location is covered only by a single land cover class, followed by a separate classification or a regression model, results in more realistic land cover fraction maps representing the most likely land cover fractions, especially improving pure classes. Random forest regression was found to perform the best out of the tested

algorithms for land cover fraction mapping. This chapter provides a framework with which land cover fractions can be mapped with higher accuracy at a global scale.

Chapter 3 focuses on improving the consistency of multi-temporal fraction maps using a Markov chain model as a postprocessing method. This chapter compares the consistency of land cover fractions between a base Random forest regressor with no postprocessing, a recurrent Random forest implementation where the prediction of the previous year was used as input to predict the next year, and a Markov chain that attempted to minimise the differences between adjacent years and take the likely co-occurrences between classes into account. The results showed that the Markov chain approach significantly improved the accuracy and consistency of the yearly-updated land cover fraction maps compared to the base Random forest predictions. It was also superior to the recurrent Random forest approach, which tended to reduce in accuracy in later years. This chapter stresses the importance of postprocessing land cover maps to improve their temporal consistency.

Chapter 4 introduces BFAST Lite, an unsupervised change detection algorithm that builds upon the well-established BFAST family of change detection algorithms. Although the BFAST model is well-known, its application and adoption especially for large-scale and global applications has been low due to several inherent limitations. BFAST Lite was introduced as an alternative that focuses on applicability for larger areas due to improved speed and the ability to handle missing values, such as in cloudy areas. While BFAST relies on a computationally expensive iterative seasonal decomposition of time series using the LOESS technique, BFAST Lite performs change detection in a single iteration without the need of decomposition. The results showed that BFAST Lite is significantly faster than BFAST, without impacting the change detection accuracy. It also introduced additional tunable parameters and output statistics, such as breakpoint magnitude and goodness of fit. This chapter provides the land cover mapping community with a free software tool to perform unsupervised change detection at a global scale.

Chapter 5 focuses on improving temporal consistency in dense time series of land cover predictions using BFAST Lite as a postprocessing algorithm. BFAST Lite, introduced in Chapter 4, was used to further study the challenge raised in Chapter 3, at a much finer temporal scale where land cover maps are updated every one or two weeks. In this chapter, a new framework is proposed where a change detection algorithm, in this case BFAST Lite, can be used as a part of the land cover map production and updating pipeline directly. Rather than using an external univariate vegetation index as an input, BFAST Lite was shown to be capable of taking dense land cover predictions (fraction or probability) as input to produce more consistent land cover as a postprocessing technique, significantly improving the accuracy of land cover change. This chapter showcases the possibilities of postprocessing land cover products that could even be produced externally, such as Google Dynamic World, in order to create more temporally consistent land cover maps.

Chapter 6 is a synthesis of the findings of the thesis, answering the research questions posed in Chapter 1. In addition, the chapter proposes a future outlook and research directions for the field of land cover fraction and change monitoring.

The PhD thesis contributes to the field of remote sensing by exploring the techniques to improve mapping land cover fractions and monitoring their change to aid in producing up-to-date and temporally consistent land cover fraction products. It paves the way for a new generation of land cover products that can both better detect and express the diversity of land cover and its change, and serve the diverse users of these products better by providing means to customise them.

Contents

Chapter 1

Introduction

1.1 Global land cover and its characterisation

Land cover is the observed biophysical cover of the Earth's surface. It includes natural land cover classes such as forests (tree cover), grasslands (herbaceous cover), shrublands, bare soil, water bodies, wetlands, as well as artificial classes such as built-up land. Agricultural crops are usually included among land cover classes due to their great importance for mankind, although it is technically a class of land use, rather than land cover.

Characterisation and mapping of land cover is essential for many stakeholders, from international organisations to local landowners, in order to monitor and manage it in desired ways. Land cover monitoring is key for many indicators of the United Nations Sustainable Development Goals (SDGs), such as SDGs 2, 6, 9, 11, 14 and most importantly 15: Life on Land (Carter and Herold, 2019). It is an Essential Climate Variable, with great importance for the climate modelling community (Zemp, 2022). A number of Essential Biodiversity Variables (Pereira et al., 2013), that are important for monitoring changes in flora and fauna, and Essential Geodiversity Variables (Schrodt et al., 2019), that are important for monitoring the diversity of landscape forms and geology, are based on land cover products.

Many uses of land cover data require a consistent way of characterising land cover across the globe, both across space and across time. This includes defining a clear set of classes that would be applicable across the globe and using uniform methods to produce global land cover maps. Only then land cover can be comparable across countries, and be monitored through time.

Given the vast extent of the globe, the primary means of producing global land cover map products is by the use of remote sensing. Land cover is derived from imagery captured by sensors on board satellites. The most important sources of satellite data in recent years have been the European Space Agency (ESA) Copernicus programme, including its Sentinel-2 satellites with optical sensors, and the Landsat programme of National Aeronautics and Space Administration (NASA) and United States Geological Survey (USGS). Both of these programmes provide open and free access to global remote sensing data. Given the vast volumes of remote sensing data, machine learning algorithms are required to extract land cover information from the remote sensing data and to produce global land cover maps.

1.2 Land cover fractions

Most currently produced land cover maps are discrete: for each mapping unit, typically a pixel of a remote sensing image, the land cover map provides information about the dominant land cover class. This convention is simple but has drawbacks: information about non-dominant land cover classes in each mapping unit is lost, leading to an underestimation

of rare land cover classes. For example, urban areas are typically mixed with green areas (urban green), which often take a larger area than the built-up area. The dominant land cover class by fraction of area in such case would be the urban green, and the built-up area would be underestimated. Land cover maps often give priority to the built-up area class, e.g. by labelling any area with 10% or more built-up area as built-up, therefore underestimating urban green instead (Zhu et al., 2022). One way to tackle this problem, used by global land cover products such as ESA Climate Change Initiative Land Cover (CCI LC) (ESA, 2017), is to define a number of "mosaic" classes, which are a mix of other land cover classes at a specific proportion. However, this approach leads to an increased complexity of land cover map legends, limits the users to a set of predefined proportions for these mixed classes, and still result in an imprecise estimation of land cover areas.

An alternative option to land cover maps with discrete classes is land cover fraction maps. Instead of a label for the dominant land cover class (i.e. a thematic map), a land cover fraction map provides the fraction of each land cover class in each mapping unit. Therefore, land cover fraction maps are continuous maps, providing one layer per land cover class. In addition to better area estimates (Sales et al., 2022), land cover fraction maps provide a more realistic view of land cover distribution, including expressing gradients between classes. Land cover fraction maps also enable the tracking of trends in land cover, such as degradation or regrowth, and forms the basis for additional essential variables, such as the Essential Biodiversity Variable "live cover fraction" and the SDG indicator 15.4.2 "Mountain Green Cover Index". In addition, users of land cover fraction maps can create their own discrete land cover maps by defining the classes that they are interested in and at what mixture proportions.

Land cover fraction mapping is not a new concept. It dates back to the 1980s, when fuzzy set theory and linear mixture modelling provided a basis for many of the early land cover fraction mapping algorithms (Adams et al., 1995; Keller et al., 1985). Therefore, land cover fraction maps are sometimes also referred to as fuzzy land cover maps. Land cover fraction mapping started in the 1980s and gained prominence in the 1990s and early 2000s (Cannon et al., 1986; Foody, 1997; Trivedi and Bezdek, 1986; Zhang and Foody, 2001), when remote sensing imagery was at a relatively coarse spatial resolution and available computing power did not allow producing global fine spatial resolution land cover maps. Land cover fraction mapping would provide means to "unmix" coarse pixels, such as 500 m imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor, into more meaningful fractions. The MODIS Vegetation Continuous Fields product (Townshend, 1999) was a result of such effort and a pioneering global land cover fraction product. It provided a global 500 m annually updated estimate of three land cover fractions, namely, herbaceous vegetation, woody vegetation and non-vegetated.

With increasing computing power and spatial resolution of satellite sensors, focus has shifted to mapping discrete land cover classes. However, land cover fractions are recently

once again gaining prominence. A highly influential land cover fraction product has been the Global Forest Change product by the University of Maryland (Hansen et al., 2013), providing a tree cover percentage layer for the year 2000, as well as predictions of forest gain and loss areas since 2000. A recent land cover fraction product with comprehensive classes is the Copernicus Global Land Service Land Cover 100 m (CGLS-LC100) product (Buchhorn et al., 2020). It provides an annually updated map with ten land cover fraction layers from 2015 until 2019 at 100 m resolution. The follow-up project, Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring service (LCFM), aims to update the maps, including the fraction layers, annually until 2026, and provide discrete maps at 10 m spatial resolution (Joint Research Centre, 2023). The modern land cover products are produced using machine learning algorithms, currently primarily using algorithms based on decision trees and their ensembles.

1.3 Land cover change and change detection algorithms

In order to monitor land cover, it is important to have land cover maps that are regularly updated, or land cover change maps that directly indicate areas that have changed and how. While creating a map of potential change areas via the differencing method between two land cover maps from the same land cover product is a trivial computational task, this method of obtaining land cover change is discouraged by most global land cover map producers. The reason is that the change observed between two maps does not necessarily indicate actual change happening on the ground. Often the change comes from mapping errors caused by small variations in input data, seasonal and interannual variability, difference in atmospheric aerosols and even noise from the base classifier algorithm. Therefore, the differencing method produces a lot of spurious change that does not represent an actual lasting change in land cover.

Multiple approaches have been developed in order to reduce spurious change and to either produce more temporally consistent land cover maps or directly land cover change products. One of the approaches is the use of change detection algorithms. A large number of unsupervised change detection algorithms has been developed, including Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock, 2014), LandTrendr (Kennedy et al., 2010) and the Breaks For Additive Season and Trend (BFAST) family of change detection algorithms (Verbesselt et al., 2010a; Verbesselt et al., 2010b; Verbesselt et al., 2012). These algorithms work by taking a time series of one or more spectral bands of a satellite sensor or a vegetation index (VI) derived from a set of bands, and segmenting it into time periods with similar properties, such as periods of no change, gradual change or abrupt change. The segment boundaries define breaks in the time series and can then be used as indicators of potential land cover change. However, these breaks are indicative

of the change in the spectral signal of the sensor and not necessarily of land cover. For example, a grassland that experiences a burn or a flooding event would exhibit a large magnitude abrupt change in a VI time series, but the land cover class does not change because of the event; it stays a grassland. Therefore, using change detection algorithms to reduce spurious change in land cover time series is still challenging, as they tend to overestimate change as well. An example of a product that makes use of a change detection algorithm applied on a VI time series to reduce spurious change is CGLS-LC100 (Buchhorn et al., 2021).

Another prominent approach to reducing spurious change in land cover time series is the use of Markov models. Notably, the approach chosen by the team implementing the MODIS Land Cover Collection 6 product is to make use of the Hidden Markov Model (HMM) (Abercrombie and Friedl, 2016). Markov models are designed as (partially) supervised postprocessing models that, when applied to a time series of land cover, alter the values according to a set of rules and weights. Markov models can take into account both expert knowledge (e.g. that trees cannot naturally regrow in a year from bare soil), as well as training data from which generalised transition probabilities can be derived. Abercrombie and Friedl (2016) used an HMM in an unsupervised way, by setting all transition probabilities to an equal score, but using the class probabilities from a base classifier as a means of reducing change that the base classifier is not confident in. Markov models are also applied spatially, in the form of Markov Random Fields (MRFs), to include spatial correlation and distance-based rules, resulting in spatiotemporal postprocessing models (Cai et al., 2014; Li et al., 2014).

1.4 User needs and research gaps

While there has been considerable development in the field of global land cover mapping, linked to the advancement of the remote sensing field in general, many challenges still remain to be addressed. The climate community has created a list of user requirements for land cover maps to be used as an Essential Climate Variable (Zemp, 2022). They identify three key use cases for land cover maps: climate modelling, local land management and policy-making, with each use case having related but slightly different requirements. The requirements include spatial and temporal resolution, temporal extent, production lag, accuracy and stability. 100 m spatial resolution is considered sufficient for most use cases, with only the local land management requiring 30 to 10 m or finer spatial resolution. Recent global land cover maps such as ESA WorldCover (Zanaga et al., 2021) are tackling this challenge with the use of 10 m Sentinel-2 satellite imagery. In terms of temporal resolution, yearly updated global land cover maps are becoming the norm. There are first attempts at creating products at a monthly or even finer temporal resolution, such as the Google Dynamic World product (Brown et al., 2022). It also tackles the timeliness requirement by automatically producing land cover maps for each observation of Sentinel-2.

The largest gaps therefore remain in terms of temporal extent, accuracy and stability. The CCI LC map covers nearly 30 years of global land cover, but multiple use cases in Zemp (2022) call for 50 or even 100 years of land cover maps, to be produced using historical sources. In terms of accuracy, the use cases call for 65 % accuracy across all classes for their maps to be useful. This is already a difficult goal to reach, as self-reported overall accuracy of global land cover products has generally been around 75 % (Herold et al., 2016). Individual class accuracy varies and some classes, such as grassland, shrubland and wetland, tend to be difficult to classify and usually score lower in either user's or producer's accuracy (Brown et al., 2022; Tsendbazar et al., 2021b). Self-reported accuracy assessment is often done using cross-validation, and independent validation would generally result in even lower accuracy scores. The goal of 80 % accuracy, defined as the breakthrough goal in two use cases, is only reached as overall accuracy and only by a few global land cover maps, such as CGLS-LC100 and ESA WorldCover. The goal of 95 % accuracy across all classes is currently entirely unattainable outside single-class maps. Therefore, it is important to research methods to further improve the accuracy of global land cover maps.

Stability is the last user requirement, and it has even stricter requirements: only 25 % error is acceptable in terms of stability. However, stability does not have a definition that is universally accepted, with multiple definitions offered by different authors (Padilla et al., 2014; Tsendbazar et al., 2021a). The general concept, however, relates to the stability in the time series of land cover maps, and therefore relates to the accuracy of land cover change. Mapping land cover change is challenging due to the tendency of base classifiers and change detection algorithms to overestimate detected change, resulting in high commission error. As a result, it is uncommon for land cover products to provide indications of change directly. The products that do provide change focus on particular land cover classes, such as forests (Hansen et al., 2013) or water (Pekel et al., 2016). Therefore, it is important to study methods to improve the stability and temporal consistency of multi-class global land cover maps.

Another area in which more research is needed is the methods to perform land cover fraction mapping. Even though the importance and benefits of land cover fraction maps have been recognised several decades ago (Townshend, 1999), most research in the land cover domain has so far focused on discrete classes. There are no universal guidelines even on which statistics should be used to evaluate the accuracy of land cover fraction products, with several having been proposed (Foody, 1996; Silván-Cárdenas and Wang, 2008). More research is needed to determine which satellite sensors and machine learning or deep learning algorithms are most suited for the regression problem that land cover fraction mapping presents. Land cover fraction change mapping is an even less explored field, providing unique opportunities to track gradual processes such as forest degradation and regrowth. Research into methods to ensure stability and (spatio-)temporal consistency of land cover fraction maps is very important, especially for the user uptake of temporally dense products such as Google Dynamic World.

One way of improving the temporal consistency of land cover maps is to perform postprocessing of the initial land cover maps. Postprocessing techniques have been studied for discrete land cover classes (e.g. HMMs proposed by Abercrombie and Friedl (2016)), but rarely for land cover fractions. Souverijns et al. (2020) applied an HMM as postprocessing on the CGLS-LC100 fraction layers, but only to detect change at the discrete class level. Adapting postprocessing methods designed for discrete classification to work on continuous data is difficult and requires further research. Lastly, there is little research on postprocessing dense, i.e. fine temporal resolution, time series of land cover, especially land cover fractions.

1.5 Research Objectives

The goal of this thesis is to improve the methods of land cover fraction mapping to contribute towards the improvement of accuracy and temporal consistency of land cover fractions and their change monitoring. To address the overall objective, three research questions are defined, one with two sub-questions:

RQ1. How well can global land cover fraction maps be produced given various regression methods and their input data?

RQ2. How can the flexibility and scalability of change detection algorithms be improved for land cover change monitoring at the global scale?

RQ3. What postprocessing methods can be used to improve the consistency between frequently updated land cover fraction maps?

(a) To what extent can Markov chain and temporal variants of Random Forest improve the temporal consistency of yearly land cover fraction maps?

(b) What is the effect of using change detection algorithms to postprocess temporally dense time series of land cover predictions?

1.6 Thesis Overview

This thesis consists of six chapters, including this introduction chapter. The research questions of this thesis are answered in the four main chapters (Chapter 2 to Chapter 5). The relations between the research questions and the main chapters of this thesis are shown in Figure 1.1. Chapter 6 serves as a synthesis of the entire work.

Chapter 2 tackles the issue of mapping land cover fractions rather than discrete land cover classes. The chapter evaluates a number of machine learning methods for performing global land cover fraction mapping and compares their performance. This chapter also proposes a

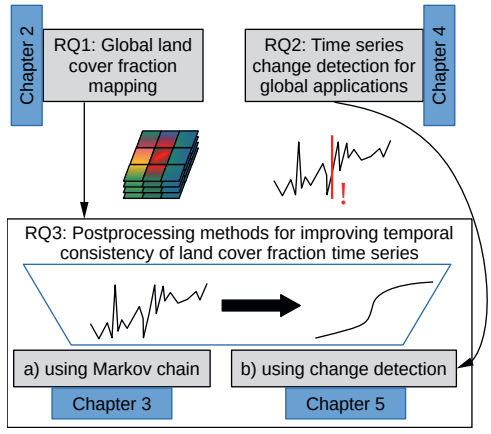


Figure 1.1: Conceptual framework of this thesis, linking the research questions with the chapters in the thesis.

novel method to deal with the zero inflation issue inherent in land cover fraction mapping, by combining regression with classification models in a multi-step approach.

Chapter 3 proposes the use of a Markov chain model to serve as a postprocessing algorithm for improving the consistency of yearly global land cover fraction maps, and compares it with two temporal variants of the Random Forest model.

Chapter 4 presents BFAST Lite, a variant of the BFAST change detection algorithm, optimised for change detection at a global scale by offering improved processing speed, tolerance for missing values in the input time series and more flexibility in its parameters.

Chapter 5 investigates the application of BFAST Lite as a postprocessing method for improving the temporal consistency of high frequency time series of land cover maps. The proposed method is tested on a time series of global land cover maps produced by both a custom Random Forest model, and the Google Dynamic World product.

Chapter 6 summarises the findings in this thesis, discusses the answers to the research questions, and provides a future outlook for the field of global land cover fraction mapping.

Chapter 2

Global land characterisation using land cover fractions at 100 m resolution

This chapter is based on:

D. Masiliūnas, N.-E. Tsendbazar, M. Herold, M. Lesiv, M. Buchhorn, and J. Verbesselt (2021a). "Global land characterisation using land cover fractions at 100 m resolution". *Remote Sensing of Environment* 259, 112409. DOI: 10.1016/j.rse.2021.112409

Supplementary material to this chapter can be found online at https://doi.org/10.1016/j.rse.2021.112409.

2

Abstract

Currently most global land cover maps are produced with discrete classes, which express the dominant land cover class in each pixel, or a combination of several classes at a predetermined ratio. In contrast, land cover fraction mapping enables expressing the proportion of each pure class in each pixel, which increases precision and reduces legend complexity. To map land cover fractions, regression rather than classification algorithms are needed, and multiple approaches are available for this task.

A major challenge for land cover fraction mapping models is data sparsity. Land cover fraction data is by its nature zero-inflated due to how common the 0% fraction is. As regression favours the mean, 0% and 100% fractions are difficult for regression models to predict accurately. We proposed a new solution by combining three models: a binary model determines whether a pixel is pure; if so, it is processed using a classification model; otherwise with a regression model.

We compared multiple regression algorithms and implemented our proposed three-step model on the algorithm with the lowest RMSE. We further evaluated the spatial and per-class accuracy of the model and demonstrated a wall-to-wall prediction of seven land cover fractions over the globe. The models were trained on over 138 000 points and validated on a separate dataset of over 20 000 points, provided by the CGLS-LC100 project. Both datasets are global and aligned with the PROBA-V 100 m UTM grid.

Results showed that the random forest regression model reached the lowest RMSE of 17.3%. Lowest MAE (7.9%) and highest overall accuracy ($72\,\% \pm 2\,\%$) was achieved using random forest with our proposed three-model approach and median vote.

This research proves that machine learning algorithms can be applied globally to map a wide variety of land cover fractions. Fraction mapping expresses land cover more precisely, and empowers users to create their own discrete maps using user-defined thresholds and rules, which enables customising the result for a diverse range of uses. The three-step approach is useful for addressing the zero-inflation issue and mapping 0% and 100% fractions more accurately, and thus has already been taken up in the operational production of global land cover fraction layers within the CGLS-LC100 project. Furthermore, this study contributes to the accuracy assessment of land cover fraction maps both thematically and spatially, and these methods could be taken up by future land cover fraction mapping efforts.

Keywords. PROBA-V, global land cover mapping, land cover fraction mapping, time series analysis, machine learning, random forest, Cubist, support vector regression, neural network, spatial accuracy, zero inflation

2.1 Introduction

Land cover, as one of the key variables for monitoring a number of SDGs, has lately received more attention due to increased availability of higher spatial and temporal resolution satellite data. In this context, the capacity for land monitoring has increased, and new global land cover maps have emerged to better map the current land cover of the world, as well as to track land cover change. Some of the recent achievements have been the CCI LC product (ESA, 2017) that provides a long-term set of consistent global annual medium-resolution land cover maps aimed at the climate community, Copernicus Global Land Service Land Cover 100 m (CGLS-LC100) product (Buchhorn et al., 2019a; Buchhorn et al., 2020) that provides a finer spatial resolution and higher quality with yearly updates since 2015, and the Finer Resolution Observation and Monitoring Global Land Cover (FROM-GLC10) product (Gong et al., 2019) that showcases the potential of land cover mapping at 10 m resolution.

Except for the cover fraction layers of the CGLS-LC100 product, all other global land cover products that include major land cover classes, such as the ones described by Arino et al. (2007), Bartholomé and Belward (2005), Chen et al. (2015), Friedl et al. (2010), and See et al. (2015), are provided with discrete classes (also known as "hard" or "crisp" classification), where each pixel of the map can only represent a single land cover class. Such discrete classification oversimplifies reality, as mixed pixels that are covered by multiple land cover classes are a common occurrence. This issue is exacerbated at coarse resolutions and over heterogeneous landscapes. It may result in biases, for instance, a sparse forest may be classified as grassland, ignoring the relatively few trees in the area, and thus underestimate tree cover in the pixel. These systematic errors then add up when scaling the result to the entire region.

A potential solution to this issue is to characterise land cover using cover fractions. In this approach, instead of a single discrete class, the proportion of every class in the legend is reported for every pixel of the map. That way, the land cover models work not on pixel labels, such as "forest", but on land cover characteristics, such as tree cover, defined as the area of the pixel covered by tree canopies, or herbaceous cover, defined as the area not covered by woody vegetation. This is also called "fuzzy" or "soft" classification, and sometimes "subpixel" mapping or "linear mixture modelling" (Okeke and Karnieli, 2006).

Land cover fraction mapping has been attempted in the past. Most of the previous research has focused on deriving land cover fractions of 3-6 classes at a local scale (Adams et al., 1995; Dwivedi et al., 2012; Foody, 1996; Gessner et al., 2013; Hansen et al., 2011; Lizarazo, 2012; Okujeni et al., 2018; Sharma et al., 2011; Uma Shankar et al., 2011; Walton, 2008), less often at a regional scale and with more detailed classes (Colditz et al., 2011). The methods for assessing the accuracy of the results vary greatly between the different studies.

In addition, global land cover fraction products have emerged, but focused on a particular class, such as tree cover (Hansen et al., 2003; Hansen et al., 2013; Townshend, 2017), water (Pekel et al., 2016; Schroeder et al., 2015) and urban area (Corbane et al., 2019; Gao and O'Neill, 2020; Gong et al., 2020). To date, only CGLS-LC100 (Buchhorn et al., 2019b) provides global maps with fractions of every major land cover class (Tsendbazar et al., 2019).

Land cover fraction mapping can be performed using a variety of different approaches and algorithms. In its core, it is a regression rather than a classification problem, as the output is a fraction of a label rather than a label itself. Methods that have been tested in previous studies include fuzzy nearest centroid regression (Zhang and Foody, 2001), spectral mixture analysis (SMA) (e.g. Adams et al., 1995; Hobbs, 2003; Shimabukuro and Smith, 1991; Yang et al., 2012), random forest (RF) regression (e.g. Walton, 2008), support vector machine (SVM) regression (Walton, 2008), Cubist regression (Walton, 2008), multi-layer perceptron (MLP) neural networks (NNs) (Zhang and Foody, 2001), genetic algorithms (Stavrakoudis et al., 2011) and wavelet transformation (Uma Shankar et al., 2011). The previous studies have only used or compared a few of these methods at once, and never with a thematically complete set of land cover classes nor at global scale.

A common issue with the use of land cover fraction data as input into regression models is data imbalance. The more classes are mapped, the more likely it is that one or more classes are not present in a given pixel (have a 0% fraction), leading to zero inflation. This is especially the case in homogeneous areas, where we can find not only an inflation of 0% fractions, but also an inflation in 100% fractions. This data imbalance leaves little training data in the middle for the regression models to learn from. Conversely, regression models tend to favour predictions closer to the mean and rarely predict extreme values. In this study, we propose the use of a hierarchical multi-step modelling approach to better predict these extreme values.

Model accuracy assessment by itself is often challenging, especially at a global scale. It requires a comprehensive dataset across the globe that would be comparable with the training data, and yet independent of it. Accuracy assessment of land cover fractions is even more challenging, as it requires a dataset that either provides fraction information, or fine spatial resolution data from which it can be calculated. Because of these challenges, the information about the accuracy of the existing global land cover fraction products is often limited, which makes it difficult for the users to decide whether a given product suits their needs. Users would also benefit from knowing the spatial variation of accuracy, as models may be more accurate at certain locations of the world and less so at others.

In this study, after investigating and comparing a variety of methods for global land cover fraction mapping, we proposed an approach to enhance land cover fraction mapping by dealing with the inherent data imbalance issue. We assessed the accuracy of the result

both thematically and spatially, by calculating model accuracy spatially across the globe. Lastly, we investigated the effect of input features on each mapped class fraction. Therefore the primary objectives of our study were to:

1. Investigate approaches for reducing bias in the model predictions with regards to zero inflation and predictions tending towards the mean.
2. Assess the accuracy of the models from a thematic and spatial point of view, comparing it to existing global land cover products.

2.2 Data and Methods

2.2.1 Reference data

The reference data (for model training and validation) used in this study was collected as part of the CGLS-LC100 project (Buchhorn et al., 2020; Tsendbazar et al., 2019). The data includes over 150 000 training points and over 21 000 validation points across the globe, describing the fractions of 12 classes in the year 2015. The classification scheme follows the United Nations Land Cover Classification Scheme (LCCS) (Buchhorn et al., 2020). However, due to the limited number of observations for some rare classes, we merged them to get a total of seven: bare land (including snow and ice), cropland (including shifting cultivation), herbaceous vegetation (including wetland, lichen and moss), shrubs, trees, built-up and inland water. The "unknown" class was discarded: points with the dominant land cover class marked as unknown were not used. For points with a minority fraction of unknown, the remaining classes were linearly rescaled to add up to 100%. Thus in the end the training dataset size became 138 164 and the validation dataset size became 20 705. See figure 2.1 for the spatial distribution of the points.

These global datasets were generated by performing high-resolution satellite imagery interpretation by a team of experts, using the GeoWiki platform. Each sampled point corresponded to a single 100 m by 100 m pixel of the PROBA-V 100 m UTM grid (Buchhorn et al., 2020). The area of each of these sampled pixels was subdivided into 100 subpixels at 10 m by 10 m spatial resolution. The subpixels were labelled by the experts, and then converted into land cover fraction estimates by calculating the proportion of subpixels that each land cover class covers in the pixel. The satellite imagery that the experts interpreted corresponded to the year 2015.

The training set was generated by a team at the International Institute for Applied Systems Analysis (IIASA), whereas the validation dataset was generated by a team at Wageningen University & Research (WUR). The validation dataset was first developed over Africa (Tsendbazar et al., 2018), and was later expanded to cover the whole world (Buchhorn et al., 2020). The class definitions and tools used to collect the data was equivalent for both datasets, but performed independently by a separate group of regional experts to

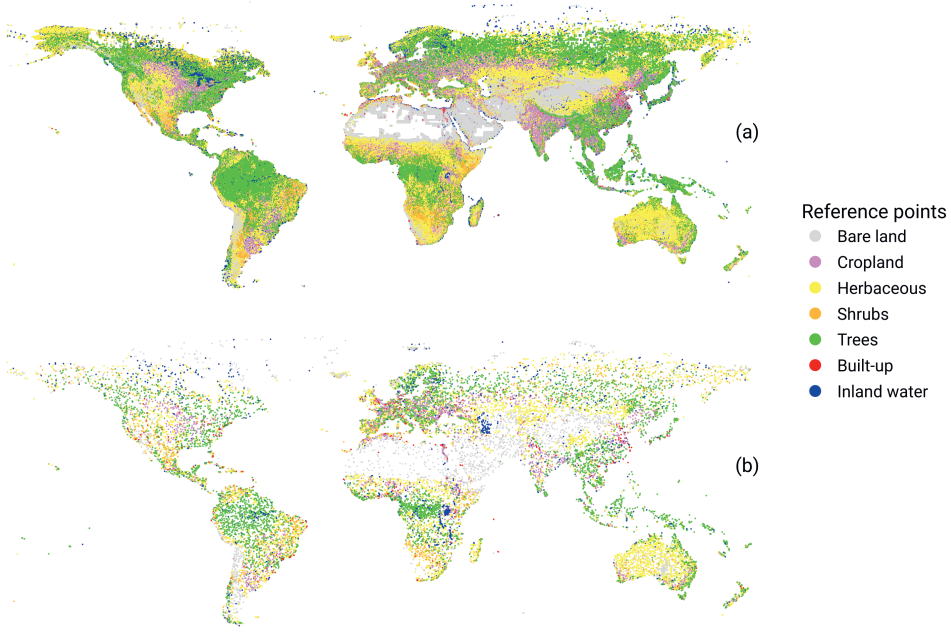


Figure 2.1: Sample points representing 100 m × 100 m areas at which land cover reference data was used in the study. The colours represent the dominant land cover class at each point. (a): training dataset, collected by IIASA, 138 164 points used in this study. (b): validation dataset, collected by WUR, 20 705 points used in this study. Both datasets were collected as part of the CGLS-LC100 project (Buchhorn et al., 2020).

ensure independence of the data, and using a different sampling method. The validation dataset was generated separately using stratified random sampling and follows the CEOS Land Product Validation guidelines, which focus on independent and statistically rigorous accuracy assessment. In contrast, the training dataset uses a mix of systematic sampling and additional random sampling in areas that exhibit higher heterogeneity, so as to provide the algorithms with enough data to train in all areas of the world.

2.2.2 Model training features

See figure 2.2 for an overview of the whole processing chain used in this study. The processing was carried out in R (R Core Team, 2021) and the resulting code has been made openly available in Masiliūnas (2020).

To train the models and predict land cover fractions in unsampled locations, six groups of features were used: vegetation indices, temporal metrics, terrain metrics, soil metrics,

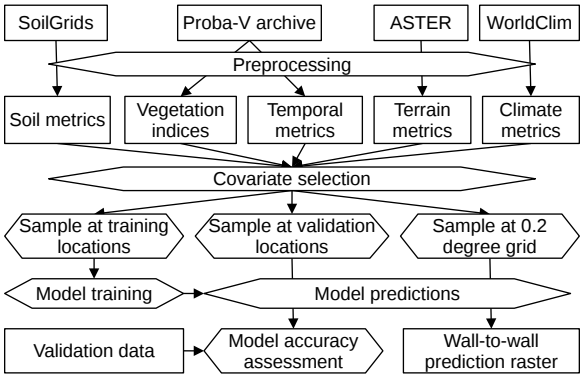


Figure 2.2: Processing workflow, from the raw input data to model accuracy assessment and wall-to-wall map output.

climate metrics and location data (see Appendix 2.A, table 2.A.1). These features had to be preprocessed before they could be input into the models.

We chose to use the entire archive (2014-03-11 to 2019-07-16) of the PROBA-V 100 m Level 3 Top-of-Canopy 5-day composite product (Dierckx et al., 2014; Wolters et al., 2016) for this study. The PROBA-V archive provides a relatively long history of frequent (daily or 2-day) observations, which is beneficial for time series analysis, as there are more observations of the land surface in cloudy areas, and a dense time series of observations can be acquired to generate robust temporal metrics. While the reference data corresponded to the land cover at the year 2015 specifically, we chose to use the whole time series of PROBA-V data to obtain more robust statistics for the temporal metrics. The long time series makes the temporal outlier filtering step more reliable, increases the robustness of the fitted harmonic model, and removes the effects of interannual variability and seasonality in the calculated descriptive statistics, as described in the following sections. Land cover change is a relatively rare phenomenon, according to tests performed in the making of CGLS-LC100 collection 3, and so we expected that the difference caused by land cover change to the input data would be smaller than the margin of error of the model output. Nevertheless, this may result in poorer model performance compared to if up-to-date reference data for each year would be available.

We masked out the clouds from the time series of each of the four PROBA-V spectral bands, first by applying the status mask provided with the product itself, and then by running a temporal cloud filter to remove the remaining outliers that were further than 2 standard deviations away from the locally estimated scatterplot smoothing (LOESS) curve fitted over the blue reflectance band. We then used the resulting cloud-free time series to generate the following vegetation indices (VIs): Normalised Difference Vegetation Index (NDVI), Normalised Difference Moisture Index (NDMI), Enhanced Vegetation Index

(EVI), Optimised Soil-Adjusted Vegetation Index (OSAVI) and near infra-red of vegetation (NIR_v). These VIs are commonly used in land cover mapping to aid in discerning the vegetation classes among each other. Next, we used the VI time series to calculate the descriptive statistics (median and interquartile range) over the whole time series, as well as for each phenological season separately. We included the resulting metrics as vegetation index features. Next, we ran harmonic analysis (Jakubauskas et al., 2001) on NDVI in order to decompose the time series into sine and cosine components for two frequency orders (annual and semiannual), as well as the trend and intercept of the model. The temporal metrics from the harmonic model quantify the seasonality of the area and allow differentiating between vegetation with different seasonality, such as crops. From this harmonic analysis we obtained the minimum and maximum values of the NDVI time series, which we included as additional vegetation index features. We also obtained the the trend and intercept components from the harmonic model. Lastly, we calculated the phase and amplitude for the two harmonic orders from the respective sine and cosine components of the model. We used these sine, cosine, trend and intercept components, as well as the phase and amplitude of the two harmonic orders, as temporal features in our study. For a list of features that were ultimately used as input to the models, see Appendix 2.A, table 2.A.1.

To generate elevation features, we obtained the ASTER GDEM v003 (NASA et al., 2019) product (30 m) and resampled it to the PROBA-V 100 m grid. We used the result directly as the terrain elevation feature. In addition, we used the Geospatial Data Abstraction Library (GDAL) (GDAL/OGR contributors, 2020) to calculate terrain parameters out of elevation: slope, aspect and Terrain Position Index (TPI).

We chose the WorldClim 2.0 30 second product (Fick and Hijmans, 2017) as a source of climate features. It includes monthly temperature, precipitation, solar radiation, wind speed and water vapour pressure data. In addition to these features, we calculated 19 bioclimatic parameters from the data, using the `dismo` package (Hijmans et al., 2017). We also calculated some additional biophysical parameters, namely all of the climate variables during the coldest, warmest, driest and wettest months of the year at each location, as well as the yearly averages of the climate variables.

We used SoilGrids (Hengl et al., 2017) to obtain features related to soils. SoilGrids is based on a random forest model that predicts soil properties at various soil depths globally at a 250 m resolution. In the creation of SoilGrids, a land cover map (based on MODIS) had been used, and so, in order to avoid circular inference, the features that are significantly influenced by the land cover map as detailed in Hengl et al. (2017) were excluded. The soil taxonomy features were also excluded, since they are categorical derivatives from the numerical soil property data and thus do not contribute to land cover fraction prediction.

Lastly, we also included the latitude, longitude and absolute latitude of the reference points as location features when training the models, so that the models could learn spatial patterns.

2.2.3 Feature selection

In total, we generated 313 features in preprocessing. However, many of them were collinear with one another. Multicollinearity prolongs training time for machine learning models and leads to unreliable coefficient estimation and increased error variance for linear models. Thus, we employed variable selection to remove collinear features before predicting the land cover fractions. Features that had a Pearson's correlation coefficient r (Pearson, 1895) above 0.9 were excluded in an iterative process. After that, features with a Spearman's rank correlation ρ (Spearman, 1904) above 0.9 were likewise excluded.

We manually selected the features to exclude, to avoid interpretation difficulties that arise from automatic selection procedures. The majority of the collinear features were soil metrics predicted at different depths. Therefore, we left in the 10 cm depth features, and excluded the other depths, as long as r was above 0.9. Similarly, climate data were collinear between subsequent months. Thus January and July data were preferred, as these months represent different extremes of the year and are less correlated with data of the other months. Our initial correlation analysis also showed that the spectral bands of PROBA-V were highly correlated with each other as well as to VIs, so we only used VIs as features.

After the feature selection process, 67 features remained. These features include data from each of the feature categories. See Appendix 2.A, table 2.A.1 for an overview of all of the features that remained and thus used in model training and prediction.

2.2.4 Land cover fraction mapping methods

We compared a wide array of machine learning regression methods for deriving land cover fractions. The tested methods can be broadly divided into three types: linear models, machine learning models based on classification and regression trees (CARTs), and machine learning models not based on CARTs. See table 2.1 for the full list of methods that we compared in this study. In addition, as a baseline we also compared the results with a trivial equal proportion model (all fractions always predicted to be equal, namely $\frac{100\%}{7} \approx 14.29\%$). Any useful model has to perform better than the equal proportion model, and by comparison to it, it is possible to tell how much better a model performs in estimating the fraction of each class. We tuned each algorithm to select optimal parameters, and postprocessed the output of each algorithm as necessary to ensure that all land cover fractions in each pixel add up to 100%. Namely, if the model output for any class was outside of the 0-100% range, the values were clamped to that range, and if the values did not add up to 100%, they were linearly rescaled so that they

Table 2.1: List of regression methods for land cover fraction estimation tested in this study.

Category	Name	Reference	R package and authors
Linear models	Fuzzy nearest centroid (FNC)	Keller et al. 1985	`GSIF` (Hengl et al., 2004)
	General linear regression model (GLM)	Neter et al. 1996	stats (R Core Team, 2021)
	Partial least squares (PLS) regression	Wold et al. 2001	`pls` (Mevik et al., 2016)
	Lasso regression	Tibshirani 1996	`glmnet` (Friedman et al., 2010)
	Multinomial logistic regression (MLR)	Theil 1969	`nnet` (Venables and Ripley, 2002)
Machine learning models based on decision trees	Random forest (RF) regression	Breiman 2001	`ranger` (Wright and Ziegler, 2017)
	Cubist regression	Quinlan 1992	`Cubist` (Kuhn and Quinlan, 2020)
Other machine learning models	MLP neural networks (NNs)	Dreyfus 1990	`keras` (Allaire and Chollet, 2018)
	Support vector machine (SVM) regression	Suykens and Vandewalle 1999	`liquidSVM` (Steinwart and Thomann, 2017)
Ensemble learning	Super Learner	Van der Laan et al. 2007	`sl3` (Coyle et al., 2020)

would. All of the model building and data analysis was performed using the free and open-source statistical software R (R Core Team, 2021). The results showed that RF regression performed the best, with the lowest root mean squared error (RMSE) value of 17.3%. For more technical details and in-depth results of the regression model accuracy comparison, see Appendix 2.B.

2.2.5 Multi-step approach to account for zero inflation

As RF regression performed the best by RMSE in the regression model comparison, we tested whether it could be improved further by attempting to solve the training dataset balance issue, namely the high frequency of 0% and 100% fractions. As the dataset describes fractions of each land cover class at each point, most of the locations consist of a mix of only a few classes, and the fraction of the rest of the classes is zero at that location. If the pixel is pure, then one land cover class will be 100% and the rest 0%, which is also a common case. This leads to the dataset getting dominated by zeroes. In that case, minimising the objective function of a machine learning model leads to prioritising predicting 0% fractions well, and ignoring the prediction errors in the middle of the 0-100% range. This is not desirable for users of land cover fraction data, as the added value of fraction information is the information about the middle of the range; otherwise, discrete classification would be just as good. Conversely, 0% by itself is rarely predicted precisely, because when the value is uncertain, predictions tend towards the mean. Therefore we tried several approaches to deal with data imbalance by employing a hierarchical combination of machine learning models.

We compared three approaches using RF models: (a): a single regression model trained on all data. (b): a two-model approach in two steps. Step 1: a binary classification model for each class to predict zeroes, trained on a generated dataset that, based on the land cover fraction values, had labels “zero” and “non-zero”. Step 2: a model to predict non-zeroes, trained on all of the non-zero fraction values. For the combined prediction, all

points that were predicted as "zero" in step 1 were set to 0%, otherwise the value from the model in step 2 was used. (c): a three-model approach using three steps. Step 1: a binary classification model to predict pixel purity (i.e. whether we face a classification or a regression problem), trained on a generated dataset that had labels "pure" for points that had a single land cover fraction above a purity threshold, e.g. 95%, and "non-pure" for points that do not. Step 2: a model to perform regression on mixed pixels (as determined in step 1). Step 3: a model to perform classification on pure pixels (as determined in step 1), resulting in a prediction of 100% fraction of the predicted discrete class and 0% fractions for all other classes. The combined prediction is the combination of the results of steps 2 and 3, as determined by step 1.

For the three-step approach, we also tested the effect of the fraction threshold for when we consider a pixel "pure". The lower the threshold, the more pixels are considered pure and the more often the classification model will be selected, as opposed to the regression model. We also evaluated the accuracy metrics of the separate steps of the multi-step models.

Lastly, we compared the results of our proposed multi-step approach with an approach that uses the median for ensembling tree votes, instead of the mean. The median vote leads to predicting the extreme fractions of 0% and 100% more often, since if the majority of the decision trees vote for one of the two extremes, it gets selected as the output value. Finally, we also investigated the combination of both approaches.

2.2.6 Accuracy assessment

To assess the performance of the models, we used a number of statistical measures. We started with the statistics of assessing land cover fraction model accuracy that are the most commonly used in this field: root mean squared error (RMSE), mean absolute error (MAE) and mean error (ME). MAE represents the average difference between the predicted and the reference land cover fractions. In our case, its unit is percentage points. RMSE squares the errors, therefore giving a larger penalty for large errors, and thus is always higher than MAE. These statistics are relatively straightforward to calculate and interpret. In the case of land cover fraction mapping, RMSE is very sensitive to errors where a pixel is entirely mapped as a different class (i.e. 100% instead of 0%). MAE is more lenient and not as influenced by a small number of such large misclassifications. Thus it is more indicative of the overall model accuracy, whereas RMSE is more indicative of the presence of large errors.

We calculated RMSE, MAE and ME, both separately per class, and also pooled overall. These overall measures were calculated by taking the mean of all class points pooled together, rather than taking a mean of the per-class means. In addition, we calculated the relative root mean squared error (RRMSE), relative mean absolute error (RMAE) and relative mean error (RME) for each class by dividing the absolute measures by the mean

fraction of each class. The relative statistics give an extra penalty for poor predictions of rare land cover fractions (i.e. those that are absent from most pixels), to account for the issue that a prediction of constant 0% would lead to low RMSE and MAE for rare class fractions.

Next, we estimated the goodness of fit of the models by calculating the coefficient of determination R^2 of the models in two ways. The first way is the Nash–Sutcliffe model efficiency coefficient (NSE) (Nash and Sutcliffe, 1970), which is equivalent to an R^2 of a linear regression model whose intercept and slope terms are predetermined and are equal to 0 and 1, respectively. This metric shows how far away the predicted values are from the 1:1 line with the reference values. The value range of NSE is $(-\infty, 1]$, where a value of 0 means that the predicted values are no better than predicting the mean value. The second way is to calculate the R^2 of an ordinary least squares (OLS) regression that estimates the intercept and slope, rather than presetting it. This method always gives higher R^2 values than NSE, as it allows the regression line to have more flexibility. The two ways of estimating the coefficient of determination are herewith represented as R^2_{NSE} and R^2_{OLS}, respectively.

Furthermore, we calculated a subpixel confusion-uncertainty matrix (SCM) (Silván-Cárdenas and Wang, 2008), and the metrics derived from it: overall accuracy (OA), as well as producer accuracy (PA) and user accuracy (UA) per class. The SCM is an adaptation of the confusion matrix concept to fractional data. We used the MIN-PROD composite operator as recommended by Silván-Cárdenas and Wang (2008). When using this operator, the diagonal of the matrix expresses the maximum overlap (agreement) of the target and predicted class fractions, and the off-diagonal is an expression of which classes the non-overlapping fractions should belong to by calculating the expected value of overlap (product of reference and predicted class fraction). For instance, a pixel of 60% grass and 40% shrub, when predicted as 40% grass and 60% shrub, would have $min(60\%, 40\%) = 40\%$ in the diagonals and $\frac{20\% \times 20\%}{20\%} = 20\%$ in the off-diagonals. For cases where the allocation of the overestimated and underestimated parts of the fraction do not have a unique solution, the SCM indicates the expected value and an associated uncertainty measure.

Moreover, to show the variation in predicted land cover fraction values depending on the magnitude of the fractions, we produced boxplots showing the distribution of the predictions. The boxplots are binned for each 10% of the predictions. The 1:1 line indicates how well do the distributions of the predictions and the reference data match. We also created additional plots showing how RMSE, MAE and ME change over these bins.

To show how the model accuracy varies in space, we produced spatial residuals, i.e. the overestimation and underestimation of each class fraction for each validation point. We rasterised the result into global maps for ease of view. If multiple points fell into the same raster cell, we reported the mean value.

To put our results in a wider context, we compared them with the currently available global products for specific land cover classes. For this comparison, we used the same validation dataset as for our models. Since the validation dataset reflects land cover of the year 2015, we chose to compare our results with four products produced for or around the year 2015. Therefore we chose the Global Forest Cover Change (GFCC) forest cover product for 2015 (Townshend, 2017), the Global Surface Water (GSW) water occurrence history product for 2015 (Pekel et al., 2016), and two products for the comparison with the built-up class: Global Human Settlement Layer (GHSL) built-up for 2014 (Corbane et al., 2019) and FROM-GLC10 impervious surface change layer for 2015 (Gong et al., 2020). As the latter two products show the change in built-up cover over time, we reclassified them to have a value of 100 if the area was built-up in 2014 or 2015, respectively. For FROM-GLC10, we assumed that areas that had been covered by impervious surface at any time prior to 2015 continued to be covered by impervious surface in 2015. GSW data was also reclassified to 100 if the water was permanently present and 50 if it was seasonally present. Next, for validation purposes we resampled each product to the PROBA-V 100 m grid using the bilinear resampling method, so that the cover fractions get aggregated over the same areas as our own data. Then we compared the product values with our validation data point values.

Lastly, we investigated how features affect model accuracy. See Appendix 2.C for more information on feature importance.

2.3 Results

2.3.1 Zero inflation adjustment with multi-step models and median voting

Regression model comparison showed that RF regression achieved the highest accuracy: by RMSE when using a single model and a mean vote (RMSE: 17.3%, MAE: 9.4%), and by MAE when using a median vote (RMSE: 20.7%, MAE: 7.9%). Therefore we chose RF regression to test our proposed multi-step approach. For more details about the results of the accuracy comparison between the tested regression models, see Appendix 2.B.

The overall accuracy statistics of all of the RF regression models that we tested, including two-step, three-step, as well as median voting approaches and their combinations, can be seen in table 2.2. Per-class results can be seen in figure 2.3 and the relative statistics in figure 2.4.

The median voting approach resulted in significantly more predictions of 0% and 100% class fractions, therefore the output looked closer to a discrete classification map compared to mean voting. Since these two most common values were predicted precisely much more often, MAE reduced. However, in cases of high uncertainty, the median vote was also

Table 2.2: Accuracy statistics of multi-step models. Best performing statistics are highlighted. "slope" refers to the OLS-estimated slope, "int" refers to the OLS-estimated intercept.

Model	RMSE (%)	MAE (%)	R^2_{NSE}	R^2_{OLS} (slope/int)	OA (%)	Kappa
RF regression	**17.3**	9.4	**0.66**	**0.67** (1.09/-1.22)	67 ± 4	0.57 ± 0.05
″ two-step	19.9	8.2	0.56	0.60 (0.78/3.48)	71 ± 2	0.62 ± 0.02
″ three-step	19.4	8.4	0.58	0.60 (0.84/2.34)	71 ± 3	0.62 ± 0.04
″ median vote	20.7	**7.9**	0.52	0.60 (0.74/3.80)	71 ± 1	0.63 ± 0.02
″ ″ two-step	20.0	8.1	0.54	0.60 (0.77/3.67)	$\mathbf{72 \pm 1}$	0.63 ± 0.02
″ ″ three-step	20.2	**7.9**	0.54	0.60 (0.77/3.34)	$\mathbf{72 \pm 2}$	$\mathbf{0.64 \pm 0.02}$

much more likely to predict a 100% fraction compared to the mean vote, and thus was also more likely to predict 100% of the wrong class. This resulted in increased RMSE. Therefore the overall effect of using the median vote is a change in the balance of RMSE and MAE.

We observed a similar effect when using the multi-step approach for an RF regression that uses the mean statistic for ensembling the tree votes. Using a two-step approach decreased the MAE of the model, while increasing RMSE. This is because the false positives in the first step of the two-step model (the binary zero/non-zero classification) result in some high fractions nevertheless being predicted as zero. The two-step approach combined with binary relevance also had a drawback: in some cases all land cover class fractions were predicted to be 0% in the first step, making it impossible to determine the land cover fractions. Therefore, to make them sum up to 100%, we set these cases to equal proportions (100%/7), which introduced further error.

The three-step approach solved this issue, since the first step predicts purity, rather than zeroes. In this approach, pure pixels are passed to RF classification, which always predicts the most likely pure class. When the three-step approach was applied to mean vote RF, the result was a slight decrease in RMSE across most classes compared to the two-step approach. However, there was also an increase in MAE of the predicted crop cover. Therefore the three-step model offsets the increase in RMSE as seen in the two-step model case, and does not cause a high increase in MAE, thus leading to a better balance between the two measures.

The median vote had a stronger effect on decreasing MAE (and increasing RMSE) compared to the multi-step approach with mean voting. When both approaches were combined, no further decrease in MAE could be achieved, however, the combination of the three-step model and median voting decreased RMSE compared to the single-step median model, therefore reducing large errors.

We also investigated the separate steps of the multi-step model in more detail to better understand the accuracy of each model. One parameter in the three-step model is the

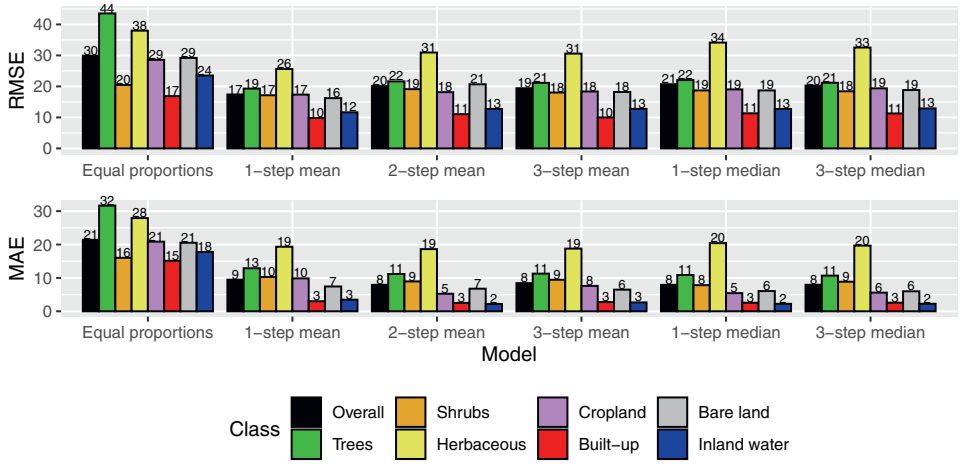


Figure 2.3: Comparison of RF regression models (equal proportion model shown as a reference). 1-step models use no adjustment for zero inflation, 2-step models perform classification on zeroes and regression for non-zeroes, 3-step models perform a classification into pure and non-pure pixels, and then a regression or classification based on that. Mean and median are the tree vote summary statistics.

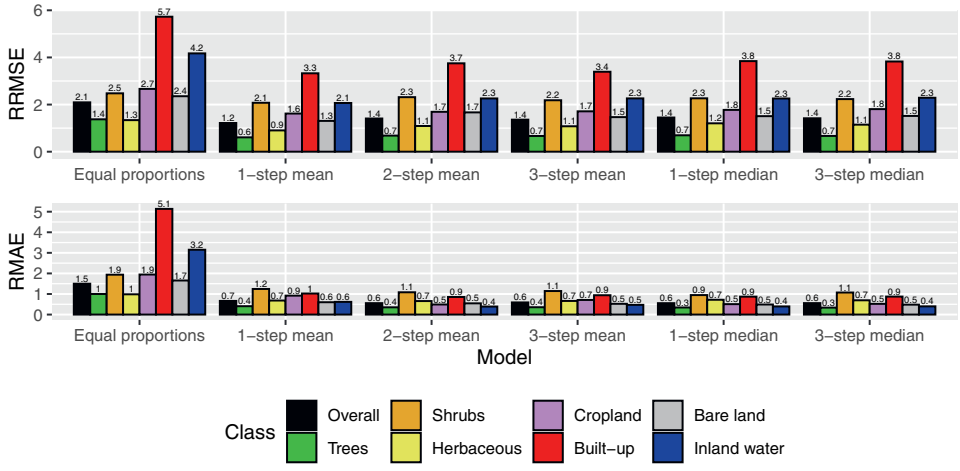


Figure 2.4: Comparison of relative accuracy metrics of RF regression models (equal proportion model shown as a reference). 1-step models use no adjustment for zero inflation, 2-step models perform classification on zeroes and regression for non-zeroes, 3-step models perform a classification into pure and non-pure pixels, and then a regression or classification based on that. Mean and median are the tree vote summary statistics.

purity threshold: how high should the cover fraction be for a pixel to be considered "pure" and be subject to classification rather than regression. Decreasing the purity threshold led to a result closer to the one-step model, i.e. reduced RMSE and increased MAE, as the classification model was used less often. The purity binary classifier, when only 100% cover is considered "pure", achieved 78% overall accuracy, and the accuracy decreased when the purity threshold was decreased. The classification model achieved 87% overall accuracy, but shrub and built-up classes had very low users' accuracy. These classes are highly heterogeneous, therefore there were too few observations to train the classifier to identify these classes. Decreasing the purity threshold resulted in a lower overall accuracy, but an increase in the users' accuracy in the classifier model for these particular classes. The regression step by itself had lower accuracy than the combined three-step model, with 21.38% RMSE and 10.49% MAE (using median voting), as the middle of the range is the most difficult to predict correctly. Decreasing the purity threshold led to a lower RMSE, as the value range of the regressor training data gets decreased, but a higher MAE, as the amount of training data for the regressor decreased. Overall for the whole multi-step model, lowering the purity threshold increased RMSE and slightly increased MAE as well.

All in all, both median voting and the multi-step approach successfully result in more correctly predicted 0% and 100% fractions, thus lowering MAE and increasing the SCM OA. While it also increases RMSE, combining the two concepts together leads to a lower increase in RMSE.

2.3.2 Spatial predictions and accuracy

We selected the three-step median vote RF model for further analysis, as it represents a good balance between the RMSE and MAE statistics, and its analysis helps further understand the three-step model. To visually demonstrate the model, we used it to predict land cover fractions at a global scale (100 m resolution, but sampled every 0.2 degrees). See figure 2.5 for a visualisation of all of the fraction layers separately, and the supplementary material for the output GeoTIFF file itself. The wall-to-wall fraction maps reveal how land cover fraction mapping is capable of expressing gradients and mixed land cover. In addition, the spatial accuracy maps that we produced based on the model predictions show the variation in the model accuracy globally, and are also presented in raster format in figure 2.5.

Biotic gradients can be seen in the global patterns of the land cover fractions. For instance, gradients are visible between communities dominated by shrubs and ones dominated by herbaceous vegetation, such as in south and east Africa. Likewise, the gradient of tree cover from 100% in the African tropics to 0% in the sub-Sahara region is evident. The tropical forest edge appears with a hard edge when using median voting and the three-step approach, as forest occurs in discrete patches due to human activity, rather than changing

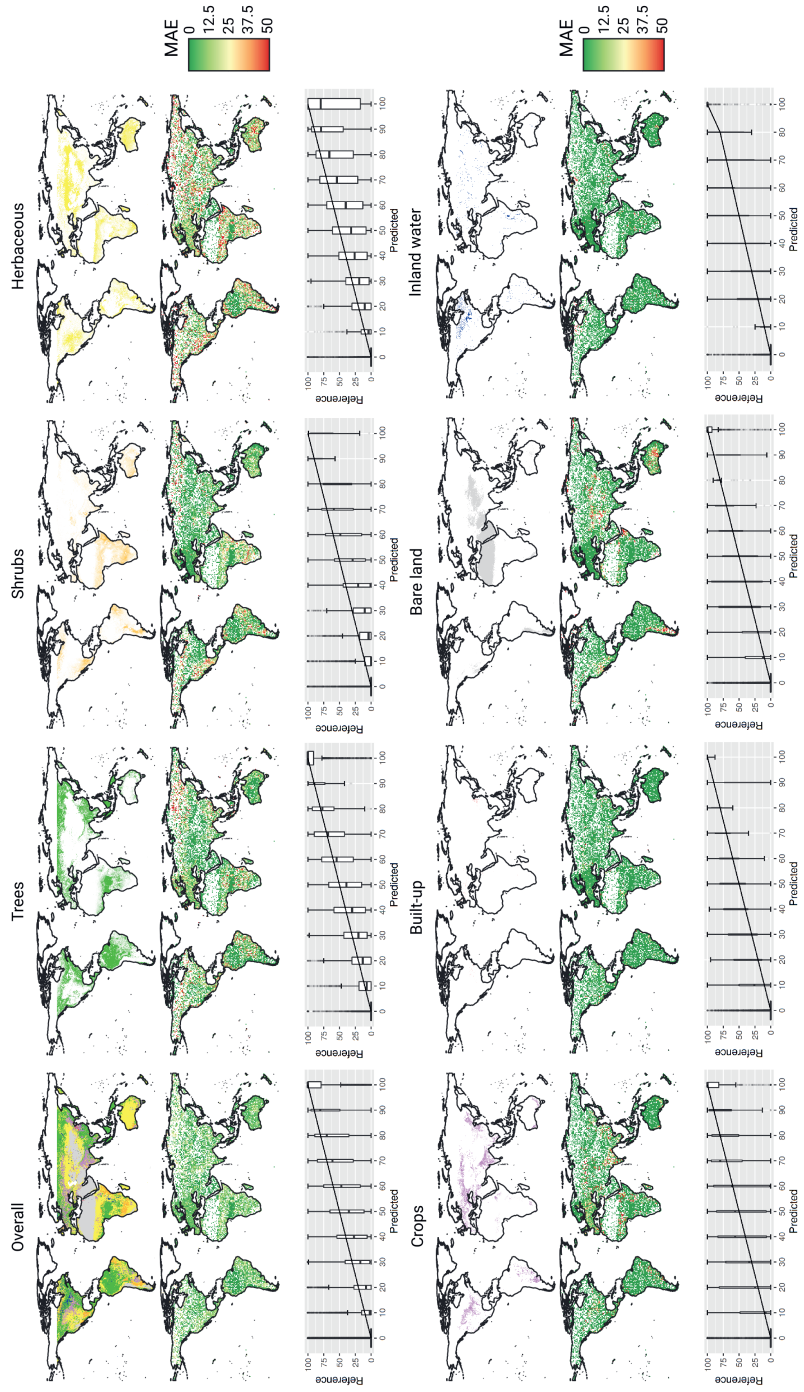


Figure 2.5: Random Forest single model predictions per class. Top row: predicted fractions, each given its own colour ("Overall" shows a hardened map of dominant land cover using the same colour scheme). Middle row: absolute errors per class, based on predictions at point locations of validation data, displayed aggregated over 1 by 1 degree cells using the mean function. Bottom row: distribution of predicted versus true values, shown as box plots with bins each 10%, widths representing sample size.

gradually over space. The tree cover in the transition zone towards savannah is much more mixed and gradual. Herbaceous cover in sub-Sahara shows an asymmetric gradient: the cover is highest at around 14-15° N and decreases quickly towards the north, becoming zero around 18° N; but decreases slowly towards the south, reaching all the way to 5° N. Inland water shows up as more discrete, as it naturally forms discrete patches. Mixed pixels that include water are uncommon. Built-up area is also relatively rare worldwide. It rarely forms a 100% fraction, as urban areas tend to include both built-up area and greenery within the footprint of the 100 m pixel.

The spatial pattern of map accuracy shows that the land cover fractions in areas with pure land cover, such as tropical forests (100% tree cover) and deserts (100% bare land), were predicted with the highest accuracy. Conversely, fractions in areas with mixed land cover were predicted less accurately, as isolating individual fractions from pixel-level information is more challenging. In addition, land cover fractions in the extreme latitudes were predicted less accurately as well. In these areas, less training data was available, owing to the lack of high-resolution imagery there.

As we can see from the box widths of the boxplots in figure 2.5, the distribution of the predictions was relatively even across the whole range for herbaceous vegetation and trees, but uneven for bare land, inland water and built-up area. The number of predictions was much more even across the entire range for herbaceous vegetation compared to shrubs. The model overestimated the fractions of trees and herbaceous cover, as the medians of each box are below the 1:1 line almost throughout the entire range, but it underestimated built-up and crop fractions. The overall ME is below 0.001, which means that overall the model is not biased. For an alternative visualisation showing the change of RMSE, MAE and ME overall and per class over each prediction bin, see Appendix 2.D, figure 2.D.1.

Our analysis of the effect of training features on model prediction accuracy showed that features obtained from remote sensing were the most important, but all of the feature categories contributed to improving the model predictions. For more detailed information, see Appendix 2.C.

Lastly, the results of the comparison of the RF three-step median model predictions with existing global products that correspond to a particular land cover class in our classification are given in table 2.3. The RF three-step median model performed worse than the GSW water occurrence history product, but better than the GFCC tree cover and GHSL built-up products. There was a tie when comparing the RF three-step median model with the FROM-GLC10 impervious surface product when it comes to MAE, but FROM-GLC10 was slightly more accurate according to RMSE and less biased according to ME.

Table 2.3: Accuracy comparison between our results and existing global land cover fraction maps. Highest accuracy and lowest bias results are highlighted.

Fraction and source	RMSE (%)	MAE (%)	ME (%)	RRMSE	RMAE	RME	R^2_{NSE}	R^2_{OLS}
Inland water (RF 3-step median)	12.90	2.25	-1.17	2.29	0.40	-0.21	0.65	0.67
GSW, year 2015 (Pekel et al., 2016)	**10.22**	**1.94**	**-0.63**	**1.81**	**0.34**	**-0.11**	**0.78**	**0.78**
Trees (RF 3-step median)	**21.18**	**10.68**	**1.81**	**0.67**	**0.34**	**0.06**	**0.72**	**0.74**
GFCC, epoch 2015 (Townshend, 2017)	28.80	18.33	-12.57	0.91	0.58	-0.40	0.48	0.61
Built-up (RF 3-step median)	11.28	**2.57**	-2.33	3.83	**0.87**	-0.79	0.19	0.22
GHSL built-up, 2014 (Corbane et al., 2018)	18.67	5.43	4.80	6.34	1.84	1.63	-1.23	0.52
FROM-GLC10 impervious surface, 2015 (Gong et al., 2020)	**11.18**	**2.57**	**0.93**	**3.79**	**0.87**	**0.32**	**0.20**	**0.65**

2.4 Discussion

2.4.1 Multi-step approach for dealing with data imbalance

One challenge in land cover fraction prediction is the tendency of the models to favour the mean over the extremes, as that minimises RMSE that may arise from incorrect predictions. However, that leads to increased fuzziness of the result, where pixels with high uncertainty are marked as a mix of many classes, and fractions of 0% are rarely, if ever, predicted. Our proposed multi-step approach adjusts the balance the other way. As it is a combination of one or two classification models and one regression model, it predicts significantly more pure pixels compared to a single regression model.

Therefore, the multi-step approach was successful in reducing MAE (and improving the related SCM metrics), as the particularly common case of 0% fractions was captured better. On the other hand, it comes at a cost of higher RMSE and lower R^2. This is because in highly uncertain, but pure cases, the model makes a best guess of a pure class, and due to the high uncertainty, the prediction is often incorrect. This leads to 100% error in those cases, which is highly penalised by RMSE. Due to this effect, the resulting map is closer to a discrete classification map, with less expressive transition gradients between land cover classes.

A similar effect was seen when using techniques such as median voting for RF regression. Like the three-step model, median voting resulted in more correct 0% fraction predictions, but likewise increased the chances of predicting 100% of the wrong class. Our results showed that median voting has a stronger effect in reducing MAE than the multi-step approach. Therefore it may be a suitable choice in cases where it is computationally infeasible to train three models. However, this approach is only applicable to RF models, whereas the three-step approach is generic and can be used with any combination of models. In addition, when the median voting approach was combined with the three-step approach,

the RMSE decreased without affecting MAE. Thus, three-step median RF achieved the best combination that is optimised towards reducing MAE, doing so without increasing RMSE as much as the single-step median vote model does.

These findings show that the use of a multi-step approach depends on what is more important for the user. If an occasional prediction of 100% of the wrong class is acceptable, then a multi-step model provides an overall more accurate result, especially for zero fractions. On the other hand, a single-step approach emphasises the strength of land cover fraction mapping by expressing gradual changes over space better, and avoids large errors. The latter is more likely to be useful for the modelling community that deals with uncertainty with probabilistic frameworks, and the former may be more useful for policymakers and land owners who are more concerned with what land cover is most likely to be present on the ground. In addition, we expect a multi-step approach to be more suitable for fine resolution mapping, where more pure pixels can be expected, and a single-step approach to be more useful for coarse resolution mapping, where mixed pixels are the norm.

It is also worth noting that the multi-step approach is flexible and can be used with any algorithm that provides both classification and regression modes, or with two separate unrelated classification and regression algorithms. Therefore there may be some combinations of models that have not been tested yet, but could achieve even higher accuracy.

2.4.2 Comparison with global land cover products

To gain insight into how well our proposed multi-step median vote RF model performs, we compared it to existing global land cover fraction products. These products, that only focus on a single land cover class, had varying accuracy compared to our model (see table 2.3). Some products, like GSW water occurrence (Pekel et al., 2016), had a higher accuracy. Others, like GFCC forest cover (Townshend, 2017) and GHSL built-up (Corbane et al., 2019), had a lower accuracy. The accuracy of the impervious surface class fraction from FROM-GLC10 mostly matches that of the built-up cover fraction of our proposed model. This shows that our proposed method, coupled with the large training dataset that we used, achieves similar and sometimes even better performance compared to specialised land cover fraction products, but brings the advantage of producing fractions for a variety of land cover classes that sum up to 100%.

The accuracies of other global products reported in literature likewise varied compared to our results. However, these comparisons are much more difficult to make, as the validation methods and scope vary significantly between the studies. For example, our RF three-step median model had a higher RMSE for the tree class (21.1%) than the one reported by Sexton et al. (2013) for their vegetation continuous fields product (16.8%). However, Sexton et al. (2013) validated their data using lidar datasets within several local study areas, rather than using global image interpretation data as we did. The study by

Montesano et al. (2009) that used a validation approach closer to ours to validate the MODIS tree cover product, reported an R^2 of 0.57, RMSE of 13.4%, root mean squared deviation (RMSD) of 21.3%, slope from a linear regression of 0.5 and intercept of 18.4. In comparison, our results for the three-step median RF model for the tree cover class had an R^2_{OLS} of 0.74, RMSE of 21.1%, slope of 0.87 and intercept of 5.82. Thus while our model appears to perform better, Montesano et al. (2009) only evaluated boreal regions rather than the entire globe, and used MODIS 500 m data rather than PROBA-V 100 m data.

2.4.3 Challenges and future outlook

Machine learning algorithms pose several challenges that are inherent to how the models are constructed. The trade-off between minimising RMSE and minimising MAE comes from the chosen loss function. Typically, in cases of high uncertainty, the loss function is minimised when the predictions tend towards the mean. In that case, the models predict in areas with a high cover of a fraction, such as mixed shrublands, a lower fraction of the class than expected, whereas in cases with a low cover, such as for fraction of built-up, higher fractions are predicted than expected. This is due to a prediction of the mean being less penalising than predicting the extremes incorrectly; e.g. for a case of 50% shrub cover, predicting 100% shrubs would be a larger mistake (and thus lead to higher RMSE) than predicting 15% shrubs. Likewise, predicting 0% built-up in areas covered by dark bare soils risks a case where it truly would be 100% built-up, so on average predicting 15% built-up in this case lowers the possible error. The three-step approach (or median vote) tilts the balance in the other direction, as e.g. the first step determines that the pixel is pure, and the classification step determines that it is more likely to be bare soil than built-up area. But there may also be cases when the classifier predicts the wrong class. If the pixel is not pure, the regression step still tends to predict towards the mean due to the loss function, so the challenge of underestimating large fractions and overestimating small ones remains. However, with the multi-step approach, it is now possible to influence the decision process of the model to tweak it towards the desired outcome. In addition, using a median vote in tree-based ensemble models makes the model tend towards the median, which is often one of the extreme values. In that case, we see the inverse pattern compared to mean vote: small fractions are underestimated (i.e. predicted as 0%), and large fractions are overestimated (i.e. predicted as 100%). On average, the model is still not biased, as the predicted values are more polarised but balance each other out. The multi-step approach could be improved further by exploring the option of using different machine learning models for each step, and by gathering more features that would increase the accuracy of the models at each of the steps.

Another challenge inherent in land cover mapping is discerning classes that are related, e.g. herbaceous vegetation and shrubs. These classes were particularly difficult to map, in part due to their heterogeneity, and in part due to confusion between herbaceous vegetation,

cropland and shrubs. As shown in Appendix 2.C, figure 2.C.1, specific types of features can be used to discriminate between these classes better. Specifically, the identification of herbaceous vegetation is primarily based on vegetation indices, identification of shrubs relies mostly on climate data, whereas cropland identification is more data-intensive and makes use of vegetation indices, temporal metrics, climate, soil and location information. These three classes are challenging to discriminate between not just for regression algorithms, but also for expert interpreters, which may lead to higher uncertainties also in the training and validation data for these classes. It is even more challenging to discern between these classes if they mix within the area of a single pixel, which is common in grasslands and shrublands, as well as in smallholder agriculture. The difference between the definition of trees, shrubs and herbaceous vegetation largely comes down to plant height, therefore dynamic information about vegetation height would allow mapping these classes more accurately. However, this would require either photogrammetry techniques (that typically rely on much finer spatial resolution and more overlap between the scenes) to reconstruct vegetation height, or the use of non-optical sensors such as synthetic aperture radar (SAR) interferometry or lidar data. Another way to differentiate between the vegetation classes could be to make use of hyperspectral data, which allows differentiation between different kinds of vegetation based on e.g. their photosynthesis processes or water content, which affect light absorption. Emerging new high level hyperspectral products, such as sun-induced chlorophyll fluorescence (SIF) or gross primary productivity (GPP), such as ones based on the upcoming FLEX satellite, could allow for a straightforward way to incorporate this extra information into land cover models. Hyperspectral data could be useful for better differentiation of non-vegetated classes as well, such as bare soil from urban, e.g. by separating the spectral signature of sand from concrete or asphalt. In addition, land cover time series information could help track land cover change over time, as the land cover at one time step depends on the land cover at the previous step. This information would allow the regressor to limit the predicted values to a smaller range, and thus reduce the noise in the predictions.

Another challenge is class imbalance. For example, the built-up class rarely forms a 100% fraction. That makes it simple to achieve a high prediction accuracy according to absolute statistics, as a fraction of 0% is in most cases not far off from the true value. However, a prediction of 0% in every pixel makes the fraction map not useful for user needs. This challenge is further exacerbated by the training dataset containing relatively few points in built-up areas to begin with. Therefore, having a more balanced training dataset may further increase the accuracy of the models. However, the issue of value imbalance within the class will always remain for land cover fraction mapping, therefore the multi-step model approach will be relevant, especially if the legend involves even more classes, or if the land cover is more homogeneous at the level of the mapping unit.

Several more challenges are yet to be tackled in this field, but doing so is becoming more and more feasible over time. Finer spatial resolution mapping, such as 10 m mapping using

Sentinel-2 data, is a future research direction, where the pixel footprint will more likely cover homogeneous land. Therefore, due to an increase in 0%/100% fractions, such future developments would be more likely to benefit from a multi-step approach or optimisation for MAE. The multi-step approach is fully portable to finer scales, but more research is needed to determine the effect of the different scales on the purity of the pixels, and how much benefit does a combination of classification and regression bring compared to doing only classification or only regression. Finer spatial resolution sensors can also be used for mapping fractions at coarser resolutions in a more precise way, by performing aggregation of the finer resolution pixels to estimate the land cover fraction at a coarser resolution. This is likely to become the norm if even finer resolution data (e.g. 1 m) becomes available globally.

Another benefit of using different optical sensors is higher spectral resolution. As PROBA-V only measures four spectral bands, the amount of information that can be retrieved from them is limited. Sensors such as Sentinel-2 MultiSpectral Instrument (MSI) have a much wider range of spectral bands that could be used both directly as features, as well as enable computing a more diverse range of VIs, such as ones based on the slope of the red edge. This could potentially improve the distinction between different land cover classes.

With more availability of such additional remote sensing data, it becomes increasingly more feasible to perform land cover monitoring and change mapping. Mapping land cover fractions is a great opportunity to track gradual changes, such as regrowth, better. The challenge here is that the higher uncertainty about fraction estimates may cause the time series of land cover fractions to fluctuate, making it difficult to determine robust trends.

2.5 Conclusions

We investigated ways to tackle the issue of accurately predicting the extreme fraction values of 0% and 100% by proposing a hierarchical multi-step approach combining classification and regression models. This approach was applied to an RF regression model, which our tests showed to have the highest accuracy (RMSE: 17.3%) for land cover fraction mapping compared to other regression algorithms. We also combined this approach with RF median voting. The combined RF median three-step approach obtained the best results for MAE (7.9%) and SCM OA ($72\,\% \pm 2\,\%$). The proposed approach results in predictions of the most likely pure land cover class, when the class is uncertain. This is in contrast to predicting a mix of classes when a standard one-step model is used, and therefore is useful for users who are more interested in the most likely class, rather than class probabilities. Based on this model, we created a demonstration map showing the global distribution of land cover in separate land cover fraction layers. Remote sensing features were the most important for model accuracy, although all other types of features

(climate, soil, terrain) also contributed significantly for some classes and thus could not be omitted without negatively affecting the overall model accuracy.

These findings directly contribute to the operationalisation of global land cover fraction mapping by analysing and advancing currently available methods for thematically exhaustive global land cover fraction mapping. Information on land cover fractions offers better precision than discrete land cover maps, and allows the users to manually define thresholds to generate discrete classifications of their own choosing, based on their classes on interest. Furthermore, given the recent advances in optical sensor spatial resolution and the resulting increase in pixel purity, the multi-step model approach may become more important in the future. Lastly, due to the advances in spatial and spectral resolution, longer imagery time series of existing sensors, and the availability of non-optical sensor data, this work paves the way towards operational land cover fraction change mapping, which would allow monitoring gradual land cover change.

Appendices

2.A List of features used in regression models

Table 2.A.1: List of features used as input for the models tested in this study, and their data sources.

Category & data source	Features	
Location (3, intrinsic)	longitude latitude	absolute latitude
Vegetation indices (22, derived from PROBA-V 100m top-of-canopy reflectance v1.02 (Dierckx et al., 2014))	minimum NDVI maximum NDVI median NDMI NDMI yearly IQR NDMI March-May IQR NDMI June-August IQR NDMI September-November IQR NDMI December-February IQR OSAVI March-May IQR OSAVI June-August IQR OSAVI September-November IQR	OSAVI December-February IQR EVI March-May IQR EVI June-August IQR EVI September-November IQR EVI December-February IQR median NIRv NIRv yearly IQR NIRv March-May IQR NIRv June-August IQR NIRv September-November IQR NIRv December-February IQR
Temporal metrics (9, derived from a harmonic model over time series of PROBA-V 100m top-of-canopy reflectance v1.02 (Dierckx et al., 2014))	NDVI order 1 cosine NDVI order 1 sine NDVI order 2 cosine NDVI order 2 sine NDVI trend coefficient	NDVI order 1 phase NDVI order 1 amplitude NDVI order 2 phase NDVI order 2 amplitude
Terrain (4, ASTER GDEM V003 (NASA et al., 2019))	elevation slope (log-transformed)	aspect terrain position index
Climate (21, WorldClim 2.0 (Fick and Hijmans, 2017))	January precipitation (log) April precipitation (log) July precipitation (log) October precipitation (log) January solar irradiance July solar irradiance mean temperature temperature monthly range isothermality temperature annual range annual precipitation (log)	temperature seasonality minimum solar irradiance maximum solar irradiance mean solar irradiance mean windspeed mean water vapour pressure coldest month precipitation (log) warmest month precipitation (log) wettest month solar irradiance driest month solar irradiance
Soil (8, SoilGrids (Hengl et al., 2017))	soil available water soil bulk density soil cation exchange capacity (log) soil clay fraction	soil coarse fragments (log) soil pH soil sand fraction soil water wilting point

2.B Regression model comparison

2.B.1 Regression models tested in this study

Before applying the multi-step approach, we tested the performance of various regression models for land cover fraction mapping. We tested four types of models: linear models, models based on decision trees, machine learning models not based on decision trees, and ensemble learning (for an overview, see table 2.1).

First, we chose to compare five types of linear models, to have a baseline for a comparison with the nonlinear machine learning models. The most simple model we selected was the general linear regression model (GLM), also known as multivariate linear regression. It is an extension to the standard linear regression that allows for multiple outcomes. Next, we tested two linear models that include input data regularisation in the model itself: lasso regression and partial least squares (PLS) regression. We also tested multinomial logistic regression (MLR), which is usually used for classification and is fit using land cover class labels rather than fractions, but the output includes probabilities for each class that add up to 100%. We fit MLR using the dominant land cover class as a label for the pixel and used the class probabilities as a proxy for land cover fractions. Lastly, we tested fuzzy nearest centroid (FNC) regression, also called fuzzy nearest prototype, fuzzy c-means or fuzzy k-means. It is a simple regression method, where the land cover fractions in a pixel are determined by the distance of the pixel from the centroids of each class in feature space.

The second group tested in this study was machine learning methods not based on decision trees. Neural networks (NNs) are a promising technique for land cover fraction mapping, as they allow both multiple inputs and multiple outputs, and, using the softmax activation function, also ensures that the result sums up to 100% with no need for additional postprocessing. In this study, after performing tuning, we ended up using a multi-layer perceptron (MLP) with three hidden layers with 128, 64 and 32 neurons per layer respectively. We used the Nadam optimiser (Dozat, 2016) with MAE as the loss function to optimise the NN, and softmax activation for the output. The models were trained using the `keras` package, built upon TensorFlow that enables the use of a graphics processing unit to accelerate the NN training process. The other method in this category that we tested was Support vector machines (SVMs), which are machine learning models that attempt to find the optimal boundary between the class clusters in feature space by constructing a dividing hyperplane. For land cover fraction classification, we used SVM regression based on Least Squares SVM. As SVM models are univariate, we used the binary relevance method (Karalas et al., 2016): training separate models per class that predict a single class, and then combining the results.

The third group we tested were tree-based machine learning models. Random forest (RF) regression is a popular method for land cover classification that works by building a number

of CART decision trees based on random subsets of the input training data, and taking the mean or median of the "votes" of these individual decision trees. RF is univariate, therefore we again used the binary relevance approach. Next, we tested Cubist regression. It is based on RF regression, but instead of using a threshold of a feature to split the decision tree, Cubist uses a linear regression based on a subset of the data relevant for the split in question. In addition, it features committees, a boosting technique that iteratively trains trees so as to learn from the previously generated ones. Model tuning led to us using 10 committees. Cubist predictions were also made using the binary relevance method.

Lastly, we made an ensemble from the two machine learning models that produced the lowest RMSE and MAE respectively: RF and Cubist regressions. We used the super learner algorithm (Van der Laan et al., 2007) to create a hierarchical ensemble, where the two models were cross-validated using 10-fold cross-validation to obtain relative weights of each model for each land cover class. The predictions of the two models, along with the weights of the models, were then used as input features for another RF regression metalearner. The output from this metalearner is the final prediction of the whole ensemble. As the ensembled methods are univariate, we used the binary relevance method in this case as well.

2.B.2 Comparison results and discussion

The overall accuracy statistics of the compared models are reported in table 2.B.1, and per-class statistics in figure 2.B.1. The results show that the R^2 statistics, especially R^2_{NSE}, are in agreement with the RMSE statistics, and the overall accuracy (OA) from the subpixel confusion-uncertainty matrix (SCM) is in agreement with the MAE statistics. R^2 is a representation of how closely the predictions correlate with the validation data, therefore big outliers have a large effect on the value. The SCM does not take this into account, as it does a comparison on the overlap of fractions in each pixel, which doesn't apply an extra penalty for large errors.

The baseline models that performed the best were the two tree-based machine learning models: RF regression and Cubist. Both of these models are univariate, therefore predictions per class had to be made using the binary relevance method (one model per class). This shows that the disadvantages of the binary relevance method, namely that each model is only trained on the fractions of its own class and has no knowledge of the fractions of the other classes, and the need for a rescale step to make sure the fractions sum up to 100%, are outweighed by the advantages of these machine learning models and the flexibility of training separate models for each class.

The super learner method that combines both RF regression and Cubist regression resulted in a model that is in between the two ensembled models in terms of accuracy. Its RMSE was below that of Cubist regression, but above that of RF regression. Conversely, the MAE was below that of RF regression, but above that of Cubist regression. This shows that

Table 2.B.1: Accuracy statistics of the tested regression models. Best performing statistics are highlighted. "slope" refers to the OLS-estimated slope, "int" refers to the OLS-estimated intercept. "Only RS features" stands for a model trained only with VIs and temporal metrics.

Model	RMSE (%)	MAE (%)	R^2_{NSE}	R^2_{OLS} (slope/int)	OA (%)	Kappa
Equal proportions	29.9	21.4	0	—	26 ± 5	0.13 ± 0.07
FNC	24.4	13.5	0.33	0.35 (0.80/2.89)	53 ± 4	0.42 ± 0.06
GLM, PLS, Lasso	21.6	12.7	0.48	0.49 (1.10/-1.42)	56 ± 4	0.43 ± 0.05
MLR	21.6	12.1	0.48	0.48 (0.96/0.62)	58 ± 4	0.46 ± 0.06
MLP NNs	22.7	9.2	0.43	0.52 (0.70/4.25)	68 ± 1	0.57 ± 0.02
SVM regression	20.7	8.9	0.52	0.56 (0.79/3.02)	69 ± 2	0.58 ± 0.03
Cubist regression	18.1	**8.1**	0.63	0.65 (0.88/1.77)	$\mathbf{72 \pm 2}$	$\mathbf{0.63 \pm 0.03}$
RF regression	**17.3**	9.4	**0.66**	**0.67** (1.09/-1.22)	67 ± 4	0.57 ± 0.05
″ only RS features	18.4	10.3	0.62	0.64 (1.09/-1.25)	64 ± 4	0.54 ± 0.05
Cubist + RF ensemble	17.7	8.6	0.65	0.65 (0.95/0.73)	70 ± 3	0.61 ± 0.04

the metalearner (another RF model) could not differentiate between the two ensembled models well enough to select the model that is the most accurate for a given class. Rather, it weighted in both of the models' predictions, losing the advantages of the models when taken separately.

The linear statistical models (GLM, PLS, Lasso regression and MLR) performed the worst, as they are limited to a linear approach. The results of all of the linear models were very similar. There was no added value to PLS and Lasso regressions over the basic GLM. This is likely due to the additional preprocessing step of feature selection, as detailed in section 2.2.3. Lasso and PLS regressions are extensions to GLM that include a regularisation step, which proved to be unnecessary if all of the highly correlated features are already manually removed.

Fuzzy nearest centroid (FNC) was the model that performed the worst out of the non-trivial models tested. This indicates that the shape of the feature data in feature space is too complex to capture merely with a centroid approach. This is further evidenced by a much better result obtained by the binary relevance SVM model, which works on a similar principle but can capture more complex shapes.

MLP NNs, despite being a multivariate machine learning model, performed worse than all of the tested binary relevance methods. Nevertheless, it performed better than all of the multivariate linear models.

All tested models had lower accuracy when estimating the cover fractions of vegetation classes (herbaceous, trees, crops, shrubs), compared to the non-vegetated land cover (inland water, built-up, bare land), as shown in figure 2.B.1. This was exacerbated by the imbalance in class distribution: the built-up class is rare and very rarely forms a majority, therefore a prediction of 0% leads to a perfect prediction most of the times, and even if it does not, the error is low. In contrast, tree cover had a lot more balanced

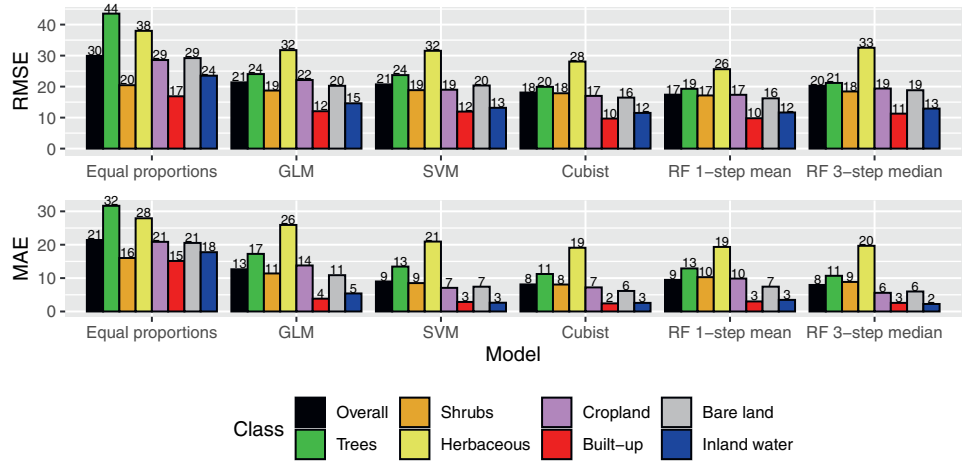


Figure 2.B.1: Comparison of absolute RMSE (top) and absolute MAE (bottom) per class of the best performing models in their category, and the equal proportions solution as a reference.

distribution of fractions, including a large number of pure pixels, which leaves no such trivial solution. The RRMSE and RMAE statistics show this (see figure 2.B.2): the tree cover prediction has low relative error given its high mean value (32%), whereas the built-up class is the most challenging to predict according to RRMSE given its low mean value (3%). The shrub cover fraction is also very challenging to predict, since it had the highest RMAE and none of the models showed large improvements in RMSE compared to the equal proportions model. There were only moderate improvements in the prediction of the herbaceous vegetation fraction as well. In contrast, most models were a significant improvement in predicting tree, built-up, water and bare cover fractions compared to the equal proportion model, indicating that the features used to train the models were useful to distinguish these classes from the others and to quantify their proportions in each pixel.

To sum up, we tested a number of regression methods for characterising land cover by means of predicting thematically comprehensive land cover fractions at the global scale, and found that RF regression and Cubist regression produce output of the highest accuracy by RMSE and MAE, respectively. Our findings agree with those of Li et al. (2018), who compared Cubist and RF regression for water fraction classification and found that Cubist performs slightly better than RF regression for this particular class. Their Cubist regression result achieved 7.52% MAE and 10.39% RMSE. Our respective results using Cubist regression were 2.57% MAE and 11.51% RMSE. While Cubist works best for the inland water class, when considering all of the classes, RF regression nevertheless results in higher accuracy (see table 2.B.1). The difference in the reported numbers may

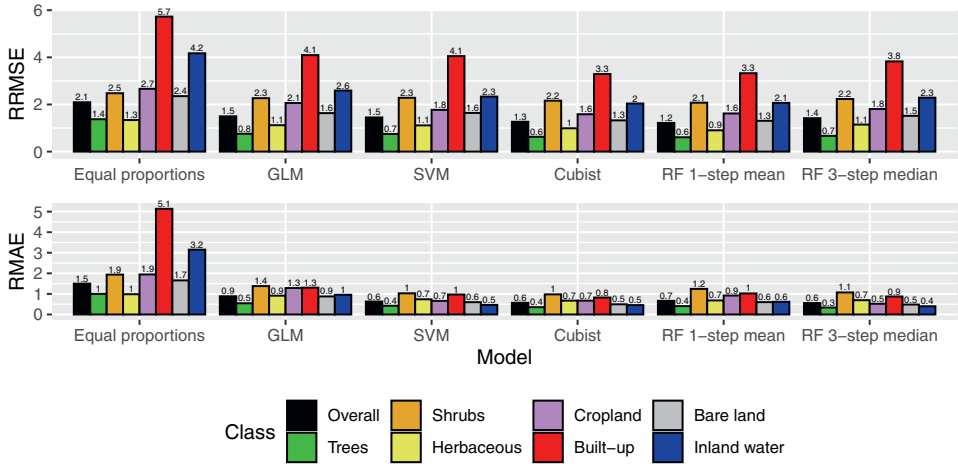


Figure 2.B.2: Comparison of relative RMSE (top) and relative MAE (bottom) per class of the best performing models in their category, and the equal proportions solution as a reference.

be due to the differences in scale (regional vs global) and training data balance. As we focused on predicting multiple classes, our training dataset intrinsically had higher zero inflation.

Our results showed that the binary relevance method works well to facilitate the use of univariate regression methods for the global land cover fraction mapping task, as the highest accuracy was achieved by algorithms that used the binary relevance method. This further expands the field of possible algorithms that could be used for the task. In addition, both the binary relevance method and the multi-step method can accept different algorithms for its submodels, i.e. for different classes in the case of binary relevance and for different steps in the case of the multi-step approach. There may be a combination of models that would work even better than the models we tested. We could not cover the entire range of methods in the scope of this study, therefore there is still room for improvement. For example, classical spectral unmixing methods, such as spectral mixture analysis, has been often used in the past (e.g. in Adams et al. (1995)). However, due to the limitations of the spectral mixture analysis methods (Somers et al., 2011), we could not use it with the features we selected, as they do not form a linear mixture representing the land cover fractions. Modern machine learning methods are advantageous for their ability to model complex nonlinear relationships between variables and thus make use of all of the features.

2.C Feature importance

To gain insight on what features drive the RF three-step median model, we performed permutation importance on it: we shuffled the values of each feature in turn, and the model made predictions based on all of the features, including the shuffled one. This results in a decrease in the accuracy of predictions, as the feature in question no longer contains meaningful information. We then recorded the resulting increase in MAE compared to the validation set as a measure of feature importance, both per class and overall for all classes combined. The results are shown in figure 2.C.1.

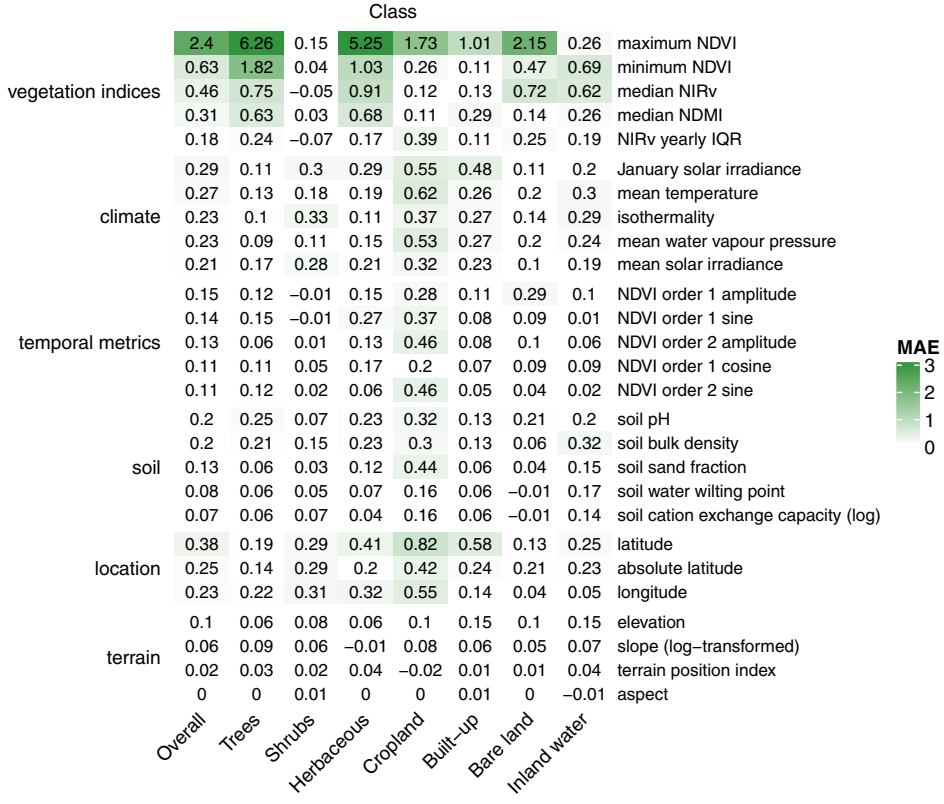


Figure 2.C.1: Random Forest three-step median approach variable importance, top 5 features per category. Categories are ordered by cumulative importance. The values represent increase in MAE for the given class when the given feature is permuted.

The overall most important variables were the maximum and minimum NDVI over the whole time series of PROBA-V imagery. They were followed by the median NIR_v and NDMI over the whole time series. The importance of remote sensing information can

also be seen from the accuracy statistics of a single-step RF regression model, when only vegetation index and temporal features were input into the model (see table 2.B.1, "only RS features"): RMSE and MAE increased only by around 1 percentage point.

The remote sensing data was of lower importance (and sometimes even confounding) only for the shrub class. This class is complicated to distinguish from other natural vegetation, and thus when trying to predict shrub cover fractions, the model benefited from a wide variety of additional features, including location and climate information.

To distinguish crops in particular, soil information, such as pH, bulk density and sand fraction, was useful for the model. Crops are usually grown in fertile soils that can sustain them, and due to the fact that cropland is further managed, the soil is also altered to be more fertile, which affects these soil properties. In addition, harmonic metrics derived from time series benefited crop cover fraction estimation the most of all the classes. The second order amplitude and sine of the harmonic model of the time series allows the model to detect areas with a double harvest throughout the year, which is indicative of crops.

Climate data, especially the mean temperature and the closely (exponentially) related mean of water vapour pressure, was also most useful for predicting crop cover fractions. It was also beneficial for estimating built-up fractions. Location features were the most important for predicting the cover fractions of these two classes as well. These classes tend to be spatially clustered, when looking at a large scale. While the models for predicting the cover fraction of these two classes made the best use of the location features, all of the classes could benefit from them to some extent, thus it is beneficial to include these intrinsic parameters in the model. Both absolute and regular latitude were used by the model to improve prediction accuracy, and latitude was more beneficial than longitude for increasing the prediction accuracy.

Terrain information was the least useful feature category. It was most useful for inland water fraction prediction, since it typically has little to no slope, and rarely occurs at high altitudes. The aspect was the one feature that does not appear to have contributed to model accuracy.

Remote sensing features were the most important features for the multi-step median RF model, especially the maximum and minimum NDVI over the entire time series. Note that these values are taken from a time series that has undergone temporal outlier removal, as detailed in section 2.2.2, and thus roughly correspond to the 5th and 95th percentiles of the data without additional temporal filtering. Multiple vegetation indices were useful for increasing the model accuracy: both NDMI and NIR_v median over the time series were much more important than any other feature from other groups. Nevertheless, even though a lot of the other features were of much lower importance, they contributed to prediction accuracy enough so that they could not be easily excluded from the models as redundant (after the removal of collinear features as explained in section 2.2.3). Methods based on decision trees help with effectively using features that may also have an overlap in the

information that they provide, although linear models likewise tended to not exclude any features as non-informative. This is also due to the large variety of land cover classes in the study, since a feature is useful if it helps predict any of the land cover classes better.

The result that including more training features is beneficial, but remote sensing data is the most important, is in line with the conclusions of e.g. Li et al. (2018) and Hengl et al. (2017). Remote sensing data is also unique in that it forms a time series, which enables us to both calculate additional temporal metrics and to monitor land cover change over time. In addition to remote sensing data, Hengl et al. (2017) also noted high importance of climate data, however, it is focused on soils, whereas climate has more effect on long-term processes such as soil formation than on land cover. Climate data was the second most important in our case, mostly for the crop fraction estimation, which is also closely linked with soils.

The feature importance results showed that all of the feature groups used in the study were useful, therefore leaving a feature out means sacrificing some predictive power. On the other hand, leaving out some features would be beneficial in that less time would be needed for processing, as that feature would no longer need to be downloaded, preprocessed and processed. This is an important consideration, given that for global land cover mapping, all features need to be available for the whole globe as well. In addition, model training performance (time for training and memory usage) may be an important consideration for global land cover mapping, as it may limit the scope of what models may be used for this task. For example, ensemble learning techniques like the super learner are very resource-intensive and may take weeks to train. Therefore for operational land cover product creation purposes, it is important to take into account not just the model accuracy, but also whether the improvement of the accuracy is worth the increase in processing time or computing resource usage.

2.D Model accuracy changes per predicted fraction

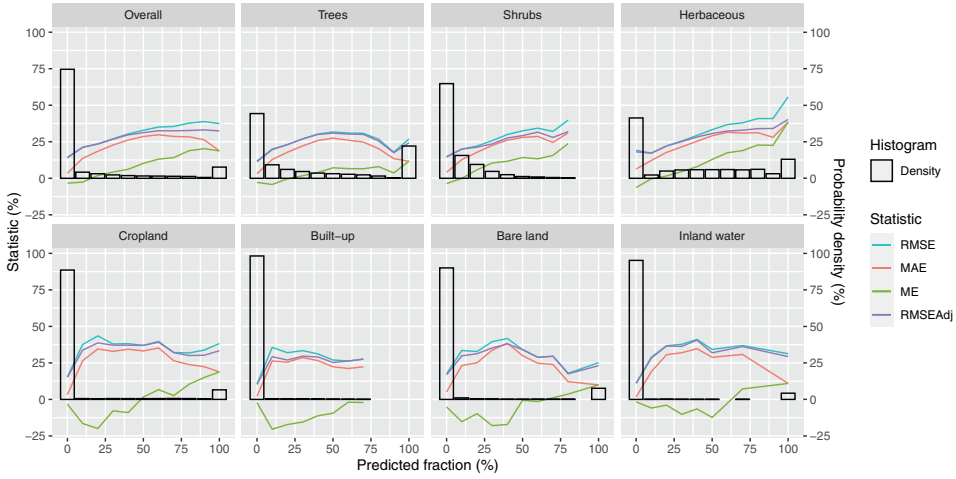


Figure 2.D.1: Change in accuracy statistics of the RF three-step median model per predicted land cover fraction, overlaid over the histogram of predictions. "RMSEAdj" stands for RMSE when the bias in predictions is adjusted for (ME is subtracted from the residuals). Bins containing less than 10 samples were omitted due to too low sample size leading to unreliable error statistics.

Chapter 3

Improving global land cover fraction change estimation using a Markov chain model

This chapter is based on:

D. Masiliūnas, R. Burger, N.-E. Tsendbazar, and D. Marcos (2024). "Improving global land cover fraction change detection using a Markov chain model". Under review.

Abstract

Global land cover mapping has aided monitoring of the complex changes of the surface of the Earth and provided vital information to understand the interactions between human activities and the natural environment. Most global land cover products are produced with discrete classes, indicating the dominant land cover class in each pixel. Fraction mapping, which expresses the proportion of every land cover class in each pixel, is able to characterise heterogeneous areas covered by multiple land cover types and enables the monitoring of gradual changes. However, land cover fraction maps have shown unrealistic inconsistent changes between subsequent yearly maps, making it difficult to monitor true change on the ground. To improve the detection of land cover fraction change, we proposed a Markov chain postprocessing model and compared it to four other models: a basic RF regression model without postprocessing, a linear smoother, and recurrent and temporal RF postprocessing models. Based on features derived from Landsat 8, initial fractions were predicted on a global scale for the years 2015 to 2018. The RF regression model was trained on over 150 000 reference points, provided by the Copernicus Global Land Service Land Cover 100 m project. The recurrent RF model incorporated the predicted fractions of its previous year to account for temporal information. The temporal RF and the Markov chain model, which incorporated temporal information from its adjacent years and class co-occurrence dependency, were applied on the initial basic RF fraction estimations as a postprocessing step. An accuracy assessment was performed with a subpixel confusion-uncertainty matrix to compare the performance of the models. All fraction estimations were validated on over 30 000 reference points that contained multitemporal land cover fraction data, and were analysed on samples with varying levels of change. The results indicate that the Markov chain model achieved better overall and change accuracies compared to the other models. Compared to the basic RF model, the Markov chain postprocessing approach consistently and significantly improved the overall and per-class accuracies. The recurrent and temporal RF models obtained considerably poorer and less consistent results, potentially due to the effects of domain transfer. These results confirm that Markov postprocessing has the potential to reduce spurious and inconsistent multitemporal land cover fraction change, which is in line with other studies. The findings highlight the importance of the addition of Markov models for future land cover mapping studies, so that multitemporal predictions provide realistic and consistent land cover change, essential for global change estimations.

Keywords. Land cover, land cover fractions, Markov chain, time series analysis, Random Forest

3.1 Introduction

Global land cover mapping has aided monitoring changes on the surface of the Earth and provided vital information to understand the interactions between human activities and natural environment (Zhao et al., 2021). Changes in land cover, such as deforestation and urban expansion, have a great influence on different Earth systems (Mahmood et al., 2014; Seto et al., 2012), and can be closely monitored with land cover maps. Land cover mapping therefore plays a critical role for applications such as climate change, biodiversity and environmental modelling (Chen et al., 2021; Pielke et al., 2011; Sharma et al., 2018). In the last decades, satellite remote sensing has enhanced the development of global land cover maps. Increased availability of high spatial and temporal resolution has provided more opportunities for enhanced global land cover products.

Most of the global land cover products are provided with discrete land cover classes. However, recent products have focused on indicating continuous fractions of land cover classes, e.g. the Copernicus Global Land Service Land Cover 100 m (CGLS-LC100) product (Buchhorn et al., 2020). With fraction mapping, instead of a single discrete class, the proportion of every land cover class is reported for each pixel of the map. Land cover fraction mapping provides more precise information about the land cover classes and, given a moderate minimum mapping unit in the order of 100 m, can depict heterogeneous areas better than discrete land cover classification (Masiliūnas et al., 2021a). Moreover, looking at land cover fraction time series allows for tracking gradual land cover fraction change over time. The detection of gradual changes is facilitated by land cover fraction maps, whereas they could potentially be missed by a discrete classification (Souverijns et al., 2020). Subsequently, this provides great opportunities for applications such as forest monitoring, land degradation, biodiversity conservation, etc.

Land cover fraction maps can be produced with various machine learning algorithms. Since the output is continuous, regression algorithms, rather than classification algorithms, are required. One of the most common regression techniques in remote sensing is random forest (RF) regression (Breiman, 2001), which has proven to be a reliable method with relatively high accuracy and fast computation time (Belgiu and Drăguţ, 2016). RF regression has been applied in several land cover fraction studies, for instance looking at fractions of specific classes (Liu et al., 2021b; Mutanga et al., 2012) and comparisons of the performance of RF regression with other machine learning algorithms, such as Cubist and SVM regression (Masiliūnas et al., 2021a; Walton, 2008).

Accurate and reliable estimations are essential for users of land cover maps. However, inconsistencies and higher uncertainty about fraction estimates may cause the time series of land cover fractions to fluctuate, making it difficult to detect actual change between subsequent years (Masiliūnas et al., 2021a; Yang et al., 2016). Inconsistent time series have been a common problem in the production of land cover maps, partly due to illogical

transitions (Cai et al., 2014; Congalton et al., 2014; Gong et al., 2017). Illogical transitions are characterised by a change in land cover (either discrete or continuous) that is unlikely to be observed and violates ecological rules within a given period of time (Cai et al., 2014; Yang et al., 2016). For instance, looking at one pixel, a transition within one year from urban to shrubs is deemed illogical. The same holds for a transition from grassland to forest within a year, especially if in previous years it was always classified as forest. Another issue arises when changes in land cover are predicted, while in reality, no change is happening. Inconsistent changes are ultimately due to classification errors of the algorithm or model (Cai et al., 2014; Gong et al., 2017; van Oort, 2005). Classification is often confounded by spectrally similar land cover classes and heterogeneous landscapes (Chen et al., 2021), and the same holds for regression in land cover fraction estimation (Masiliūnas et al., 2021a).

Several techniques have been examined in prior research to improve the accuracy and consistency of land cover mapping efforts, primarily focusing on smoothing the land cover maps spatially and/or temporally. Information in the spatiotemporal context has a lot of potential for consistency improvement and noise removal (Wang et al., 2015). Temporal context can be incorporated in a recurrent RF regression model, where a model producing a fraction map can be dependent on the fraction map of the previous year. Another way to include temporal context is by providing temporal features for training the regression model. More elaborate models that consider temporal context are Markov models, often used as a postprocessing step after initial classification in land cover mapping studies. One example of a Markov model is the Markov chain, which considers temporal context to improve the consistency of time series. A Markov chain is a probabilistic model in which the system is assumed to be a Markov process, meaning that the only temporal dependency of a cell considered is with respect to the previous state of the cell (Gagniuc, 2017). Transitions between states are usually governed by a transition probability matrix for discrete mapping, while probability densities are used for continuous variables (Hill and Spall, 2019). In its core, a Markov chain aims to model the trajectory of each pixel through the time series. Markov chains are explainable AI models that have a relatively simple and intuitive logic, making them attractive alternatives to more complex formulations of stochastic land cover models (Iacono et al., 2015).

Some studies have focused on the implementation of a Hidden Markov Model (HMM), which has proven to be a successful method to improve consistency (Abercrombie and Friedl, 2016; Bogaert et al., 2022; Gong et al., 2017; Souverijns et al., 2020; Yang et al., 2021). HMMs primarily focus on the temporal context, but can also be extended with spatial context (Liu et al., 2021a). HMMs are closely related to Markov chains, but HMMs are more complex as they attempt to model an unobserved hidden state, which has to be learned from auxiliary information. They are typically applied on the probabilities accompanying discrete classes (Abercrombie and Friedl, 2016), and more research is needed to adapt these models to land cover fractions. Another Markov Model is the Markov

Random Field (MRF) model, which considers multiple dimensions, typically both spatial and temporal context (Cai et al., 2014; Jia et al., 2014; Kasetkasem et al., 2005; Magnússon et al., 2021; Solberg et al., 1996; Wang et al., 2015). In MRF models, spatial context is incorporated by accounting for spatial relations between the fractions of neighbouring pixels. Although the spatial context is able to significantly improve the accuracy and consistency of land cover maps, it also requires reference data that includes spatial relations between neighbouring pixels, and it brings additional computation time. Besides that, one weakness of MRF models is that they can just act on the maximum clique, instead of the whole chain (Yang et al., 2021). Therefore, while more complex models such as HMM and MRF are promising, more research is needed to adapt these algorithms to the case of land cover fractions rather than discrete classes.

Markov models have not often been explored for continuous land cover fraction mapping, even though accurate and reliable fraction estimations are of great importance for tracking gradual changes and trend estimations. Souverijns et al. (2020) used an HMM for reducing variability in a land cover fraction time series, but applied it on discrete classes and only transferred information about change or no change to the land cover fractions, therefore ignoring gradual changes.

RF regression models are commonly used for land cover classification tasks, including predicting land cover fractions, but their ability to act as a postprocessing method in land cover time series has rarely been explored either. Xu et al. (2022) has found that RF is able to reduce the overprediction of change in time series analysis, but more research is needed on which problems RF is capable of solving as a postprocessing algorithm. In this study, we proposed and studied two RF models made to reduce the variability in the time series of land cover fractions: a recurrent RF model that learns from its previous year prediction of land cover fractions, and a temporal RF model that purely postprocesses a time series of land cover fractions predicted by a base classifier.

The focus of this paper is on exploring ways to improve the accuracy and consistency of yearly land cover fraction maps by using machine learning techniques combined with a Markov chain, adapted for land cover fraction postprocessing. The objectives of this study were to assess the accuracy of land cover fraction predictions and fraction change when predicted using (1) a simple Random Forest regression model without postprocessing, (2) postprocessing using a Markov chain model, (3) postprocessing using a temporal Random Forest and (4) a recurrent Random Forest regression model.

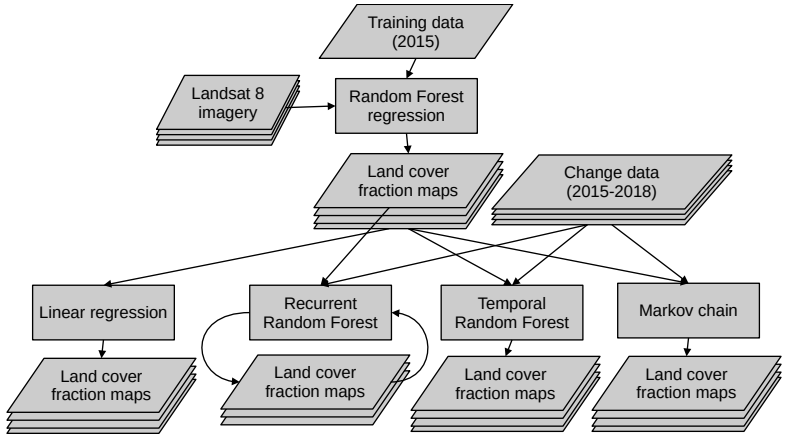


Figure 3.1: General workflow used in this paper. All of the output land cover fractions were assessed using the validation dataset (2015-2018).

3.2 Methodology

3.2.1 Reference Data

The reference data used for this study was provided by the CGLS-LC100 project (Buchhorn et al., 2020). There are three datasets, one dataset was used for training random forest models ("2015 training data"), the second dataset was used for validating all of the tested models ("validation data"), and the third dataset was used to train the Random Forest postprocessing models and to compute the co-occurrence matrix for the Markov chain ("multitemporal training data").

The RF training data contains more than 150 000 sample points across the globe and was generated by a team of experts at the the International Institute for Applied Systems Analysis (IIASA). These sample points describe the fractions of land cover classes in each sample site in the year 2015. This dataset was created using systematic sampling with extra sample points for underrepresented classes (Buchhorn et al., 2020).

The validation data contains more than 30 000 sample points across the globe including yearly multitemporal (2015-2018 inclusive) fraction data and was generated by a team of experts coordinated by WUR. The validation dataset is independent from the training dataset, as it was collected by a different team and used spatially representative stratified random sampling (Tsendbazar et al., 2021b).

The multitemporal training dataset contains more than 33 000 sample points across the globe, providing yearly multitemporal (2015-2018 inclusive) fraction data, and was generated by IIASA (Buchhorn et al., 2021). This dataset had a clustered sampling design.

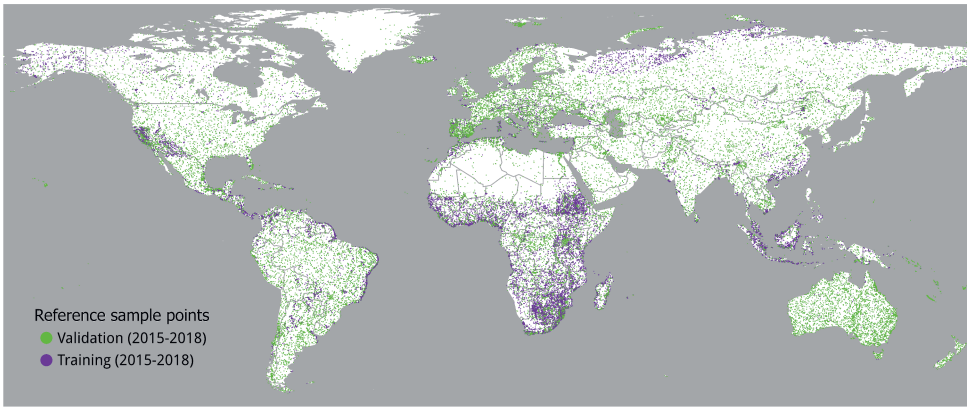


Figure 3.2: Global distribution of the reference sample points. The validation dataset contains 30 489 multitemporal sample points, and the multitemporal training dataset contains 33 251 multitemporal sample points. Note that the validation sample points have been drawn on top of the multitemporal training sample points.

The spatial distributions of these multitemporal datasets for training and validation are shown in Figure 3.2.

All sample sites have an area of 100 m × 100 m, aligned to the Sentinel-2 grid with Universal Transverse Mercator (UTM) projection. Each site is divided into 10 m × 10 m subpixels, which were labelled using discrete classes and then converted into land cover fraction estimates by calculating the proportion of subpixels covering the sample site for each class (Tsendbazar et al., 2021b). All reference data are provided with 12 land cover classes, which follow the definitions as suggested by the United Nations LCCS (Buchhorn et al., 2020). In this study, a few rare classes were discarded or merged, in line with the classes used by Masiliūnas et al. (2021a), to get a total of seven classes: bare land, cropland, grassland, shrubs, trees, built-up and inland water.

3.2.2 Landsat 8–based features

This study used Landsat 8 satellite imagery to derive relevant features for model training and prediction. The benefit of using Landsat 8 imagery is that the satellite has been operational since 2013, therefore its imagery covers the entire temporal range of the reference data without gaps or needing to combine several sensors. This was an important consideration for deriving features for training, as we used a 3-year sliding window to obtain robust statistics and therefore needed data going back to 2014. With the launch of Landsat 9, the time series could be extended further, if the reference datasets were to be updated in the future.

Table 3.1: List of used vegetation indices, their corresponding formula and spectral bands from Landsat 8.

VIs	Formula	Bands used
NDVI	(NIR – Red) / (NIR + Red)	B4, B5
NDMI	(NIR – SWIR1) / (NIR + SWIR1)	B5, B6
NBR	(NIR – SWIR2) / (NIR + SWIR2)	B5, B7

The entire time series (March 18, 2013–July 14, 2021) of the USGS Landsat 8 Operational Land Imager (OLI) Level 2, Collection 2, Tier 1 product was derived from Google Earth Engine. To align Landsat imagery with the sample sites, values from each band (apart from the thermal band) were extracted using the polygon footprints of each $100\,\mathrm{m} \times 100\,\mathrm{m}$ sample site, using a weighted mean reducer. The time series was filtered for clouds with the CFMask algorithm provided by the Landsat product, keeping only pixels marked as no fill, no cloud, no dilated cloud, and no cloud shadow.

The next step involved preprocessing the time series with a temporal filter to remove outliers, i.e. noise that was undetected by the CFMask algorithm. A LOESS curve was fitted on the time series of the blue reflectance band, and outliers over 2 standard deviations from this fitted curve were removed. After preprocessing the time series, the following VIs have been generated: Normalised Difference Vegetation Index (NDVI), Normalised Difference Moisture Index (NDMI) and the Normalised Burn Ratio (NBR) (Table 3.1). These VIs were chosen because they contain vital spectral information in the red, near infra-red (NIR) and shortwave infra-red (SWIR) bands. Then from these VIs, temporal statistical attributes were extracted. The temporal statistics were derived from the time series of each year plus its adjacent years, i.e. using a 3-year window, over each reference site. The statistics included the median and interquartile range of the VIs. In addition, harmonic analysis was performed on the NDVI time series (Jakubauskas et al., 2001). Sine and cosine components were derived for two frequency orders (annual and semiannual) in order to account for seasonality. From these two harmonic orders, phase and amplitude were calculated, next to the intercept, trend and minimum and maximum NDVI, over each 3-year period. Lastly, the longitude and latitude of the reference points were included, so that the RF model could learn spatial patterns. All features that were used for model training and prediction are listed in Table 3.2.

3.2.3 Random Forest models: basic, linear smoothing, recurrent, temporal

After obtaining the features listed above, a baseline RF regression model was created. First, the training dataset with land cover fractions from 2015 was used for training the RF model using the binary relevance method, i.e. building one model per class (Masiliūnas et al., 2021a). This trained set of RF models then predicted land cover fractions for

Table 3.2: List of features used as input for the basic RF model.

Category	Feature
Location	longitude
	latitude
NDVI	median
	IQR
	minimum
	maximum
	order 1 cosine
	order 1 sine
	order 2 cosine
	order 2 sine
	intercept
	trend coefficient
	order 1 phase
	order 1 amplitude
	order 2 phase
	order 2 amplitude
NBR	median
	IQR
NDMI	median
	IQR

the points of the validation data for each year (2015-2018). Usually, RF models use the mean to ensemble tree votes, but in this study median voting was chosen in order to predict extreme fractions of 0% and 100% more often. For most classes, extreme fractions of 0% and 100% are more probable, because most 100 m × 100 m areas often are only dominated by a few land cover classes, while the remaining classes only have fractions of 0% (Masiliūnas et al., 2021a). The RF models were created with the "ranger" package (Wright and Ziegler, 2017) in the software R (R Core Team, 2021).

The output of the time series of the basic RF model includes unrealistically high fluctuations in its predicted land cover fractions over the years. Therefore, we investigated several postprocessing techniques to reduce this spurious variation and compared the models against each other. As a first postprocessing baseline, i.e. to test how well postprocessing models perform when applied to the basic RF model output, we created a simple postprocessing model using a linear regression. For each pixel and each class fraction, a linear model was fit on the four points of the 2015–2018 time series of basic RF predictions, and its fitted values were used as the postprocessed fraction values of the linear smoother. This

simple method serves as an example of a postprocessing algorithm that does not require any additional training data.

In addition to the basic RF model, two additional RF models were created, both of which incorporate temporal information to be comparable with the Markov chain method: the recurrent and temporal RF models.

The recurrent model incorporates information about the past state of the land cover fractions. The 2015 predictions of the recurrent model were based on the basic RF regression model. But for 2016 and later, in addition to the features from Table 3.2 for the given year, the predicted fractions from the previous year were also included as features for training the recurrent model. Since the basic RF regression model was trained only on 2015 data, we used the multitemporal training dataset (2015–2018) to provide labels for training the model for 2016–2018. After training, predictions were made for 2016, 2017 and 2018 with this recurrent model by including the previous year's predicted fractions from the same model.

The temporal RF model was designed to work similarly to the Markov chain model (Section 3.2.4) by acting purely as a postprocessing model that includes temporal information. Namely, the features for training the temporal RF model were all seven land cover class fraction predictions from the basic RF model, including the given year, the previous year and the next year, therefore totalling $7 \times 3 = 21$ training features. To obtain labels for training this model, we used the multitemporal (2015–2018) training dataset. The temporal RF model was also a binary relevance model, but all classes were included as features so that the model could learn class co-occurrence information in addition to temporal information. When using this model for prediction, it would take the predictions of the basic RF model and produce a postprocessed set of land cover fraction predictions, without using any additional features. In the case of edge years (2015 and 2018) where only two adjacent year predictions were available, the centre year prediction was repeated for the missing year.

All RF models used median voting and had the output land cover fraction predictions linearly rescaled to add up to 100%.

3.2.4 Markov Chain

A Markov chain model was implemented as a postprocessing step to improve the temporal consistency of the yearly land cover fraction predictions from the basic RF model. The Markov chain functions as a model, which can be iteratively refined by updating the land cover fractions of each class per pixel.

The trajectory of each pixel over time was modelled with a Markov chain (Markov, 1907). An energy function was defined to encode known and desirable properties in the trajectory of pixels. To design this energy function, we created the graph $G = (V, U)$, where V is the

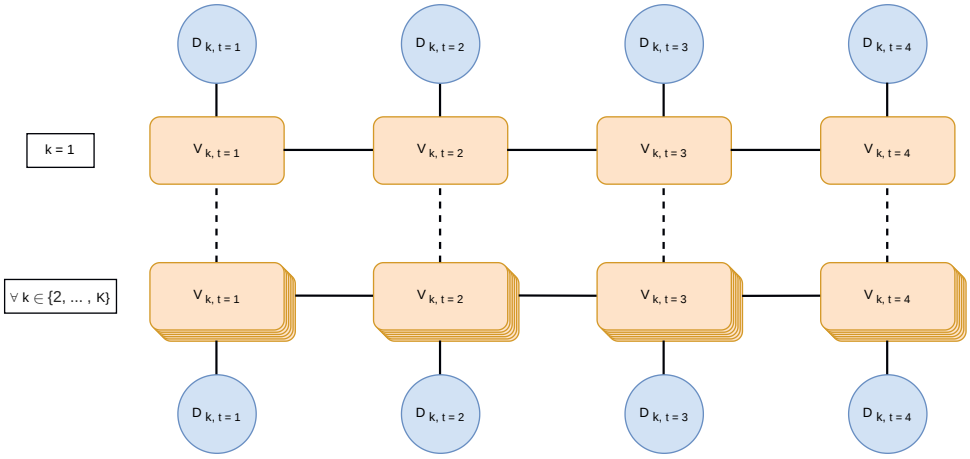


Figure 3.3: Markov chain representation used in this study illustrating the nodes and defined edges. Each node (V, indicated in yellow) is connected to its unary term (D, indicated in blue), the pairwise edges (U, horizontal lines between years) and the within-year edges (U, dashed lines) that connect to the nodes of all remaining classes ($\{2, \ldots, K\}$) of the given year.

set of nodes, each representing the land cover fraction of one class in a particular year, and U the set of edges. To model the fractions of one particular class (k) from 2015 until 2018, we define four nodes that represent the land cover fraction of each year (t). Figure 3.3 provides a schematic overview of the used Markov chain. Each node (V) is connected with its temporally adjacent node(s) and the original RF predicted fraction (D) of that corresponding year (t). Each node is also connected to the other classes' nodes to model the likelihood of co-occurring land cover classes.

Based on the defined Markov chain, a global energy function was made (see Equation 3.1), which is divided into unary, pairwise and within-year parts. The first part is defined by the sum of all dissimilarities between the original RF predicted fractions (D) and the node's fractions (V). The second part, looking at the pairwise edges between the nodes, aims to penalise fractions that deviate too much from the fractions from its adjacent year(s). To compute the energy of the entire chain from Figure 3.3, the first and second part were calculated for all years ($T = \{2015, 2016, 2017, 2018\}$) and for all classes ($K$). The third part, looking at the within-year edges between the nodes, aims to penalise classes that are unlikely to co-occur with each other. A smoothness parameter (λ) was added to balance the three constraints.

Table 3.3: Co-occurrence probability matrix showing the likelihood that one class (k) co-occurs with another class (j).

	Bare	**Crop**	**Grassland**	**Shrub**	**Tree**	**Urban**	**Water**
Bare	1	0.021	0.063	0.044	0.019	0.031	0.019
Crop	0.021	1	0.030	0.054	0.038	0.047	0.003
Grassland	0.063	0.030	1	0.200	0.099	0.057	0.020
Shrub	0.044	0.054	0.200	1	0.125	0.050	0.009
Tree	0.019	0.038	0.099	0.125	1	0.080	0.013
Urban	0.031	0.047	0.057	0.050	0.080	1	0.006
Water	0.019	0.003	0.020	0.009	0.013	0.006	1

$$
\begin{aligned}
E_{global} = \lambda_{unary} & \sum_{k \in K} \sum_{t \in T} \sqrt{(V_t - D_t)^2} \\
& + \lambda_{pairwise} \sum_{k \in K} \sum_{t \in T} \left(\sqrt{(V_t - V_{t-1})^2} + \sqrt{(V_t - V_{t+1})^2} \right) \\
& + \lambda_{within} E_{within}
\end{aligned}
\tag{3.1}
$$

To characterise the within-year edges (E_{within}), a co-occurrence matrix was developed that aims to penalise classes that have a weak correlation, and apply a lower penalty to strongly correlated classes. The class co-occurrence probability is associated with a $K \times K$ matrix, where K is the number of classes. The multitemporal training dataset (2015–2018) was used to derive this co-occurrence matrix. Table 3.3 shows the co-occurrence matrix, which is defined by two kinds of elements: the main diagonals and non-diagonals. The diagonals indicate that each class always co-occurs with itself. The non-diagonals indicate the correlation or probability ($c_{k,j}$) that one class (k) co-occurs with another class (j), derived from the multitemporal fraction training data. With a value of 0.2, the classes grassland and shrubs have the highest co-occurrence probability in the matrix, meaning that this class pair is the most likely to co-occur, and should therefore be penalised the least. Looking at each class row or column, water has the lowest probabilities of co-occurrence with other classes overall. As the optimisation process is done class by class, water should thus be relatively penalised the most as long as it is not the majority class.

To quantify the penalisation of correlating classes, the third term E_{within} was included in the global energy function, calculated using Equation 3.2. For each class pair, the negative logarithm of the correlation (c) was calculated in order to apply a higher penalty to low correlated class pairs, and a low penalty to high correlated class pairs. To compute the energy of the entire chain from Figure 3.3, all possible class pairs were summed for each year (t).

$$E_{within} = \sum_{t \in T} \sum_{k \in K} \sum_{j \in K, j \neq k} (-log(c_{k,j}) \times V_{t,k} \times V_{t,j}) \tag{3.2}$$

With these energy functions, it was possible to iteratively minimise the energy, so that the fractions become more similar over time. Well-known deterministic algorithms have been used in prior research for energy optimisation, such as Iterated Conditional Modes (ICM) (Besag, 1986) and the Gradient-Descent method (Cauchy, 1847). In this study, we used the ICM method, adapted to continuous fractions. Since the fractions are a regression problem, the nodes' values had to be iteratively updated using a discrete step size to find the optimum fraction that results in the lowest energy. The step size was determined carefully, to balance computation time with output fraction precision, and set to 1%. Similarly, the values of the smoothness parameters (weights) λ in Equation 3.1 were determined by testing and it was found that the addition of the co-occurrence information was less effective than the addition of the temporal information. Therefore, the smoothness parameters were set to $\lambda_{unary} = 1$, $\lambda_{pairwise} = 1$ and $\lambda_{within} = 0.01$. This iterative optimisation technique was done node-by-node until convergence. After this, all classes' fractions were rescaled to add up to 100%.

Time series preprocessing, further processing and analysis was performed in the software R (R Core Team, 2021), and Markov chain implementation was performed in Python (Van Rossum and Drake, 2009). All of the code created for this paper can be found in Burger and Masiliūnas (2024).

3.2.5 Accuracy Assessment

An accuracy assessment was done to test the performance of the models. In total, there are five different models from which the outputs were all validated separately: 1) basic RF, 2) recurrent RF, 3) temporal RF, 4) linear smoothing and 5) Markov chain.

3.2.5.1 Statistical metrics

To measure the accuracy of the land cover fraction models, statistics were chosen that are frequently used in other land cover fraction mapping studies, e.g. Souverijns et al. (2020): RMSE and MAE. These statistics measure the difference between the predicted and the reference land cover fractions in percentage points.

Besides these statistical metrics, a subpixel confusion-uncertainty matrix (SCM) (Silván-Cárdenas and Wang, 2008) was calculated, from which the following metrics can be derived: overall accuracy (OA), producer accuracy (PA) and user accuracy (UA). The SCM is related to the confusion matrix but adapted to account for fractional data. The SCM is able to provide detailed information on the sub-pixel confusion and uncertainty. This is done by exploring the diagonals and off-diagonals of the matrix. The diagonals indicate

the maximum overlap of the predicted and reference class fraction. The off-diagonals are indicated by confusion intervals, by calculating the minimum and maximum possible overlap of classes. The accuracies can thus be described as a centre value plus-minus its uncertainty (e.g. OA: 83.33 % ± 1.12 %).

3.2.5.2 Assessment of results

In the first step of the assessment, all yearly land cover fractions of each model were evaluated with the validation data (2015-2018) from its corresponding year. The accuracy statistics were calculated for all predictions with all classes pooled together to get the overall accuracy statistics, and for each class separately to get the per-class accuracy. Furthermore, to assess the land cover fraction change over the years, the fraction difference between each year pair (2015–2016, 2016–2017 and 2017–2018) was calculated for each class. The accuracy of the fraction change predictions was assessed by calculating the RMSE and MAE pooled overall and per-class.

Secondly, the model performance was analysed with regards to the intensity of change. The multitemporal validation dataset was separated into four change bins according to the magnitude of observed change (sum of fraction change over the time series): (1) no change (0% change, 26876 sample sites), (2) gradual or partial change (1–199% change, 1682 sample sites), (3) an abrupt change from a single class to another (200% change, 944 sample sites) and (4) change in multiple classes or multiple changes (over 200% change, 986 sample sites). The accuracy of each model was assessed for each bin separately, also for each year and for each yearly transition.

3.3 Results

3.3.1 Statistical comparison of the models

The overall accuracy statistics of the annual fraction predictions by the five models are presented in Table 3.4. The per-class RMSE accuracies of the five models are presented in Figure 3.4. From Table 3.4, it can be noted that the Markov chain model resulted in higher overall accuracy, lower RMSE and lower MAE for all the years compared to

Table 3.4: Yearly accuracy statistics of the three models. a): RMSE (%). b): MAE (%). c): SCM overall accuracy (%). Best performing model statistics are highlighted in bold.

	a) RMSE				b) MAE				c) OA			
Model	**2015**	**2016**	**2017**	**2018**	**2015**	**2016**	**2017**	**2018**	**2015**	**2016**	**2017**	**2018**
Basic RF	23.1	23.4	23.7	23.6	9.1	9.3	9.4	9.4	67.7±1.1	67.1±1.1	66.6±1.1	66.8±1.2
Recurrent RF	23.1	24.5	26.4	27.6	9.1	10.0	11.0	11.7	67.7±1.1	64.8±1.1	61.1±1.0	58.7±1.0
Temporal RF	24.9	25.1	25.2	25.3	10.1	10.3	10.3	10.4	64.6±0.9	63.8±0.9	63.7±0.9	63.5±0.9
Linear smoothing	22.6	**21.9**	**22.0**	23.0	9.1	**9.1**	**9.2**	9.3	67.9±1.2	**68.1**±1.3	**67.8**±1.3	67.1±1.2
Markov chain	**22.2**	22.2	22.4	**22.6**	**9.0**	**9.1**	**9.2**	**9.2**	**68.4**±1.2	68.0±1.3	67.7±1.3	**67.6**±1.2

the other models, except for the linear smoothing, which performed slightly better in the middle years 2016 and 2017. When all years were pooled together, Markov chain showed to have performed the best, with MAE of 9.16, which was significantly better than 9.19 achieved by the linear smoother (paired t-test $t = 16.54$, $p < 0.001$).

The Markov chain greatly improved the output from the basic RF model in all statistics and in all years. This improvement can also be noted in the per-class RMSE accuracies in Figure 3.4, where all classes consistently decrease in RMSE. Classes such as grassland, crops and trees that are relatively challenging to characterise, also benefit from the Markov smoothing approach. Classes that already have a relatively low RMSE score (i.e. built-up and shrubs) seem to benefit the least from the Markov chain.

Looking at the yearly accuracies, the models score best in 2015 and poorer in the following years. The reason for this trend is that the basic RF model was trained on the 2015 training dataset, therefore it is expected that later years are not classified as accurately. This trend is even more pronounced for the recurrent RF model. The recurrent and basic RF models are identical for the year 2015, but in later years, the previous year's prediction is taken as a feature. For all three statistics, the recurrent RF model yields poorer accuracies over these years. The permutation feature importance showed that the fraction of the previous year dominated all of the other features. Therefore, the recurrent RF model suffers from error propagation over the years due to its recurrent nature. The temporal RF model did not exhibit this issue, as it could take into account the full three-year window around the target year for its predictions. However, the temporal RF model reduced the map accuracy compared to the basic RF model for every year. The year 2017 was a crossing point where the propagated errors in the recurrent RF model lowered its accuracy below that of the temporal RF model.

Comparing the basic RF model with both the Markov chain and linear smoothing model overall statistics, it seems that the postprocessing models decrease yearly RMSE more than the yearly MAE. This suggests that these postprocessing models are better at reducing large errors, resulting in a smoother time series. This result was expected, as the RF models used median voting instead of mean voting. Median voting leads to predicting more pure pixels (either 0% or 100%) (Masiliūnas et al., 2021a). This results, for example, in more correct 0% predictions, but at the same time increases the chance of getting false 100% predictions. The Markov chain and linear smoothing models tend to smooth the fractions and therefore reduces large errors.

Table 3.5 shows the RMSE and MAE of the models' predictions of the year-to-year change. The Markov chain model performed the best of all the tested models, as it successfully reduced spurious year-to-year change. The year-to-year change results also show a pattern that the transition between 2015 and 2016 had a considerably higher accuracy than the later year pairs. This pattern seems to be a combination of the higher accuracy of prediction for the year 2015, therefore also improving year-to-year change accuracy, as well

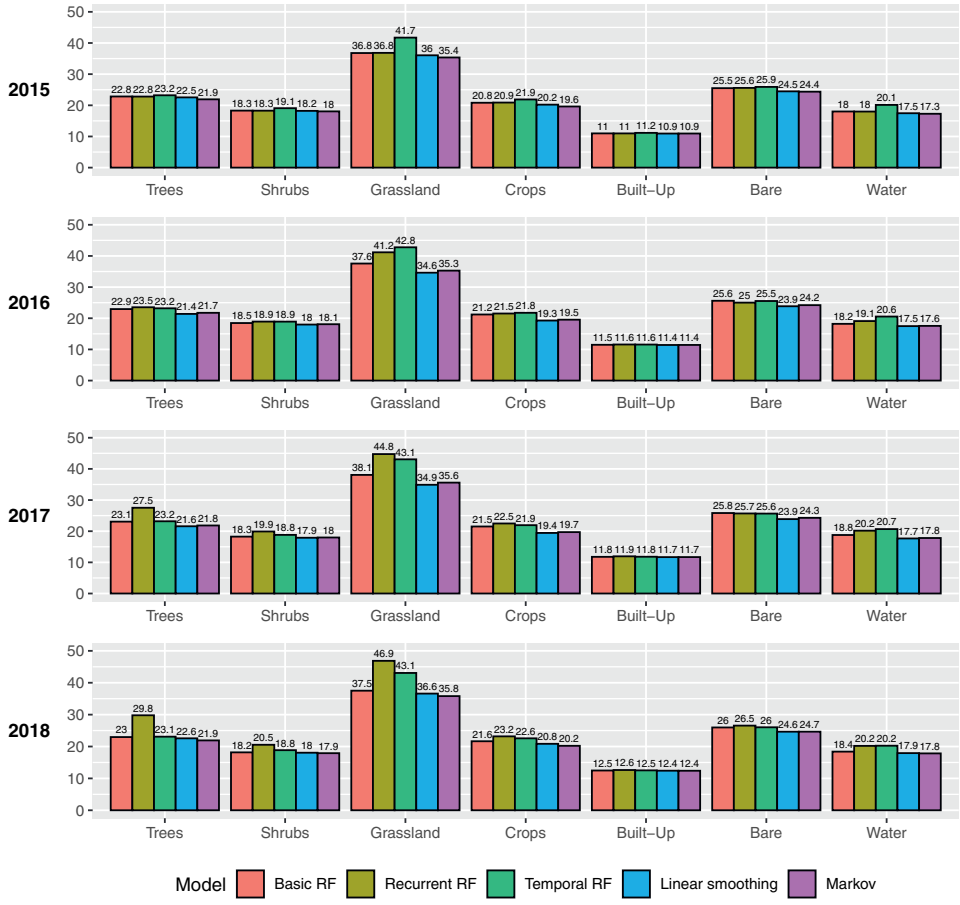


Figure 3.4: Per-class RMSE (%) of the yearly predictions of the five models, separated per year.

Table 3.5: Year-to-year fraction change accuracy statistics of the five models. a): RMSE (%). b): MAE (%). Best results are highlighted in bold.

a) RMSE				b) MAE			
Model	**2015-2016**	**2016-2017**	**2017-2018**	**Model**	**2015-2016**	**2016-2017**	**2017-2018**
Basic RF	15.0	16.1	16.6	Basic RF	4.1	4.5	4.6
Recurrent RF	17.0	14.2	13.5	Recurrent RF	5.5	3.9	3.5
Temporal RF	14.3	15.8	14.8	Temporal RF	3.5	4.3	3.8
Linear smoothing	9.1	10.7	10.6	Linear smoothing	2.4	2.9	2.7
Markov chain	**8.8**	**10.6**	**10.5**	Markov chain	**2.1**	**2.7**	**2.6**

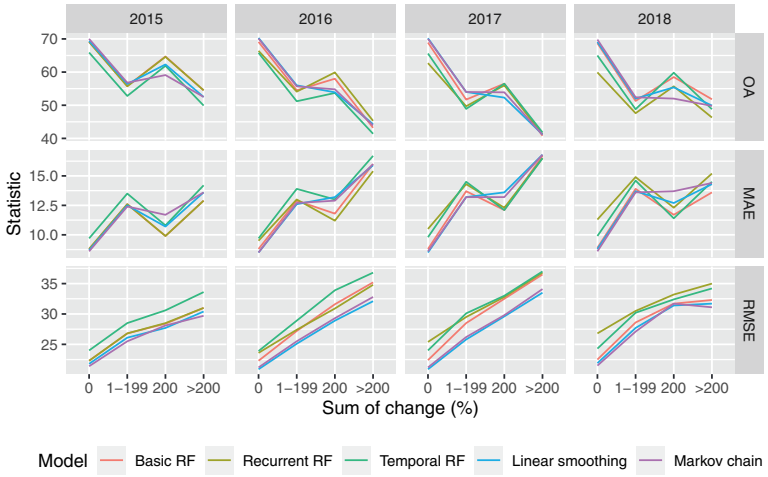


Figure 3.5: Yearly accuracy statistics of the five tested models, separated per amount of change observed in each validation site. The sum of change was calculated as absolute sum of actual fraction change (all classes) throughout the four years (2015–2018) of each validation site.

as the fact that fewer instances of change have been registered between 2015 and 2016 (2015–2016: 1652, 2016–2017: 2265, 2017–2018: 2331 sample sites with change). This pattern, however, did not hold for the recurrent RF model. In fact, it had the opposite effect, as the accuracy of later year-to-year transitions was higher than the earlier ones. Here again the transition between 2016 and 2017 was a cross-over point, though now after this point the recurrent RF model performed better than the temporal RF model. This result shows that a very conservative model is more useful for reducing spurious change, even if the yearly maps themselves reduce in accuracy. Nevertheless, even though in terms of fraction change, the RF postprocessing models improved the basic RF predictions, they did not achieve as good results as the linear and Markov chain models.

3.3.2 Model performance based on fraction change amount

The models have different performance in sample sites that exhibited no change and areas that exhibited varying degrees of change, as seen in Figure 3.5. The overall pattern shows that the models had less difficulty predicting land cover fractions in no change areas than change areas. Within change areas, it was easier to predict land cover where only one pure class transitioned into one other pure class (a change of 200%), rather than in areas with partial and mixed transitions. Areas with higher amounts of change were more difficult to predict correctly, which is reflected in the increasing RMSE as the amount of change increases. Linear smoothing and Markov chain did better in areas with no change and

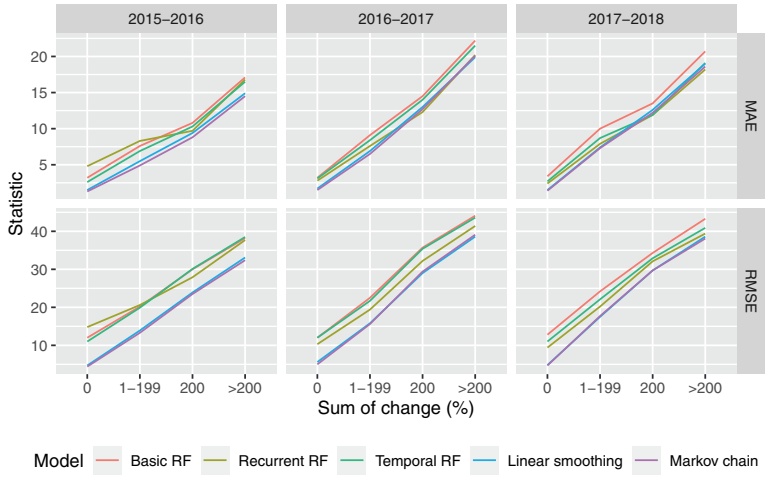


Figure 3.6: Year to year error statistics of the five tested models, separated per amount of change observed in each validation site. The sum of change was calculated as absolute sum of actual fraction change (all classes) throughout the four years (2015–2018) of each validation site. An SCM cannot be computed for change, as change events do not add up to 100%, therefore no OA calculation is available.

mixed change compared to the other models, and they especially reduced large errors. These smoothing methods did worse in areas of a single pure change, as they tended to oversmooth this abrupt change. None of the RF models were consistently better at predicting the fractions in areas with a single pure change, as their accuracy varied among the years. The recurrent RF model was better at the start of the time series and worse at the end.

The year-to-year change statistics per change amount (Figure 3.6) show that correctly predicting change in areas with a large amount of change is challenging. Characterising pure change correctly was not easier than smaller amounts of gradual or mixed change, in contrast to the results of yearly accuracy statistics. The results also show that linear smoothing and especially Markov chain performed well in all change areas compared to the other models. Only the recurrent RF model managed to achieve lower MAE for the abrupt single change case and the highly mixed change case in the later year pairs. But the recurrent RF model was also the only model that did not always improve upon the basic RF predictions: in the 2015–2016 transition, its no change and gradual change (1-199%) predictions were worse than the basic model predictions.

3.4 Discussion

3.4.1 Comparison of model performance

The results of this study (Tables 3.4 and 3.5) showed that postprocessing yearly land cover fraction predictions can yield substantial improvements in their accuracy and fraction change accuracy, especially when using the Markov chain approach. However, there were large differences between the tested models. To better understand the behaviour of the different approaches, Figure 3.7 gives several examples of land cover fraction time series, the basic RF model predictions and its postprocessed time series.

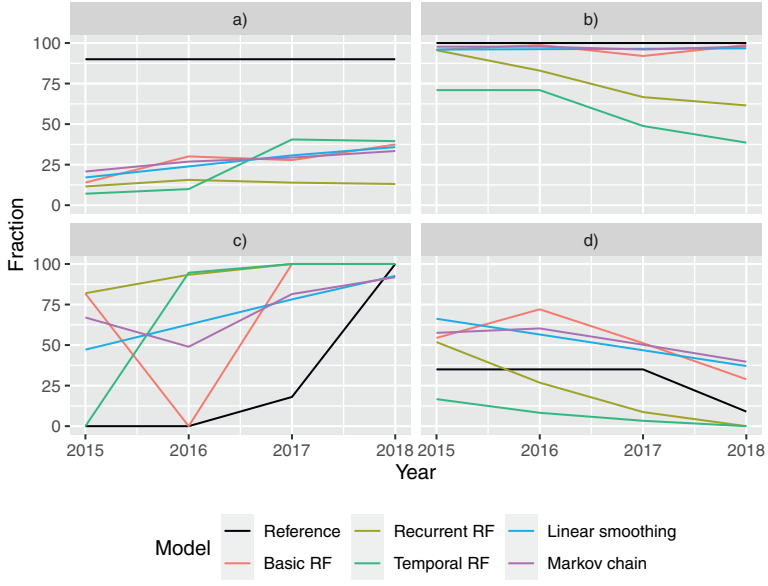


Figure 3.7: Example time series of the five tested models, no change (a-b) and change (c-d). a) Shrub fraction in a shrubland in Eritrea (lat: 14.51299, lon: 36.95253). b) Tree fraction in Anhui, China (lat: 30.2037, lon: 118.5329). It did not experience change in 2015–2018, though it has since been deforested. c) Water fraction of a shallow part of Eagle Lake, California, United States (lat: 40.71386, lon: -120.6944). d) Tree fraction in a tree plantation in Calvão, Portugal (lat: 40.49238, lon: -8.707942).

The time series trajectory examples show that the basic RF model tends to overestimate change, as it consistently predicts some change even when there is none, or more change than actually occurred. In turn, the lack of consistency results in lower accuracy of year-to-year change, as evident from Table 3.5. The recurrent RF model starts at the same point as the basic RF model in 2015, but afterwards diverges. It tends to be conservative with change and retain values close to the initial year, therefore smoothing large changes

over (e.g. Figure 3.7 (a) and (c)), however, it also sometimes creates change where none has occurred (e.g. Figure 3.7 (b) and (d)). This conservative behaviour leads to the accumulation of errors and therefore a decrease in yearly map accuracy over time, as seen in Table 3.4 and Figure 3.4. The temporal RF model behaves similarly to the recurrent RF model, but it tends to produce even more change, especially since it is not bound to basic RF predictions at the start of the time series. The behaviour of introducing change explains the relatively poor performance of these models in terms of yearly map accuracy seen in Table 3.4. Nevertheless, both models overall reduced the amount of predicted change and obtained better change accuracy (Table 3.5). Both the recurrent and temporal RF models generally exhibit similar trajectory patterns, which likely comes from the training data used for both models, namely the multitemporal (2015–2018) training dataset. It appears that the models are overfitted on this dataset, or that the dataset is not representative enough of the validation dataset. While these models did not perform well in our study, there may still be some merits in exploring the ideas further, e.g. by training these techniques on more representative and larger datasets, or by having the models predict change rather than the fraction directly, which could prevent the overfitting issue due to a larger number of reference points.

The linear smoothing model and the Markov model performed better than the basic RF model, and generally had similar patterns. These models reduce the amount of change in the time series, by either reducing the whole time series to a line (linear smoothing) or reducing the change extremes towards the mean (Markov chain). Figure 3.7 (c) is a good example of Markov chain smoothing, where the temporal pattern from the basic RF model is still visible, but the change is reduced. This allows the overall amount of change to be reduced, but abrupt change in the time series to still be represented.

The linear smoother performed slightly better than Markov chain for the middle years of the time series (Table 3.4). This result was surprising, also given that for these years Markov chain generally performed better in areas that experienced one pure class change to another pure class. The result seems to mostly come from areas with no change, as well as the further reduction of large errors (stronger smoothing) in the middle years (Figure 3.5). At the start and end of the time series, Markov chain performed better than linear smoothing, as it is less bound to the values predicted by the basic RF model. Linear regression is well known to be highly affected by outliers (Hampel et al., 2005), in this case, the fraction values at the years 2015 and 2018 have a higher leverage on the linear regression.

The optimal Markov model used $\lambda_{within} = 0.01$, which gives a low weight to the effects of the co-occurrence matrix (Table 3.3). This result is likely linked to the issue that the RF postprocessing models have experienced, namely, that the multitemporal training dataset is not necessarily representative of the validation dataset, also in terms of class co-occurrence. The multitemporal training dataset used a clustered sampling design at the

global level, i.e. selectively targeting regions and with a higher emphasis on Africa to focus collection efforts on areas that are known to have higher incidence of change (Figure 3.2), however, that means that large areas of the world are entirely unrepresented. In contrast, the validation data uses a stratified random sampling design across the globe and is more representative of actual change on the ground.

Overall the Markov chain produced the best results of all the tested models. Even though the performance of the Markov chain was similar to that of the linear smoother, nevertheless it achieved significantly better results overall. This result shows that the Markov chain is a promising postprocessing method for land cover fraction maps.

3.4.2 Comparison with other studies

This study builds upon the work of Masiliūnas et al. (2021a), who used RF regression and other machine learning algorithms to produce a global land cover fraction map in 2015. The same 2015 fraction training data was used in both studies, but this study used annual validation data from 2015 until 2018, whereas Masiliūnas et al. (2021a) particularly used validation data from 2015. They reached an overall accuracy of 72% using RF regression with median voting and a three-step model. This study managed to get the highest overall accuracy (68.4%) in 2015 using the Markov chain model. This difference in overall accuracy can be explained by the fact that this study used the annual validation dataset from 2015 until 2018, as this dataset contains a larger proportion of land cover fraction samples that experienced change, which are often difficult to capture well. In addition, this study lacks climate, soil and terrain features, which were used by Masiliūnas et al. (2021a) and could certainly improve the model performance. However, the focus of this study was to compare land cover fraction postprocessing models to improve yearly predictions and changes in fractions, rather than to achieve the highest absolute overall accuracy. The remote sensing–derived features are the most important for an RF regressor and therefore are sufficient for the model comparison. In production of a land cover product, additional features would be used to improve the final result further.

There are few studies on using RF models as a postprocessing method, and our study is the first one to compare RF models with a Markov chain postprocessing approach. RF was used as a postprocessing model in Xu et al. (2022) to reduce falsely detected changes from a change detection algorithm by ensembling the results of several change detection models and runs. RF showed promise in their study, as it successfully reduced the number of false changes and resulted in a better balance between user's and producer's accuracy. Our results (Table 3.5) also showed that RF in most cases, and temporal RF in all cases, improves the change accuracy. However, it was less effective than a linear smoother. Temporal RF is more similar to the RF implementation in Xu et al. (2022) than recurrent RF, but nevertheless there are considerable differences between the studies due to different objectives and methods that makes the results difficult to generalise. In

addition, RF postprocessing methods decreased the map accuracy in every year (Table 3.4). Therefore, further research is needed if RF could be used as an effective postprocessing model.

Several studies have focused on the implementation of Markov models for land cover maps, and generally successfully increased the accuracy of land cover maps. However, different Markov implementations have been used, for instance including both spatial and temporal context (Cai et al., 2014; Wehmann and Liu, 2015; Yang et al., 2021), or just spatial (Kasetkasem et al., 2005; Xie et al., 2022) or temporal context (Abercrombie and Friedl, 2016; Souverijns et al., 2020), and with or without co-occurrence dependencies (Wang et al., 2017). Moreover, all of the above-mentioned studies focus on discrete land cover mapping, while in this study the Markov chain was applied on land cover fractions. Therefore, in terms of accuracy numbers, the results of this study cannot directly be compared to other studies that used a Markov model. However, generally Markov model optimisation has a great potential to reduce spurious and inconsistent multitemporal land cover change, and this study confirms this principal idea. The findings highlight the importance of the use of Markov models for land cover mapping studies. In addition, Markov models can be constructed quite flexibly by including different terms, and the penalisation of edges can be calculated in accordance with the user's needs. Markov models can also be used for a variety of multitemporal problems. While this study only focused on four single years, Markov models can also be applied to time series with a longer time period or on time series with a shorter interval. As satellite imagery products continue to grow and land cover time series become more common, it is essential to ensure that the multitemporal predictions provide realistic and consistent land cover change.

3.4.3 Recommendations

Although the Markov chain model obtained higher accuracy than the other tested models after postprocessing, it is still essential to explore practices that further improve the accuracy of the base model's initial prediction. Additional features used by Masiliūnas et al. (2021a), such as climate and soil metrics, could be included in further research in order to improve the distinction between certain land cover classes. Features such as precipitation, temperature and soil properties could considerably be associated with the presence of certain land cover classes. However, spectral data most likely remains the most important element for achieving accurate land cover maps.

Using the Landsat 8–derived features, it was possible to obtain initial land cover fraction predictions. But, due to temporal filtering of outliers and cloud masking in combination with a 16-day revisit interval, the harmonic features extracted from the 3-year temporal window may not be robust enough, as the fit relies on the three years having sufficient data to fit a harmonic seasonality model with three orders without overfitting. Therefore it might be challenging for the RF models to produce consistent predictions. Additional

spectral sources and longer time series could be considered in future studies to improve prediction consistency. Landsat 9 imagery could help more recent land cover mapping studies to obtain an 8-day revisit interval. A combination of Landsat and Sentinel-2 sensors, such as the NASA Harmonized Landsat and Sentinel-2 (HLS) product (Claverie et al., 2018) or the ESA Sen2Like (level-2H products, Saunier et al. 2022) could be a strategy to get even more robust data.

The Markov chain model worked well, showing that explainable AI models can lead to a substantial improvement of land cover time series. The rules used to define the Markov chain were relatively simplistic, and therefore it is easy to understand and track back the reason for any change the Markov chain applies to the basic model predictions. Additional rules could be added to the model while keeping it explainable. Firstly, the model could be further expanded with a transition probability matrix. For each class, yearly transitions are computed independently from other classes in this study. However, it could be beneficial to examine the potential of penalising uncommon class-transitions in further research. In addition, information from the spatial context could be useful to explore. No spatial neighbouring pixels were used in this study, however, other studies that incorporated spatial information in Markov models have been shown to improve the accuracy of land cover maps (Cai et al., 2014; Kasetkasem et al., 2005; Wang et al., 2015; Xie et al., 2022). Therefore, it is recommended for follow-up studies to examine the potential of spatial information for land cover fraction and fraction change mapping. HMMs and MRFs would be suited for this task, but they need to be adapted for handling fractions rather than discrete classes. Another important consideration is that the expansion of Markov models could also bring additional complexity. The different components and their smoothing parameters need to be determined carefully. In addition, more complex models will usually have a longer computation time and higher computer memory demands, especially with the addition of spatial information, which would require at least a nine-fold increase in input data size.

One limitation of the Markov chain model in this study is that even though it can follow the change patterns of the basic RF model, it nevertheless smooths out abrupt change, as we see from the poorer results for 200% change samples in Figure 3.5. To solve this issue, a combination with a time series segmentation and change detection algorithm, e.g. from the BFAST family of change detection algorithms (Verbesselt et al., 2010a), could be beneficial. A simple approach would be to build separate Markov models for areas with detected change and areas without detected change, which would have different weights for the pairwise and unary terms, optimised for change and no change. A more complicated approach would be to analyse stable and change segments within the time series. In time series segments that have no detected change, a Markov model could smooth out the time series to make it more consistent. On segment boundaries where change is detected and segments with rapid change, a Markov chain with additional rules on transition likelihood could be used to confirm whether the detected change is true or

spurious, and therefore reduce the commission errors of the change detection algorithm. If both models agree that the change is likely, the full change would get preserved, and therefore an abrupt change would be visible between the pair of years. As Markov rules are adaptable, such a framework could be made of two separate Markov models, optimised for different outcomes (smoothing vs confirming change). Lastly, Markov models could be adapted run on temporally dense time series, such as the original Landsat 8 time series without yearly compositing. More research and more detailed reference data are needed to determine if the added temporal information could prove beneficial in a dense Markov model.

3.5 Conclusion

The main objective of this research was to improve the temporal consistency of land cover fraction change by exploring several temporal postprocessing models. The results indicate that the Markov chain model achieved better overall and change accuracies of land cover fractions compared to the other models. The basic RF model produced initial fraction predictions, but the Markov chain postprocessing approach consistently improved the overall and per-class accuracies. The recurrent RF model obtained considerably poorer results, especially in the later years. The temporal RF model also obtained higher errors in yearly predictions compared to the basic RF model, but it did not lose accuracy over time and it reduced errors in predicted land cover fraction change, although not as well as Markov chain and linear smoothing. The linear smoothing technique performed well in improving yearly land cover fraction predictions, but it was overall still significantly worse than the Markov chain model, especially in terms of year-to-year land cover change. The Markov chain model has the ability to smooth large interannual variability, which forms a great combination with RF median voting that primarily predicts extreme fraction values of 0% and 100%. These results confirm that Markov postprocessing has the potential to reduce spurious and inconsistent multitemporal land cover change, which is in line with other studies. The Markov chain model could be expanded in further research by exploring the potential of penalising uncommon class-transitions and by accounting for the relationships of spatial neighbouring pixels. Furthermore, the Markov chain model could be explored in combination with a change detection algorithm in order to improve the monitoring of land cover change. The findings highlight the importance of the addition of Markov models for future land cover mapping studies, so that multitemporal predictions provide realistic and consistent land cover fraction change, essential for consistent and user-centric global land cover fraction mapping.

Chapter 4

BFAST Lite: a lightweight break detection method for time series analysis

This chapter is based on:

D. Masiliūnas, N.-E. Tsendbazar, M. Herold, and J. Verbesselt (2021b). "BFAST Lite: A Lightweight Break Detection Method for Time Series Analysis". *Remote Sensing* 13.16, 3308. DOI: 10.3390/rs13163308

Abstract

BFAST Lite is a newly proposed unsupervised time series change detection algorithm that derives from the original Breaks For Additive Season and Trend (BFAST) algorithm, focusing on improvements to speed and flexibility. The goal of the BFAST Lite algorithm is to aid the upscaling of BFAST for global land cover change detection. In this paper, we introduce and describe the algorithm, and then compare its accuracy, speed and features with other algorithms in the BFAST family: BFAST and BFAST Monitor. We tested the three algorithms on an eleven-year-long time series of MODIS imagery, using a global reference dataset with over 30 000 point locations of land cover change to validate the results. We set the parameters of all algorithms to comparable values, and analysed the algorithm accuracy over a range of time series ordered by the certainty of the input time series having at least one abrupt break. To compare the algorithm accuracy, we analysed the time difference between the detected breaks and the reference data to obtain a confusion matrix and derive statistics from it. Lastly, we compared the processing speed of the algorithms, using both the original R code as well as an optimised C++ implementation for each algorithm. The results showed that BFAST Lite has similar accuracy with BFAST, but is significantly faster, more flexible, and can handle missing values. Its ability to use alternative information criteria to select the number of breaks resulted in the best balance between user's and producer's accuracy of detected change of all the tested algorithms. Therefore, BFAST Lite is a useful addition to the BFAST family of unsupervised time series break detection algorithms, which can be used as an aid in narrowing down areas with changes for updating land cover maps, detecting disturbances or estimating magnitudes and rates of change over large areas.

Keywords. Time series, land cover, change detection, BFAST, MODIS

4.1 Introduction

The lengthening of remote sensing satellite data archives is opening up new opportunities for the field of Earth Observation. It is now possible to monitor changes of the Earth's surface better than ever before, as long time series facilitate data-driven analysis methods. A long time series provides information about the usual variability over time within a monitored area, providing an opportunity to detect deviations from the norm in near real time. It also gives the opportunity for change detection algorithms to identify historical changes with more confidence.

These developments have resulted in an increase in the number of algorithms for change detection in satellite imagery time series. Algorithms such as LandTrendr (Kennedy et al., 2010), Breaks For Additive Season and Trend (BFAST) (Verbesselt et al., 2010a) and BFAST Monitor (Verbesselt et al., 2012), Continuous Change Detection and Classification (CCDC) (Zhu and Woodcock, 2014), Exponentially Weighted Moving Average Change Detection (EWMACD) (Brooks et al., 2014), Time-Series Classification approach based on Change Detection (TSCCD) (Yan et al., 2019) as well as Jumps Upon Spectrum and Trend (JUST) (Ghaderpour and Vujadinovic, 2020) have been introduced with the aim of aiding the efforts of land cover change detection.

Some of these algorithms have been well established and widely used, e.g. the BFAST family, LandTrendr and CCDC. Recent developments on these algorithms include improvements to upscaling and ease of use, such as implementing them on big data platforms like Google Earth Engine (Hamunyela et al., 2020; Kennedy et al., 2018). In contrast, many other change detection algorithms are at the proof-of-concept stage. Such proof-of-concept algorithms are not yet accessible as widely, as they either do not have a publicly available, efficient and easy to use implementation, or have not yet been proven at a large scale and using real remote sensing data. As a result, it is complicated for end-users to adopt proof-of-concept algorithms, whereas optimising them and making them user-friendly requires considerable software development effort. Thus, focusing development efforts on improving algorithms that are well-established already is beneficial for rapid dissemination of the improvements to a wide user audience.

The BFAST family of algorithms has proven to be a particularly popular choice for change detection in satellite imagery time series. The BFAST family includes the namesake BFAST (Verbesselt et al., 2010a), focused on detecting multiple breaks in a time series; BFAST Monitor (Verbesselt et al., 2012), focused on detecting a single break at the end of the time series for near-real time change monitoring; and BFAST01 (de Jong et al., 2013), focused on characterising the trajectory of the time series around its largest break. These algorithms have been successfully used in many studies, ranging from semi-arid regions of Australia (Watts and Laffan, 2014) to various ecozones in Canada (Fang et al., 2018), forest disturbance in the Colombian Andes (Murillo-Sandoval et al., 2018) and turning

point characterisation in sub-Saharan drylands (Bernardino et al., 2020). The `bfast` package has even become a basis for new change detection packages that implement change detection frameworks, such as `TSS.RESTREND` (Burrell et al., 2017) and STEF (Hamunyela et al., 2016).

However, there are several limitations of the original BFAST algorithm that affects its uptake by the users. The original BFAST algorithm, as proposed by Verbesselt et al. (2010a) and implemented in the `bfast` R package, consists of several stages, of which one requires decomposition of the input time series into trend and season components. This decomposition step in turn requires a complete regular time series with no missing data, which is rare when using optical satellite imagery due to cloud cover. In addition, BFAST uses an iterative approach for convergence of change, in both the seasonal and trend components. This approach takes a significant amount of processing time.

To overcome these limitations of the original BFAST algorithm, in this paper we introduce a new unsupervised change detection algorithm, called BFAST Lite. BFAST Lite aims to improve upon the original BFAST in terms of speed and flexibility. This is achieved by omitting the decomposition of the time series entirely, using a multivariate piecewise linear regression to do model fitting in a single step. Therefore, BFAST Lite can handle missing data in the time series without any interpolation. It is also much faster, as it omits both the computationally demanding decomposition step, and the iterative model refitting approach. Speed improvement is important for upscaling break detection to larger areas, especially for global scale; whereas better handling of missing data is particularly important at locations where cloud cover is common. In addition, BFAST Lite has more tunable parameters, which enables calibration of the algorithm for particular locations, or for use with machine learning algorithms to combine multiple BFAST Lite models for automatically increasing change detection accuracy.

BFAST Lite is easily accessible for users as a function in the version 1.6 of the `bfast` R package. BFAST Lite is particularly simple to use for current users of the `bfast` package, who are already familiar with other BFAST algorithms provided by the package.

The objective of this paper is to detail the implementation of the BFAST Lite algorithm and compare it with the existing algorithms in the BFAST family. In this paper, we first described the implementation of the BFAST Lite algorithm, the differences from the BFAST algorithm, and its tunable parameters. We then compared BFAST Lite to two existing algorithms in the BFAST family, BFAST and BFAST Monitor, in terms of break detection accuracy and processing time. For this purpose, we used a global land cover change reference dataset created as part of the CGLS-LC100 project, which includes information about all land cover transition types.

4.2 Data and methods

4.2.1 The BFAST Lite algorithm

The BFAST Lite algorithm is an adaptation of the BFAST algorithm as described in Verbesselt et al. (2010a). As a brief overview of the original algorithm, BFAST is an unsupervised time series change detection algorithm, specialised in detecting multiple breakpoints within a multi-year time series. It works by first decomposing an input vegetation index time series Y_t into trend (T_t), seasonal (S_t) and error (ϵ_t) components using Seasonal decomposition of Time series by LOESS (STL) (Cleveland et al., 1990). Next, the trend and season components are tested for at least one significant break in the whole time series using an ordinary least squares residual moving sum (OLS-MOSUM) statistical test. If there is significant evidence ($p < 0.05$) of a break in either the trend or season component, then a process of fitting a univariate piecewise linear regression to determine the locations of the breakpoints (as defined by Bai and Perron (2003)) is carried out for each of the components. The decomposition and fitting process is repeated iteratively until convergence is reached.

In contrast, BFAST Lite involves only two steps. The first is the generation of regressors for the piecewise linear regression. This is done by curve fitting on the input time series: the trend is modelled by a monotonous line, harmonics by a sine and a cosine component for each harmonic order (Jakubauskas et al., 2001), and seasonality by seasonal dummies (Makridakis et al., 1997). There is also an option to include autoregressive (lag) terms, as well as to manually include external regressors, for example spatial neighbourhood features as presented by Dutrieux et al. (2015). This step is highly user-configurable. The user chooses which of these regressors to include in the model to be fit on the time series. In addition, the harmonic component order and the number of seasonal dummies per year are both user-configurable parameters.

The second step is breakpoint estimation following the approach of Bai and Perron (2003), implemented by Zeileis et al. (2003), and equivalent to the last step within an iteration in the original BFAST algorithm. To estimate breakpoint timing, this approach creates a number of piecewise linear regressions, where each segment has its own estimates for the regressor coefficients. The break timing is determined by minimising the residual sum of squares (RSS) of the piecewise model, repeated for each possible number of breaks in the time series. The optimal number of breaks is then selected using an information criterion. Whereas BFAST prescribes the use of the Bayesian Information Criterion (BIC), BFAST Lite defaults to a more conservative metric developed specifically for piecewise linear regression by Liu et al. (1997) (LWZ). While the information criterion of Liu, Wu and Zidek (LWZ) is the default, BFAST Lite allows the user to choose between LWZ, BIC, Akaike's Information Criterion (AIC), and minimum RSS. Support for custom user-defined information criteria is planned for future releases.

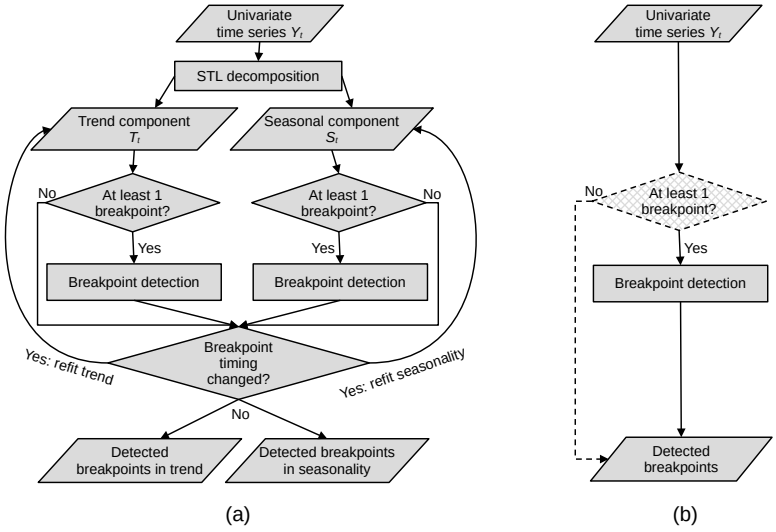


Figure 4.1: A schematic overview of the algorithms: (a) BFAST, (b) BFAST Lite. Dashed line indicates an optional step that is disabled by default. The breakpoint detection step (Bai and Perron, 2003) involves the following substeps: generation of components for model fitting, fitting the model itself, determining all possible breakpoints and selecting an optimal number of breakpoints.

The difference between the BFAST and BFAST Lite algorithms is depicted in a simplified flowchart in Figure 4.1. The design of BFAST Lite leads to several differences from the original BFAST algorithm. First, BFAST Lite implements the approach of Shao and Campbell (2002), in which the time series season and trend components are modelled in a multivariate piecewise linear regression, instead of using STL for decomposing the time series like the original BFAST. This is advantageous for two reasons: 1) it alleviates the need of an iterative decomposition step, making the algorithm faster; 2) it negates the issue of handling missing data in the time series, because a (piecewise) linear regression does not require regular interval time series. The drawback of BFAST Lite compared to BFAST is that the detected breaks can no longer be separated into seasonal and trend breaks.

Another difference is that, unlike BFAST, BFAST Lite does not prescribe the use of a structural change test (e.g. OLS-MOSUM) for detecting whether at least one breakpoint is present. It is left as an option to the user and not enabled by default. On one hand, a preliminary structural change test can be beneficial to further reduce processing time by skipping time series that do not have significant breaks. On the other hand, such a test may also cause an increase in omission error, as it is very fast but not as accurate in determining breaks as the much more sophisticated BFAST algorithms. If the structural

change test produces a false negative, none of the real breaks in the time series will be detected. Therefore, BFAST Lite gives users the freedom to not use a test, or to select any test (not limited to OLS-MOSUM) from the `strucchangeRcpp` package.

Lastly, BFAST Lite output includes new, more robust statistics to indicate the magnitude of each break. BFAST provides magnitude only based on the difference between the modelled values at the time step immediately prior to a detected break and a time step immediately following the detected break. This may lead to spurious magnitude values, as the difference between the two time steps may be influenced by a different phase in the modelled seasonality, causing either an exaggerated or a modulated magnitude value depending on the alignment of the seasonality between the adjacent time series segments. In contrast, BFAST Lite provides two more robust metrics of magnitude: the root mean squared deviation and the mean absolute deviation between the model predictions of the adjacent segments, over the time span of one year before and after the detected break. The mean deviation is also provided, and gives an indication of the change direction, that is, whether the index values increase or decrease after the break.

The BFAST Lite algorithm has been published as a function (`bfast::bfastlite()`) implemented in R in the free and open-source `bfast` package, starting from version 1.6.0, available on CRAN. The source code can be found on GitHub, at `https://github.com/bfast2/bfast` (Masiliūnas et al., 2021c). The package is now maintained by volunteers, with GitHub infrastructure facilitating more rapid update cycles, code review and continuous integration. Users are encouraged to submit bug reports and ideas for improvement via GitHub.

4.2.2 Testing the performance of BFAST Lite

4.2.2.1 Compared algorithms

We compared the performance of BFAST Lite, in both break detection accuracy as well as run time, with two other well known BFAST algorithms: BFAST and BFAST Monitor. For simplicity, all parameter values for the three algorithms were left at their default values, where applicable. The BFAST Lite default values were set to correspond to the default values of BFAST. BFAST has few customisable parameters, most parameters are preset and not adjustable; thus, for comparability, we adjusted BFAST Lite parameters to match those of BFAST. Whereas BFAST prescribes the use of BIC for selecting the optimal number of breaks, in this study we tested BFAST Lite using both criteria: BIC and the newly implemented LWZ. For both BFAST and BFAST Lite, the minimal segment size h was set to the number of observations per year. The models were fitted using both a trend component and a seasonal component. The seasonal component was represented by three orders of harmonics, that is, sine and cosine waves at the frequency of one year, half a year, and one third of a year. BFAST mandates the harmonic orders to be set to these three orders, whereas BFAST Lite allows the user to choose any number of orders.

We chose to use three orders with BFAST Lite as well for comparability with BFAST. See Appendix 4.A for a detailed list of parameter differences between the two algorithms.

The BFAST algorithm cannot handle missing values by default; thus, we tested two solutions. One is to impute the missing values by linear interpolation between the two neighbouring values in time. This has been the most common approach to handling missing values in research so far, owing to its simplicity (Awty-Carroll et al., 2019; Fang et al., 2018). The second is to use a more advanced imputation method implemented by Hafen (2016) in the package `stlplus`, which reconstructs the missing values by taking also the values in the same season within neighbouring years into account. This option has also been newly implemented in the `bfast` package for the original BFAST function, starting from version 1.6.0. This imputation approach results in a smoother time series without interruptions in seasonality, compared to the linear interpolation approach.

The BFAST Monitor algorithm (Verbesselt et al., 2012) uses a very different approach to break detection compared to BFAST and BFAST Lite. It is designed for detecting a single break at the end of a time series. BFAST Monitor works by splitting the time series into a stable history period, that contains usual variability without breaks, and a monitoring period, within which the algorithm attempts to detect a break. A linear model is fit on the stable history period, extrapolated over the monitoring period, and the predictions are compared with actual data in the monitoring period. While direct comparison between BFAST (Lite) and BFAST Monitor is difficult due to these differences, BFAST Monitor has been successfully used in the past to detect multiple breaks. This was achieved by running BFAST Monitor iteratively with successive monitoring periods, e.g. yearly (Dutrieux et al., 2016).

In this study, BFAST Monitor was run yearly, over the three years for which we had reference data. The model values were left at their defaults when possible. The linear regression components (regressors) and harmonic order were set to match the other models. When testing the run time of the algorithms, we considered the sum of all three runs of BFAST Monitor to represent the run time of the algorithm.

With the release of the `bfast` version 1.6.0, a new backend engine written in C++ was introduced. It replaces the previous R engine by default, with the option to switch back to the R engine if desired by the user. This change was done to substantially decrease the time needed to run all of the BFAST family algorithms.

The run time of the algorithms was measured using the `microbenchmark` package (Mersmann, 2019). It runs all of the algorithms interleaved, to avoid any effect of random busy periods of computer activity on the benchmark results. It also repeats the benchmark multiple times to ensure robust results.

For time benchmarking purposes, we selected a smaller subset from the reference data (see section 4.2.2.3) using principal component analysis, to retain a diverse dataset. We further

filtered out points with too many missing values that prevented any of the algorithms from running successfully, resulting in a set of 288 points. Each algorithm was run five times on this dataset, and the time needed to run was compared between the algorithms. We included both interpolation options for BFAST, and tested both the iterative approach as originally proposed by Verbesselt et al. (2010a), and running the algorithm for a single iteration only. The benchmark was run on a virtual machine on the Terrascope cluster, running on an Intel Xeon (Skylake) processor with 4 cores and 8 gigabytes of memory. The benchmark was run sequentially on a single thread.

4.2.2.2 Satellite data

BFAST algorithms are flexible with regards to the input time series that is required for carrying out break detection within it. For best results, a vegetation index time series derived from a long archive (more than five years) of satellite imagery should be used. The input time series should have as dense observations as possible (at least monthly) to better model the seasonality, although yearly observations can also be handled if no seasonality is modelled. The observations should also be cloud-free, if necessary by performing temporal outlier removal. This way, changes in vegetation can be detected with less uncertainty thanks to the length of the time series, and with higher temporal accuracy thanks to dense observations. The time series may be irregular, although specifically the original BFAST algorithm requires a gap-filled time series with no missing values.

In this study, as input to the algorithms, a 16-day composite time series of MODIS surface reflectance at 250 m (MOD09Q1 v006) was used. The time series covered the period from 2009 until 2020. The imagery was preprocessed to remove pixels marked as clouds, cloud shadows, ocean, snow and invalid/missing using the included product state quality assurance layer. Next, the NIR_v vegetation index (Badgley et al., 2017) was calculated from the surface reflectance, using the formula:

$$NIR_v = \frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}} \cdot \rho_{nir} \tag{4.1}$$

In this formula, ρ is surface reflectance at certain wavelengths: ρ_{nir} at 860 nm and ρ_{red} at 645 nm. NIR_v was chosen over other indices due to a lower saturation effect and closer link to gross primary productivity (Badgley et al., 2017), and the fact that it does not require any coarser resolution bands than those provided by the MOD09Q1 product. Lastly, the data was resampled using bilinear interpolation to align exactly with the 300 m × 300 m grid that the reference samples used. The time series of the pixels at reference data locations were used for all subsequent analysis.

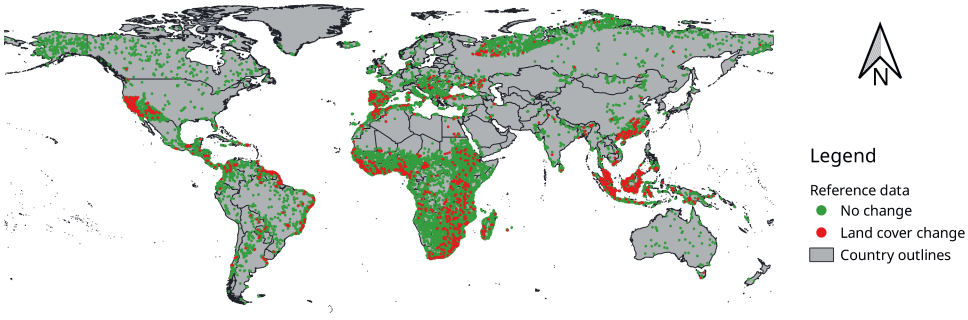


Figure 4.2: Location of all land cover change reference sites. Each site covers a 300 by 300 m area, aligned with the PROBA-V 300 m UTM grid, and indicates whether more than 50% of this area underwent a land cover change between each pair of years between 2015 and 2018. Sites with at least one year of land cover change are marked in red, others in green. Background: World Bank Official Boundaries.

4.2.2.3 Reference data and validation

To measure the accuracy of break detection, we used the land cover change reference data that was collected by the the International Institute for Applied Systems Analysis (IIASA) as part of the CGLS-LC100 project (Buchhorn et al., 2021). The reference dataset is visualised in Figure 4.2. It consists of 33881 300 by 300 m sample sites worldwide, of which there are 2594 sites with land cover change (7.7%). Experts at IIASA collected the land cover change reference data by visual interpretation of very high resolution and Sentinel-2 imagery of the years 2015-2018, assisted by time series of Sentinel-2, MODIS and PROBA-V vegetation indices. The reference data covers a broad range of land cover change, e.g. flooding, deforestation, agricultural expansion, land abandonment, etc.

An important point to consider is that while land cover change and breakpoints in time series are correlated, they are distinct from one another. For instance, forest regrowth is a gradual change process in terms of vegetation index time series and involves no breaks. In contrast, in terms of land cover, forest regrowth involves three changes: bare soil to grass, grass to shrubs, and shrubs to trees. Conversely, a break in time series may not result in land cover change, e.g. a burnt grassland stays a grassland next year, even though its vegetation index experiences an abrupt change between the years. Therefore, we used the p-values of the OLS-MOSUM structural change test, the same test that BFAST employs to filter out time series with no change prior to running, to assign probabilities of change to every time series. We then analysed the accuracy of each algorithm, binning the results based on the probability of there being at least one break in the time series as reported by the OLS-MOSUM test. That way we can be more confident at the higher and lower

ends of the probability gradient that breakpoints (or lack thereof) in these time series correspond to land cover change (or lack thereof).

Calculating the accuracy of detected change is challenging, and defining methods and statistics for this purpose is still an active field of research. This is due to the fact that it is highly unlikely that an algorithm could predict the timing of a break in perfect agreement with the assessment of an expert interpreter. The issue is worse in cloudy areas, where fine spatial resolution observations available to the interpreter differ in frequency from the coarse spatial but high temporal resolution observations input to the algorithm. For instance, an interpreter may only have one image per season, which only allows pinpointing the time of break at a half-year precision. Thus, even though our reference data included an indication of the timing of the change within the year, it was not always highly accurate due to the aforementioned reason, and if the time of change could not be determined, it would default to January 1st. Another issue is data imbalance, as the majority of the reference data indicates no change, since land cover change is a relatively rare phenomenon.

For the purpose of assessing how well time series break predictions match the reference data, we developed a distance-in-time approach similar to that proposed by Zhou and del Valle (2020). We treated each reference and predicted break as an event, and then determined whether it was a true positive, false positive, false negative or true negative. If both the predicted and reference breaks overlap within a tolerance distance, it is considered a true positive; if a predicted break does not overlap with a reference break, it is considered a false positive; and if a reference break does not overlap a predicted break, it is considered a false negative. We chose $\pm$ 1 year as a tolerance threshold, as per previous studies (Cohen et al., 2017). True negatives are more difficult to determine. Even though Zhou and del Valle (2020) recommend treating all the rest as true negatives, that approach does not hold in our case, as there is no predetermined number of change events that can happen during a given time series. The maximum number of events, given a one-year spacing interval, for the three years of reference data, would be seven (four predicted breaks and three reference points of change). However, the majority of the time series do not exceed two, and the most common case is no change with zero events; thus, using a maximum of seven would lead to an inflated true negative count. Therefore, we chose three as the number of events to expect in a time series (i.e. the number to count down from when calculating true negatives), since the reference data covered three years of potential land cover change per reference site. This leads to a total count of positives and negatives similar to the total number of sample sites multiplied by the three years of reference data. If a time series had more than three events, the number of true negatives was set to zero.

While this method leads to potentially unreliable counts of true negatives, it avoids the more serious pitfalls of other approaches, such as event double counting. By themselves,

true negatives are not very important in an imbalanced data setting, where the user is interested primarily in correctly detected breaks. For this reason, the F1 score is a useful statistic, as it ignores true negatives altogether, and so avoids any problems that may arise from unreliable true negative counts. We also separately calculated the overall accuracy, sensitivity (producer's accuracy of change), precision (user's accuracy of change), specificity (producer's accuracy of no change) and an absolute distance between precision and sensitivity as an indicator of bias between the two statistics, which we called "beta".

4.3 Results

4.3.1 Accuracy comparison

We first compared the three algorithms using bins of 1000 points with lowest and highest probability of change based on the p-value of the OLS-MOSUM test, selected from points that were marked as having no land cover change and at least one land cover change event, respectively. In these points we were relatively confident that land cover change and time series breaks should match. The results are presented in Table 4.1.

The results show that the accuracy of all of the tested algorithms is comparable. BFAST Lite was the most conservative algorithm, with the highest overall accuracy, precision and specificity, and the lowest bias between sensitivity and precision. With LWZ breakpoint number selection, BFAST Lite had a higher omission error, while with BIC it had a higher commission error, reversing the balance between sensitivity and precision. In contrast, all the other algorithms had a much higher commission error, as seen by lower precision and specificity scores and a higher beta statistic.

Using stlplus interpolation in BFAST resulted in a lower sensitivity and a higher precision compared to linear interpolation. This is because stlplus interpolation results in smoother time series, so fewer breaks are found. It also improved the beta statistic.

Table 4.1: Accuracy statistics of the tested algorithms. "OA" stands for overall accuracy.

Algorithm	Sensitivity	Specificity	Precision	F1 score	OA	Beta
BFAST Lite (BIC)	0.737	0.881	0.607	0.666	0.852	0.130
BFAST Lite (LWZ)	0.569	0.943	0.715	0.633	0.868	0.146
BFAST (linear int.)	0.880	0.807	0.518	0.653	0.821	0.362
BFAST (stlplus)	0.811	0.832	0.540	0.648	0.827	0.271
BFAST Monitor	0.852	0.851	0.587	0.695	0.851	0.265

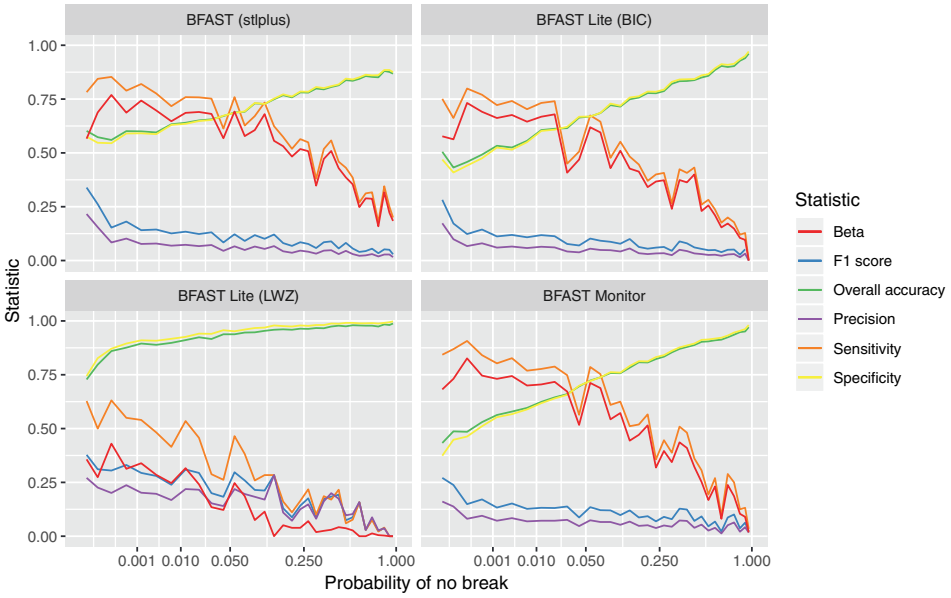


Figure 4.3: Algorithm accuracy comparison across the range of confidence in the time series having at least one break (i.e. ordinary least squares residual moving sum (OLS-MOSUM) test p-value), among the entire reference data set. Towards the left on the x axis are time series with high confidence of a break, towards the right with a high confidence of no break. The data is binned to ∼1000 locations per bin.

BFAST Monitor achieved the highest F1 score and generally outperformed BFAST in each category. Nevertheless, it still had relatively low precision, indicating an overestimation of change.

When looking further into the accuracy statistics and their change across the whole spectrum from high confidence of break in the time series to high confidence of no break in the time series, the results are given in Figure 4.3 for all of the reference data and in Figure 4.4 for reference points marked as land cover change.

The results show that the statistics of all of the algorithms are close to one another, and change consistently across the break confidence range. BFAST Lite with breakpoint number selection by LWZ is the only algorithm that stands out as different from the others. This shows that there is little difference between the algorithm results when comparable parameters (i.e. breakpoint selection by BIC) are used. However, the flexibility of BFAST Lite allows significant customisation of the algorithm when desired. When LWZ was used for selecting the number of breakpoints, it resulted in significantly better results: higher overall accuracy, F1 score, precision and specificity, as well as lower beta, throughout most of the range. However, it also had a lower sensitivity compared with other algorithms. This

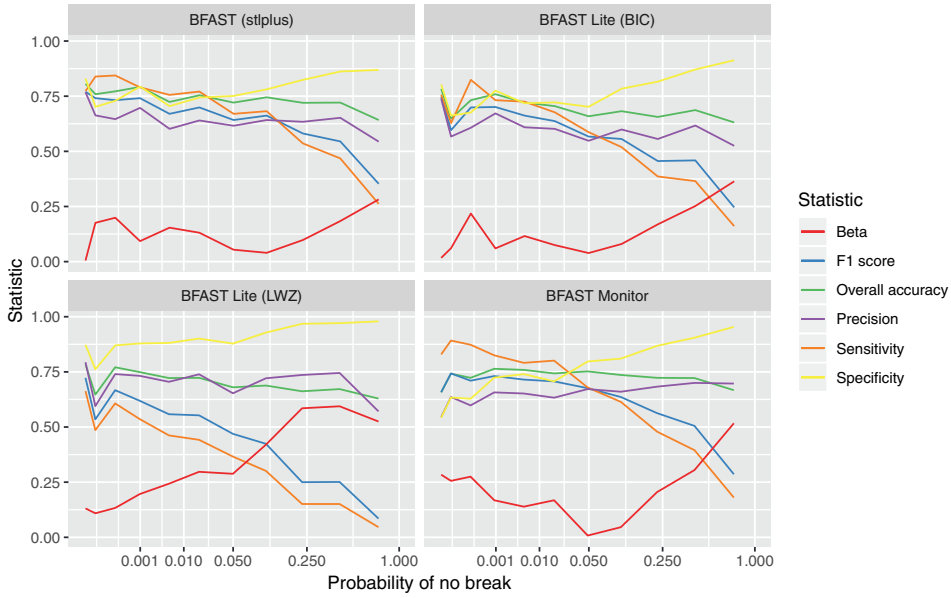


Figure 4.4: Algorithm accuracy comparison across the range of confidence in the time series having at least one break (i.e. OLS-MOSUM test p-value), among time series that have at least one land cover change event. Towards the left on the x axis are time series with high confidence of a break, towards the right with a high confidence of no break. The data is binned to ∼300 locations per bin.

shows that break selection by LWZ makes the algorithm considerably more conservative, which means it does not overestimate the number of changes as often, albeit at the cost of some additional omission error.

If we only consider locations marked as having land cover change in any of the years in the reference data, we see the same pattern: all of the algorithms perform similarly except for BFAST Lite with LWZ. However, in this case BFAST Lite with break number selection by LWZ performed worse than with BIC. Since this dataset has a lot higher incidence of change, the conservative nature of the LWZ criterion results in a decrease in sensitivity that is not adequately compensated by the increase in precision. Since sensitivity is the limiting factor in this case, the use of LWZ also results in higher bias and lower F1 score. In this case, all other algorithms had a higher F1 score and a lower beta statistic. BFAST achieved the highest F1 score and lowest beta throughout the range, but BFAST Lite with BIC had similar results as well.

4.3.2 Run time benchmark

A summary of the run time benchmark results is shown in Figure 4.5. The results show that BFAST Lite run time when using the original R engine is comparable to BFAST when using a single iteration, but up to 2.9 times faster than BFAST using the original iterative approach. However, when using the optimised C++ engine, BFAST Lite is significantly faster than BFAST, regardless of the number of iterations. It was 1.7–2.2 times faster than BFAST using a single iteration and even 4.3–6.0 times faster than BFAST using the original iterative approach.

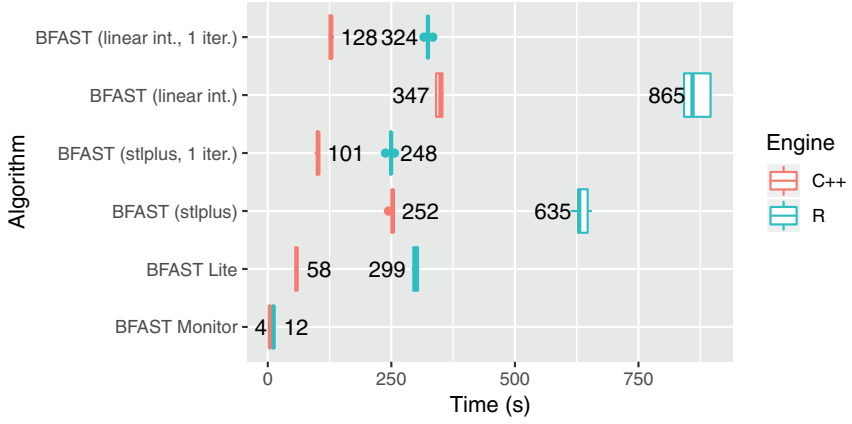


Figure 4.5: Run time benchmark between the tested algorithms. Each algorithm has been run 5 times on a dataset of 288 time series. The time axis represents the total time needed to process this dataset. "1 iter." indicates that the algorithm was run for a single iteration rather than iteratively, "linear int." indicates filling in missing data using linear interpolation.

The C++ engine introduced in the new `bfast` release was faster than the R engine in all cases, with the gains consistently ranging between 2.5 and 3 times the speed of the R implementation.

Comparing across the algorithms, BFAST was the slowest and BFAST Monitor was the fastest. BFAST Monitor was 14.5 times faster than BFAST Lite and even 86.8 times faster than the original BFAST implementation when using the C++ engine. Comparing the two interpolation methods for BFAST, processing time series with stlplus interpolation was consistently faster than using linear interpolation, by 1.3–1.4 times. This is because linear interpolation increases the number of observations that need to be processed by STL. Using a single iteration rather than an iterative approach in BFAST was consistently 2.5–2.7 times faster.

4.4 Discussion and outlook

Our newly proposed algorithm, BFAST Lite, proved to be a useful addition to the existing family of BFAST algorithms. The results showed comparable accuracy with other BFAST algorithms, and in some cases better than the original BFAST, when comparable parameter settings were used. In addition, BFAST Lite was faster and more flexible than BFAST in terms of parameters. The additional parameters introduced in BFAST Lite proved to have a strong effect on the predicted time series breaks, which empowers users to customise the algorithm to fit their study area and use case. To avoid overloading users with too many adjustable parameters, all of the parameters default to reasonable starting values for common tasks, such as forest disturbance detection. See Table 4.A.1 for a list of parameters and their default values. As a future research direction, it is planned to develop a supervised method for optimising parameters and the choice of the input vegetation index based on machine learning. Such a method would build upon the added parametrisation in BFAST Lite and could help users to automate its use for a wide variety of applications and locations.

The newly introduced LWZ criterion of BFAST Lite helps to reduce commission error when the number of samples with no change exceed the samples with land cover change, which is usually the case due to the relative rarity of land cover change. The choice of the information criterion can shift the balance between commission and omission error drastically; therefore, selecting a criterion appropriate for a given task is important. We introduced the LWZ criterion due to its theoretical properties as described in Liu et al. (1997). However, it is not necessarily optimal; information criteria other than AIC, BIC and LWZ may be better suited for specific tasks and result in a better balance between commission and omission error. Therefore, it is planned in upcoming releases of the package to allow users to pass a custom function to act as an information criterion, increasing the flexibility of the BFAST Lite algorithm further.

In addition to the aforementioned benefits, BFAST Lite also brings several more advantages over BFAST. First, it handles time series with missing values without the need for interpolation, unlike BFAST. Second, it provides additional metrics that can be used to further narrow down the detected changes: goodness-of-fit metrics such as R^2, RSS, BIC and LWZ, as well as several indicators for estimating breakpoint magnitude. All of these features are practical benefits that ease the use of the algorithm and the interpretation of its output.

The faster speed of BFAST Lite means that it is now feasible to detect all breaks in time series over large areas, including running the algorithm globally. This was demonstrated in the production of the CGLS-LC100 maps (Buchhorn et al., 2021), where BFAST Lite was used for reducing the number of changes between the yearly maps. BFAST Lite was run on the same set of MODIS imagery as presented in this study, but globally and

Table 4.2: Feature comparison between the three BFAST family algorithms. See Table 4.A.1 for a detailed comparison of parameters between BFAST and BFAST Lite.

Algorithm	Advantages	Disadvantages
BFAST (Verbesselt et al., 2010a)	Detects breaks in trends and seasonality separately	Slow, limited number of parameters, overestimates change
BFAST Monitor (Verbesselt et al., 2012)	Designed to detect breaks at the end of time series (near real-time), fastest, many tunable parameters	By default not designed for multiple breakpoints, overestimates change
BFAST Lite (this study)	Faster than BFAST, designed for multiple breakpoints, many tunable parameters, lowest bias between sensitivity and precision, highest OA	Needs parameter tuning to optimise performance, does not differentiate between breaks in seasonality and trend

wall-to-wall. It was feasible to process the global time series data and produce a map of detected changes in approximately a week, running the algorithm on the Terrascope cluster, which provides approximately 1000 processor cores and 1.5 terrabytes of memory. To build upon these large-scale processing capabilities, BFAST Lite can further be ported to Google Earth Engine, similar to BFAST Monitor (Hamunyela et al., 2020). This would allow running the algorithm over higher spatial resolution time series, such as Landsat, and would further ease the algorithm accessibility for a wider user audience.

Our study also showed good performance of BFAST Monitor compared to BFAST, even though BFAST Monitor was not designed for detecting multiple breaks. It is faster than BFAST Lite and the accuracy is also comparable with BFAST. However, the limitation of BFAST Monitor is a high commission error. It is less suitable for dealing with data with relatively little change, as is the case at the global scale. See Table 4.2 for a summary comparison of the advantages and drawbacks of the three algorithms that we compared in this study.

Despite the recent advances in time series break detection, there is still a research gap remaining between the concepts of unsupervised break detection and land cover change detection. Since these concepts are not interchangeable, break detection algorithms alone are not enough to identify all types of land cover change. Therefore, more research is needed for improving land cover change detection. One approach is to use supervised algorithms. Such algorithms can make use of the output from unsupervised break detection algorithms, such as the BFAST family, as one of the inputs to help determine whether a land cover change has occurred. This approach was implemented by Xu et al. (2022), who found that combining the output of two BFAST Lite and one BFAST model in a Random Forest model increased the F1 score compared to BFAST Lite alone. This shows that BFAST Lite, due to increased flexibility and additional statistics on model fit and

breakpoint magnitude, can help to further narrow down the detected breaks to those that actually correspond to land cover change.

4.5 Conclusion

We have introduced a new time series break detection algorithm, BFAST Lite, which is a faster and more flexible variant of the BFAST algorithm. The algorithm implementation in R is openly available in the `bfast` package version 1.6 on CRAN. BFAST Lite can handle missing data (irregular time series) and has a variety of user-customisable parameters, including the newly introduced LWZ information criterion for automatic determination of the number and timing of breakpoints. Our results showed that BFAST Lite performs similar to BFAST when using comparable parameters, with a higher user's accuracy but a lower producer's accuracy of change, and with a better balance between the two statistics. The choice of parameters has a big influence on the results, which allows tuning the model, e.g. decreasing commission errors when there is a lower incidence of change. In addition, BFAST Lite provides more statistics about the breakpoints and goodness-of-fit compared to other BFAST methods, which allows postprocessing of the results.

These outcomes, including the increased processing speed, show that BFAST Lite is a good candidate for upscaling break detection to larger areas. It has already been adopted in the production of the CGLS-LC100 product for annual land cover map updating at a global scale. BFAST Lite can help to detect disturbances on the ground and estimate their magnitude, as well as to determine long-term change trends, over larger areas than could previously be achieved with the original BFAST algorithm.

4.A Appendix

Table 4.A.1: Comparison of parameters between BFAST and BFAST Lite.

Parameter	BFAST	BFAST Lite
Minimum segment size h	15% by default	15% by default
Trend component	Always included	Included by default, can be disabled
Seasonality component	Either harmonics or seasonal dummies	Harmonics, seasonal dummies, both, or external regressor
Maximum harmonic order	Preset to 3	Customisable, defaults to 3
Number of seasonal dummies	Preset to equal to observations per year	Customisable, defaults to the number of observations per year, but can be fewer
Maximum number of iterations	10 by default	1
Structural change test type	OLS-MOSUM by default	None, from version 1.7: optional with none by default
Structural change test significance threshold	0.05 by default	None, from version 1.7: optional with 0.05 by default
Decomposition algorithm	STL by default or stlplus	None by default, STL or stlplus on any of the components
Break number selection criterion	Preset to BIC	LWZ by default, BIC, RSS
Autoregressive components	None (not supported)	None by default, seasonal lag, trend lag, or both

Chapter 5

Postprocessing high frequency global land cover maps for change assessment

This chapter is based on:

D. Masiliūnas, N.-E. Tsendbazar, M. Herold, D. Marcos, and J. Verbesselt (2023). "Postprocessing high frequency global land cover maps for change assessment". Under revision.

Abstract

The field of land cover mapping has recently been shifting towards more rapidly updated land cover maps, with the Dynamic World product from Google classifying every Sentinel-2 scene as soon as it is published. There is also a move towards mapping land cover probabilities and fractions, in addition to discrete classes. This new volume of data is very valuable, but it is challenging for land cover products to accurately represent land cover change, as the time series of land cover maps exhibits spurious change and seasonal effects. In this study, we propose to use high frequency land cover maps as input into change detection algorithms, such as BFAST Lite, in order to improve land cover change estimation by reducing spurious variability in the time series, while retaining abrupt change events. We applied the proposed method to both our own Random Forest regressor predicting land cover fractions, and the per-class probabilities of the Google Dynamic World product. We compared the results with simpler postprocessing algorithms, as well as with a common approach of change detection based on a spectral index. The results show that our method reduces the variability of dense land cover time series, but also provides additional semantic information on land cover change facets that so far are rarely available, such as land cover change metrics (trends, magnitudes) and which fraction of each land cover class changes into which other for every time step. However, postprocessing only adjusts the output of a base classifier, therefore to improve the map accuracy, further improvements in base land cover classification algorithms are needed. We recommend producers of high frequency land cover map products to include postprocessing in their processing chain, making use of a processing level system similar to that of satellite imagery processing chains.

Keywords. BFAST, postprocessing, land cover, time series, change detection

5.1 Introduction

Land cover is being mapped at an increasingly high spatial and temporal detail, as evidenced by recent developments such as the Dynamic World land cover product developed by Google (Brown et al., 2022), which provides a discrete land cover map with nine classes for every Sentinel-2 image globally. For the first time it is possible to track land cover change in near-real-time for all of the land cover classes. Such high frequency maps are also called dense time series of land cover, because of the high density of points when an area of a map is plotted as a time series.

Land cover maps are improving not only in frequency, but also in their ability to incorporate continuous information, rather than only presenting discrete classes (Masiliūnas et al., 2021a). The Dynamic World product (Brown et al., 2022) includes probabilities of each land cover class in every pixel, which can be visualised as continuous layers of each land cover class. Typically, the model confidence decreases around the borders of a land cover patch, where multiple classes are mixed in a single pixel of the satellite sensor. To visualise this effect, Google used hill shading, where valleys indicate low probability of a class, and peaks indicate areas of high model confidence in the mapped class. Other products, such as CGLS-LC100 (Buchhorn et al., 2020), include explicitly mapped land cover fractions. Land cover fraction models predict the fraction of each class in a pixel, therefore also resulting in continuous layers of land cover classes. In border areas, fraction layers explicitly capture land cover mixtures. Therefore, fractions and probabilities are highly related, as they both provide continuous layers in percentage units that have high values in areas covered by a single class and low values in mixed areas, and all classes sum up to 100%. In fact, class proportions predicted from Random forest have been successfully used as an estimator of the area covered by each class (Sales et al., 2022). However, fractions and probabilities are distinct concepts and result from different processes. Namely, probabilities are an intrinsic product of model confidence and do not involve any training data, whereas fractions are obtained by explicitly training a model on labels containing land cover fraction percentages. The advantage of maps that provide continuous land cover data, especially in a dense time series, is the possibility of tracking not only abrupt change, but also gradual change, such as forest regrowth (Asner et al., 2002; Broadbent et al., 2006) and degradation (Wang et al., 2005).

The major remaining challenge in such temporally dense time series of land cover maps is their consistency: between subsequent time steps, observed changes are often spurious, especially as the interval between the time steps decreases (Abercrombie and Friedl, 2016). This noise in the time series of land cover, primarily caused by small variations between images and randomness in the classifiers, results in an overestimation of change, and makes real change on the ground more difficult to identify. In turn, the lack of consistency causes confusion for the users as to how to make use of the maps, especially for scientific

modelling communities in which consistent land cover data is essential (Bontemps et al., 2012).

The temporal consistency issue can be tackled in two ways. The first is to improve the algorithm that makes initial land cover predictions, which is called the base classifier, or base regressor when the output is continuous (Masiliūnas et al., 2021a). The second is to postprocess the output of the base classifier using a second model that attempts to remove the noise from the base classifier output. Postprocessing methods range from simple spatial and temporal filters (Souza et al., 2020) and expert rules to smoothing and change detection algorithms. Smoothing algorithms, or smoothers, result in a time series with reduced variability based on neighbouring points in time. Examples of smoothers used in land cover postprocessing include Savitzky-Golay filters (Seo et al., 2016), LOESS, Markov Chain, Hidden Markov Model (Abercrombie and Friedl, 2016) and Markov Random Fields (Wang et al., 2017; Xie et al., 2022). Some of the commonly used unsupervised change detection methods used in land cover postprocessing are BFAST family algorithms (Verbesselt et al., 2010a; Verbesselt et al., 2012), e.g. in Buchhorn et al. (2020), and LandTrendr (Kennedy et al., 2010), e.g. in Shimizu and Saito (2021). Supervised change detection methods learn from training data of labelled change events, and due to the training data requirement are expensive to apply (Abercrombie and Friedl, 2016). Postprocessing methods can also be combined. For example, Buchhorn et al. (2020) combined Hidden Markov Model and BFAST Lite with a vegetation index input, whereas Xie et al. (2022) combined Markov Random Fields and CCDC.

The change detection algorithm used in postprocessing typically takes a univariate time series as an input, often of the NDVI. In a simple application of change detection, if the algorithm does not detect change in the time series, the pixel is considered to be stable, and the previous prediction of the base classifier is retained. If a change is detected, the prediction after the end of the detected change is used as the new value in the updated map. This approach reduces the amount of observed change between time steps, but it relies on the accuracy of the change detection algorithm, which in turn relies on its single-index input data. Even though NDVI is sensitive to changes in vegetation, no single index is capable of capturing the complexity of the whole diversity of land cover change. One option to overcome this challenge is to perform change detection on multiple indices in parallel (e.g. Xu et al. (2022)) or to use a multivariate change detection algorithm (e.g. Zhu and Woodcock (2014)), but in either case there is a domain adaptation issue of a spectral index change not necessarily matching the change in land cover.

In this study, we propose the use of a time series change detection algorithm as a postprocessing method, taking dense time series of land cover fractions or probabilities as input. We explored the effects of directly using predicted land cover fractions or probabilities, rather than spectral indices, as an input into the change detection algorithm BFAST Lite. We then compared the results with that of a change detection approach

using NDVI as input to BFAST Lite, as well as with a simple smoothing technique, namely LOESS. We assessed the map, change, trend and variability errors of two base classifiers, namely the Google Dynamic World product and our own Random Forest regressor, as well as the aforementioned postprocessing techniques applied to the base land cover maps.

5.2 Data and methods

5.2.1 Reference data sets

In this study, we made use of the global CGLS-LC100 land cover fraction reference data, collected by high resolution image interpretation by experts at sample sites of 100 m × 100 m (Buchhorn et al., 2020; Tsendbazar et al., 2021b). To train Random Forest regressors for predicting land cover fractions, we used the 2015 data by IIASA, and for validating the errors of all products and model outputs, we used the 2015-2019 yearly data collected by WUR.

To harmonise the legends across the datasets and products, we followed the procedure of Masiliūnas et al. (2021a), merging moss and lichen, flooded vegetation, herbaceous grassland and burnt areas into the herbaceous class; shifting cultivation into the cropland class; and snow and ice into the bare land class. The resulting legend consists of 7 classes: herbaceous, cropland, trees, shrubs, bare land, built-up and water. Sample sites that were dominated by the class "not sure" or whose fractions were all reported to be zero were dropped.

After this filter, the 2015 training data consisted of 138 308 sample sites. The 2015-2019 validation data consisted of 30 712 sample sites; with five years' land cover recorded per site, it amounted to 153 560 data points. See Figure 5.1 for an overview of the sample locations.

5.2.2 Landsat 8 imagery

For training the Random Forest model, we used the whole Landsat 8 OLI Collection 2 surface reflectance archive (Masek et al., 2006; U.S. Geological Survey, 2023) from 2013-03-18 until 2021-07-29. The data was masked of clouds using the mask provided by the CFMask algorithm, by keeping only pixels marked as no fill, no cloud, no dilated cloud, and no cloud shadow. We extracted the 16-day time series of every band of OLI, aggregated exactly to the 100 m × 100 m footprints of the sampling sites using a weighted mean reducer. The data extraction was done in Google Earth Engine (Gorelick et al., 2017).

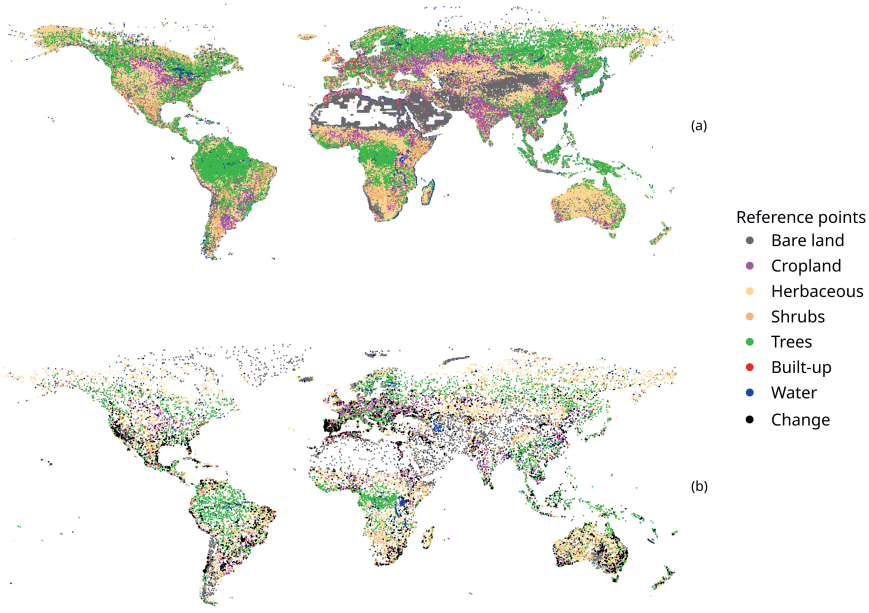


Figure 5.1: The global distribution of the reference data: the 2015 training data (a) and the 2015-2019 validation data (b). The colours indicate the dominant land cover fraction of the given reference point, if the fractions were stable throughout the time series, otherwise the point is marked as Change.

5.2.3 Base classifiers

5.2.3.1 Random Forest regression for dense land cover fraction mapping

As a base classifier predicting dense time series of land cover fractions, we implemented a binary relevance Random Forest regression model following Masiliūnas et al. (2021a) (`ranger` package in R, Wright and Ziegler (2017)). The regression model was trained on the fraction labels provided by the IIASA training data of 2015 (see Figure 5.1 (a)), assuming that the fractions represent the land cover at the peak of vegetation season, i.e. middle of July in the northern hemisphere and middle of January in the southern hemisphere. The features that were matched with the labels included Landsat 8 OLI surface reflectance and vegetation indices closest to the peak of season. The trained model was then used to predict the whole dense (16-day) time series of land cover fractions using observed reflectance values and vegetation indices at each time step.

We included multiple vegetation indices and a snow index to help the model discern between land cover classes, and the harmonic and temporal features to take into account the variability of the land cover over time. All these values were calculated by making use of a moving window through each of the 16-day observations of Landsat 8 OLI. We chose the ±1.5-year interval for temporal features as a recommended choice for robust estimation of trends (Tyukavina et al., 2018) as a compromise between the robustness of the derived statistics and the need to preserve short-term changes in the time series. The harmonic statistics in particular needed longer time series to be robustly estimated, therefore we included an extra feature with a longer temporal window for one of the vegetation indices, namely Tasselled Cap (greenness) (TCg). For the list of all input features used, see Table 5.1.

Table 5.1: Input features for the Random Forest regression model. The first column indicates the number of input features that come from each feature category.

#	Name	Formula	Time span
1	Surface reflectance band 4 (red)	ρ_{655}	Instantaneous
1	Surface reflectance band 5 (near infrared)	ρ_{865}	Instantaneous
1	Surface reflectance band 6 (shortwave infrared 1)	ρ_{1610}	Instantaneous
1	Normalised Difference Vegetation Index (NDVI) (Kriegler et al., 1969)	$\frac{\rho_{865}-\rho_{655}}{\rho_{865}+\rho_{655}}$	Instantaneous
1	Enhanced Vegetation Index (EVI) (Huete et al., 1999)	$2.5\frac{\rho_{865}-\rho_{655}}{\rho_{865}+6\rho_{655}-7.5\rho_{480}+1}$	Instantaneous
1	Normalised Burn Ratio (NBR) (Key et al., 2002)	$\frac{\rho_{865}-\rho_{2200}}{\rho_{865}+\rho_{2200}}$	Instantaneous
1	Normalised Difference Snow Index (NDSI) (Hall et al., 1995)	$\frac{\rho_{560}-\rho_{1610}}{\rho_{560}+\rho_{1610}}$	Instantaneous
1	Tasselled Cap (greenness) (TCg) (Kauth and Thomas, 1976)	$-0.2941\rho_{480} - 0.243\rho_{560} - 0.5424\rho_{655} + 0.7276\rho_{865} + 0.0713\rho_{1610} - 0.1608\rho_{2200}$	Instantaneous
4	NDVI, EVI, NBR and TCg medians	Q_2	±1.5 years
4	NDVI, EVI, NBR and TCg interquartile ranges	$Q_3 - Q_1$	±1.5 years
4	NDVI, EVI, NBR and TCg 5th percentiles	P_5	±1.5 years
4	NDVI, EVI, NBR and TCg 95th percentiles	P_{95}	±1.5 years
5	NDVI, NBR, NDSI and TCg amplitudes of order 1 (Jakubauskas et al., 2001)	$c_1 = \sqrt{a_1^2 + b_1^2}$	±1.5 years and ±2.5 years for TCg
5	NDVI, NBR, NDSI and TCg phases of order 1 (Jakubauskas et al., 2001)	$\phi_1 = \arctan\frac{b_1}{a_1}$	±1.5 years and ±2.5 years for TCg

The mean voting rule was used and the output fractions were normalised to add up to 100% at each time step. The predicted fractions were validated using the validation data at the sampling locations, as in Figure 5.1 (b).

5.2.3.2 Google Dynamic World product

In addition to testing postprocessing on our own base classifier, we also tested how well the postprocessing algorithms can generalise to different base classifiers, and especially whether they are applicable to both land cover fractions as well as land cover probabilities. Therefore, we made use of the Google Dynamic World V1 product (Brown et al., 2022), which is based on Sentinel-2 imagery and provides probabilities of each land cover class. Specifically, we extracted the probability of each land cover class over each reference site, aggregating it to their 100 m footprints using a weighted mean reducer. The Dynamic World product is based on Sentinel-2 MSI data with a joint revisit time of approximately 5 days, therefore we made use of the whole archive of Dynamic World from 2015-06-01 to 2023-01-01 at 5-day time steps.

Equivalent to the harmonisation as per section 5.2.1, we merged the probabilities of flooded vegetation into grass and merged snow and ice into bare. Note that while we attempted to harmonise the legends of Dynamic World and CGLS-LC100, there may still be small differences in class definitions. In addition, Dynamic World provides probabilities, rather than fractions, which are generally more fuzzy (rarely 0% or 100%) compared to land cover fractions. Since the validation data followed the definitions of CGLS-LC100 fractions, extra error is inevitably introduced merely due to these differences.

5.2.4 Postprocessing methods

A quick overview of the tested postprocessing models and their parameters can be found in Table 5.2. A visual example of some of the time series and the result of postprocessing using the tested models can be seen in Figure 5.3.

All of the processing in this study has been done using R and Google Earth Engine. The code is freely available online (Masiliūnas, 2023).

5.2.4.1 Postprocessing using BFAST Lite

BFAST Lite is a univariate algorithm that detects breaks in the trend and seasonality of an input variable (Masiliūnas et al., 2021b; Verbesselt et al., 2010a). It is based on a piecewise linear regression, where the input time series is modelled using selected temporal features such as trend, harmonics, or seasonal dummy variables. BFAST Lite performs break detection by segmenting the input time series into stable segments in terms of regression coefficients, i.e. each segment has its own set of coefficients. A break is then the boundary between adjacent segments, and the method allows deriving metrics such as magnitude of

Table 5.2: Tested models and their parameters. See Figure 5.3 for a visual example of time series of these models.

Model name	Model description	Model parameters
Dynamic World	Google Dynamic World V1 class probabilities (base classifier).	None
Dynamic World + LOESS	Dynamic World postprocessed using LOESS	$\alpha = 0.75$
Dynamic World + BFAST Lite	Dynamic World postprocessed using BFAST Lite by directly using class probability time series.	$h = 0.16$, magnitude (RMSD) $> 30\%$, formula = trend
Random Forest regression	Land cover fractions predicted by Random Forest regression using binary relevance (base classifier).	Defaults of the `ranger` package
Random Forest + LOESS	Random Forest fraction predictions postprocessed using LOESS.	$\alpha = 0.75$
Random Forest + BFAST Lite	Random Forest regression postprocessed using BFAST Lite by directly using fraction predictions.	$h = 0.16$, magnitude (RMSD) $> 30\%$, formula = trend
RF + NDVI-only BFAST Lite	Random Forest regression fractions, only updated if BFAST Lite detects a break in Landsat NDVI.	order = 3, magnitude (RMSD) $> 20\%$, formula = trend + harmon
Random Forest + BEAST	Random Forest regression postprocessed using BEAST, combining the trend and seasonal components.	Defaults of the `Rbeast` package

change and model fit over each segment, as well as overall. This way all breakpoints in the time series are detected, with the optimal number of breaks determined by the chosen information criterion, by default LWZ (Liu et al., 1997). BFAST Lite can handle missing data, which is an important consideration in areas with frequent cloud coverage. It is also significantly faster than its predecessor BFAST, which is an important consideration when upscaling land cover change detection to the globe and using denser time series. However, it cannot detect breaks at the start and end of the time series within the minimal segment size h. In this work we used the R version of the BFAST Lite algorithm, freely available in the `bfast` package in R (Masiliūnas et al., 2021c).

A vegetation index such as NDVI is commonly used as input into such change detection algorithms, and the output breakpoints are then used as a proxy indication of potential land cover change. The time series of predicted land cover (fractions) is then postprocessed, so that if no change is detected, the previous class or fraction is retained. If a change is detected, the class or fraction is updated (see Figure 5.2 (a)). The updated value

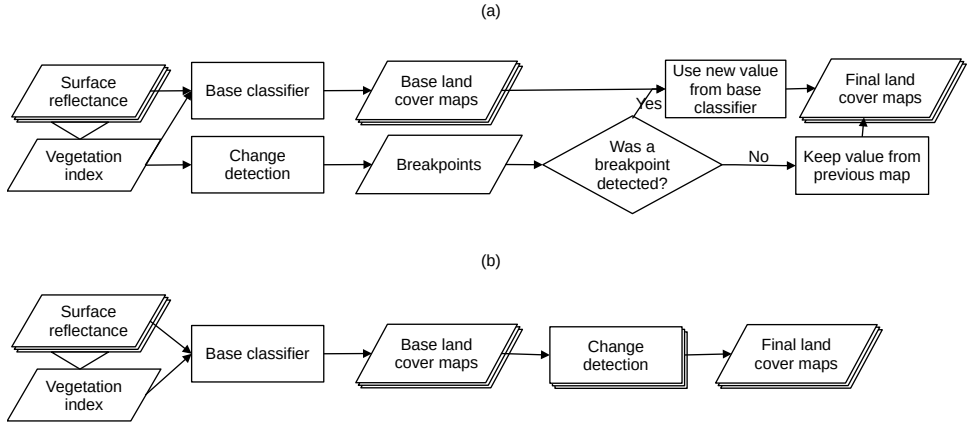


Figure 5.2: A comparison of two processing chains for a time series of a pixel of a satellite sensor: the traditional land cover map processing chain using breakpoints from a change detection algorithm run on a vegetation index (a), and our proposed land cover map processing chain using a change detection algorithm to directly postprocess land cover (b). The chain is repeated for each pixel of the map. In this study, the surface reflectance was taken from Landsat 8 OLI, base classifiers were Random Forest and Google Dynamic World, and the change detection algorithm was BFAST Lite.

may or may not be different from the previous value, as some breaks in an NDVI time series do not correspond to a land cover (fraction) change. We tested this approach using NDVI as input, and refer to it as "NDVI-only BFAST Lite" in the results section and in Table 5.2.

BFAST Lite can use any univariate time series as input, and the fitted piecewise linear regression is an idealised model of the change in the time series. Therefore, we proposed and tested the use of land cover fractions or probabilities directly as input into the change detection algorithm, in this case BFAST Lite. This makes it possible to use BFAST Lite as a postprocessing algorithm directly, as its fitted model retains trends, breakpoints, and (optionally) seasonality of the input land cover class, but discards noise and outliers in the time series, resulting in a smoothed output. The magnitude of detected breaks can be further used to filter out breaks that are too small in terms of fraction change percentage to be considered abrupt land cover change, therefore filtering out some of the remaining spurious change.

We applied both the traditional NDVI-only approach and the proposed direct approach of using BFAST Lite as a postprocessing algorithm of the random forest regression fraction predictions, in order to compare the two approaches. We optimised the model parameters (see Table 5.2), and the traditional NDVI-only BFAST Lite method performed best using trend and harmonics of order 3 as components, and when filtering out breaks with a

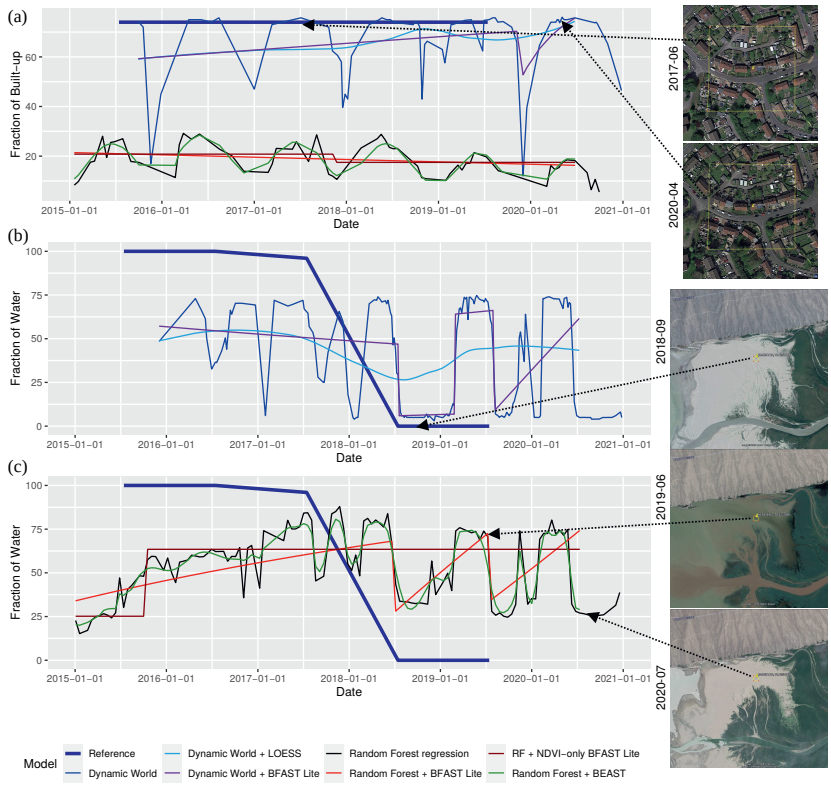


Figure 5.3: Example time series. a) the built-up class fraction in a suburban area of Kingswood, Bristol, UK (longitude: -2.4911165, latitude: 51.464915). There was no change in built-up area (74%, reference line), but there is considerable seasonality due to urban green space, resulting in fluctuations predicted by both base classifiers. Postprocesing smooths them: BEAST retains seasonality, LOESS predicts a smoothed curve, BFAST Lite predicts sloped lines and retains breaks (such as in 2019-12 due to consecutive unfiltered clouds in Dynamic World), NDVI-only BFAST Lite predicts straight lines with jumps when a break is detected. b & c) the water fraction of the Ili river draining into the Kapchagay reservoir in Kazakhstan (longitude: 78.1237458, latitude: 43.8591412). Its water level fluctuates seasonally, but between 2015 and 2018 the water level has been relatively stable. Abrupt drought events happened in 2014-08, 2018-08 and 2019-08 (Baranowski et al., 2020). The reference data records a change from water to seasonally flooded bare, which is recorded as bare. NDVI is not helpful to detect change between water and bare soil. LOESS oversmooths changes, BEAST does not smooth changes enough. BFAST Lite retains jumps and gradual changes when they are predicted by the base classifier.

magnitude (RMSD) lower than 0.2 NDVI. BFAST Lite directly applied on land cover fractions/probabilities performed best using a trend-only model, filtering out breaks of land cover less than 30% in magnitude (RMSD). Both models used 1 year as the shortest segment size h.

5.2.4.2 Postprocessing using BEAST

To showcase how the approach of using land cover fractions as input directly could be applied to change detection algorithms other than BFAST Lite, we also tested the Bayesian estimator of abrupt change, seasonality & trend (BEAST) algorithm (Zhao et al., 2019). The BEAST algorithm is an ensemble break detection model that combines several runs of time series decomposition models and produces an averaged model fit. The model automatically detects patterns in the time series and adapts the submodel parameters to better model the time series. The input into BEAST was the same as for BFAST Lite, i.e. the land cover fractions predicted by Random Forest. We used the default model parameters (see Table 5.2). To obtain postprocessed land cover fractions, we summed the estimated trend and seasonality components, or if seasonality was not detected, used the estimated trend component by itself.

5.2.4.3 Postprocessing using LOESS

We compared the postprocessing results of the change detection methods to a relatively simple locally estimated scatterplot smoothing (LOESS) technique. LOESS (Cleveland, 1979) is a variant of the Savitzky-Golay filter (Savitzky and Golay, 1964) that fits quadratic polynomials in a local window using weighted least squares, where the weight decreases further away from the centre of the window (tricubic weighting). Unlike change detection methods, LOESS does not attempt to detect breaks, but simply smooths the input time series, resulting in a more consistent output, therefore reducing the interannual variability observed in the time series of land cover fractions or probabilities. Due to the simplicity of the method, it is also faster to compute. We used the default parameters of LOESS in R (Cleveland et al., 1992), namely, smoothing parameter $\alpha = 0.75$ (see Table 5.2).

5.2.5 Model validation

We assessed the errors of the models by calculating map error, change error, trend error and variability error (see Table 5.3 for the definitions; also see Appendix 5.A and Figure 5.A.1 for an example calculation of the statistics as well as a graphical representation). These error statistics were calculated at the sampling locations of the 2015-2019 validation data (Figure 5.1).

Table 5.3: Error statistics derived from the time series of a single sample site.

Name	Formulae	Interpretation
Map error	$\text{RMSE}_t = \sqrt{\frac{\sum_{t=1}^{T}(y_{pt}-y_{rt})^2}{T}}$ $\text{MAE}_t = \frac{\sum_{t=1}^{T}\lvert y_{pt}-y_{rt}\rvert}{T}$ $\text{ME}_t = \frac{\sum_{t=1}^{T} y_{pt}-y_{rt}}{T}$	Error of the individual map produced for time t.
Change error	$\text{RMSE}_c = \sqrt{\frac{\sum_{t=1}^{T}(\Delta y_{pt}-\Delta y_{rt})^2}{T}}$ $\text{MAE}_c = \frac{\sum_{t=1}^{T}\lvert \Delta y_{pt}-\Delta y_{rt}\rvert}{T}$ $\text{ME}_c = \frac{\sum_{t=1}^{T} \Delta y_{pt}-\Delta y_{rt}}{T}$ where $\Delta y_t = y_{t+1} - y_t$	Error of change between time t and $t+1$.
Trend error	$e_a = a_p - a_r$	Error in the 5-year trend a, as estimated by an OLS regression.
Variability error	$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^{T} r_t^2}{T}}$ $e_v = \text{RMSD}_p - \text{RMSD}_r$	Error in the variability of the time series, estimated by the square root of the RSS.

To obtain statistics for the whole class and overall for all classes, the RMSE, MAE and ME statistics were calculated by pooling each error statistic e from each time series. Examples of RMSE:

$$\text{RMSE}_{\text{class}} = \sqrt{\frac{\sum_{l=1}^{L} e_l^2}{L}} \tag{5.1}$$

$$\text{RMSE}_{\text{overall}} = \sqrt{\frac{\sum_{c=1}^{C}\sum_{l=1}^{L} e_{cl}^2}{C \times L}} \tag{5.2}$$

where l is the location number, L is the total number of sample locations, c is the class number and C is the total number of classes.

In addition, we calculated a relative version of the class error statistics by dividing them by the average fraction size of each class. The relative statistics adjust for the fact that estimating rare classes is inherently easier, as their fractions are dominated by zeroes for most locations, thus a constant zero prediction leads to lower error. However, the absolute values of relative error statistics are more difficult to interpret compared to absolute error statistics. Therefore, the relative statistics are reported in Appendix 5.B.

To determine whether there was a significant difference between the tested models, we performed a one-way repeated measures ANOVA on the absolute residuals of both sets of models (Dynamic World and Random Forest). The analysis was implemented using a mixed effects framework (Bates et al., 2015) where the model was the fixed effect and the

sample site, class and year were nested random effects. The difference between individual models was analysed using Tukey's all-pairwise comparison (Hothorn et al., 2008). The differences between class predictions were analysed in the same way, with the class as the fixed effect and sample site and year as the random effects.

To analyse the effect of postprocessing on land cover transitions, we calculated transition matrices, where rows correspond to classes in the starting year and the columns to the ending year. The values represent the sum of all fractions that have experienced the corresponding transition. Overestimation and underestimation can be assessed by comparing the transitions of the reference dataset with those of the model predictions.

5.3 Results

5.3.1 Map error

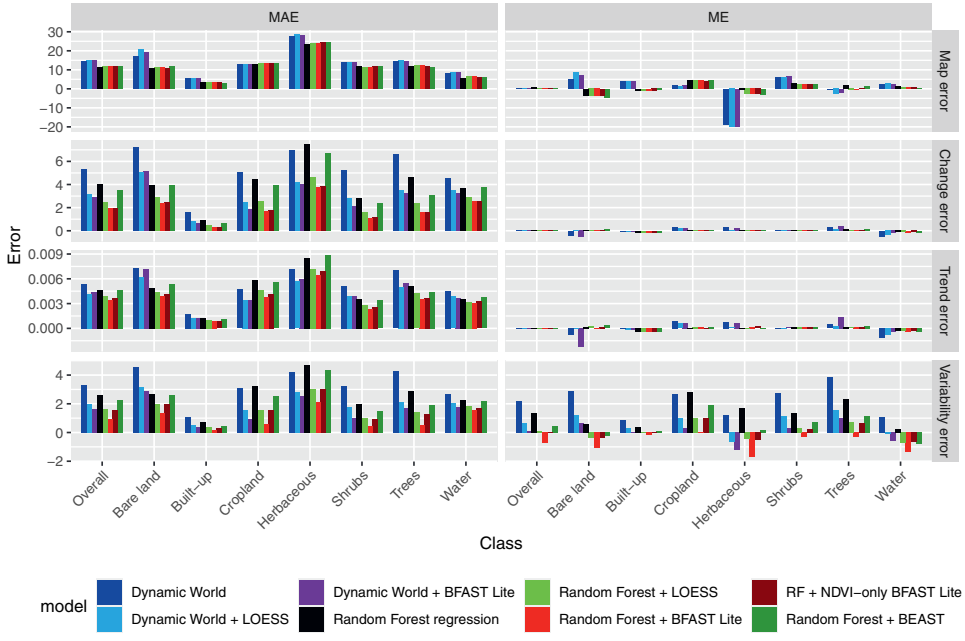


Figure 5.4: Validation statistics of base classifiers and postprocessing models, with locations and time steps pooled together. For the relative statistics, see Figure 5.B.1.

Results show that when postprocessing is applied, map error overall slightly, though significantly increases for both base classifiers ($F = 4489.9$, $p < 0.001$ for Dynamic World models and $F = 8582.2$, $p < 0.001$ for Random Forest models, see Figure 5.4).

Postprocessing mostly hindered the accuracy of the predicted bare land class. The effect was similar on both the Random Forest regressor, and the Dynamic World product.

The difference between the base classifiers was significant ($F = 44358$, $p < 0.001$) and larger than the differences between the postprocessing methods. Random Forest had a significant advantage over Dynamic World in every class, even in cropland ($z = -9.8$, $p < 0.001$), which was the most similar between the two base classifiers. Dynamic World also showed a strong underestimation of the herbaceous class, and a weaker overestimation effect of all the other classes, except for trees. These effects most likely come from the fact that Dynamic World uses probabilities rather than fractions, which are generally more fuzzy than actual fractions in the reference data, and the herbaceous class has particularly low confidence due to low base classifier accuracy for this class (Brown et al., 2022).

5.3.2 Change error

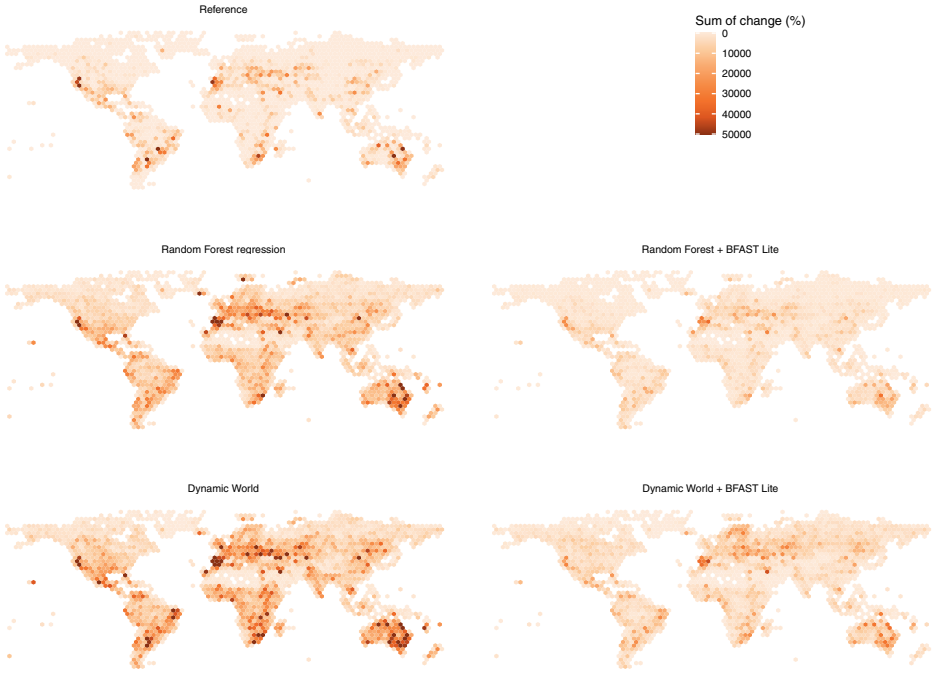


Figure 5.5: Comparison between select model predictions and the reference dataset on land cover fraction change, aggregated to the sum per hexagon (diameter: 3.6 degrees). Both base classifiers overestimate change. After postprocessing, change overestimation is reduced, although the results are less selective and more diffuse than the reference data.

In terms of change error, there was a highly significant difference between the results with and without postprocessing, as well as between the postprocessing algorithms ($F = 41368$ for Random Forest models and $F = 52296$ for Dynamic World, $p < 0.001$). Applying postprocessing in some cases halved the change error. The BFAST Lite algorithms performed the best of all tested postprocessing algorithms, both when applied on Dynamic World probabilities and on Random Forest fractions. Even though LOESS performed significantly worse than BFAST Lite ($z = 99.1$, $p < 0.005$ for BFAST Lite on Random Forest fractions), it still reduced change error considerably, given the simplicity of the algorithm. There was an almost insignificant difference in performance between BFAST Lite applied on NDVI, and BFAST Lite applied on land cover fractions ($z = 3.0, p = 0.022$).

Similar to the map error, the results showed that Dynamic World had an overall significantly higher error compared to the Random Forest regressor in terms of land cover change ($z = -210.6$, $p < 0.001$), except for the herbaceous class, where the error was lower in Dynamic World than in Random Forest. However, postprocessed Dynamic World performed better than non-postprocessed Random Forest, which shows the strong impact that postprocessing has on change error. Figure 5.5 shows the spatial distribution of land cover change, both reference and modelled, showing the same pattern in space.

5.3.3 Trend errors

Postprocessing significantly decreased the trend errors as well ($F = 4273.9$ for Random Forest and $F = 4905$ for Dynamic World groups, $p < 0.001$). In terms of algorithms, the results show a diverging effect of different algorithms on the two base classifiers. LOESS performed significantly better than BFAST Lite when applied on Dynamic World ($z = -15.7$, $p < 0.005$), but worse when applied on the Random Forest regressor ($z = 57.5$, $p < 0.005$).

In addition, BFAST Lite performed significantly better when applied on land cover fractions than when applied on NDVI ($z = -17.0$, $p < 0.001$). The reason is that when BFAST Lite is applied to NDVI, and no change is detected, the previous value is kept, resulting in a time series with a "stairstep" effect. In contrast, when applied to fractions, the BFAST Lite model produces smooth trends, which matches the trends in the reference data better (see Figure 5.3 (c)).

The difference in trend error between the base classifiers was smaller compared to the other errors. While the Random Forest regressor still overall had a significantly lower trend error ($z = -47.1$, $p < 0.001$), but Dynamic World performed significantly better at predicting trends for both cropland ($z = 12.4$, $p < 0.001$) and herbaceous ($z = 20.8$, $p < 0.001$) classes.

5.3.4 Variability error

Postprocessing had a very significant effect in reducing variability errors ($F = 20106$ for Random Forest and $F = 29093$ for Dynamic World models, $p < 0.001$). Without postprocessing, both of the base classifiers show an overestimation of variability in the predicted land cover time series. Postprocessing strongly reduced the variability, and therefore the variability error. BFAST Lite on fractions and probabilities had a significantly stronger effect of reducing variability compared to LOESS ($z = 114.1$ for Random Forest and $z = 55.6$ for Dynamic World models, $p < 0.001$), as it used a trend-only model, which discards seasonality, whereas LOESS minimised seasonality but did not entirely discard it, therefore still somewhat overestimating the time series variability. BFAST Lite applied on land cover fractions had an even stronger effect on reducing variability compared to BFAST Lite applied on NDVI ($z = -99.2$, $p < 0.001$). The effect was so strong that the variability of the predicted time series was underestimated compared to the reference values, but overall the variability error was also reduced much more compared to all of the other postprocessing techniques.

The differences between the base classifiers were similar as with the previous error types, with Random Forest overall performing significantly better ($z = -85.4$, $p < 0.001$), but Dynamic World performed significantly better at estimating the variability of the herbaceous class ($z = 18.0$, $p < 0.001$).

5.3.5 Land cover transitions

The reference dataset shows clear differences between the land cover classes in how often they transition to other land cover classes. Some classes are more dynamic than others. In particular, herbaceous is a dynamic class, as a relatively large proportion of land cover transitions (4.7% of the reference land surface) is between herbaceous and all of the other land cover classes (see Figure 5.6). The transitions between water and herbaceous cover are particularly common. Bare land is also a dynamic class, with particularly frequent changes (0.9% of all change) between water and bare land. The transition from trees to herbaceous cover is common, but the opposite is relatively rare, as it takes multiple years for trees to regrow. The built-up class is particularly stable and rather rare, with few transitions to it and even fewer transitions away from it. Nevertheless, even though these transitions are rare, they still sometimes occur. The only transitions that were never observed were from built-up to natural vegetation (trees and shrubs), and from water to trees.

The transition matrices of model predictions (Figure 5.6) show a clear overestimation of predicted change for base classifiers, with a lot more predicted transitions between all classes than observed in the reference dataset. Postprocessing greatly reduces these spurious transitions. However, the effect is general and not class-specific, therefore in cases where change was underestimated, such as transitions from herbaceous to water,

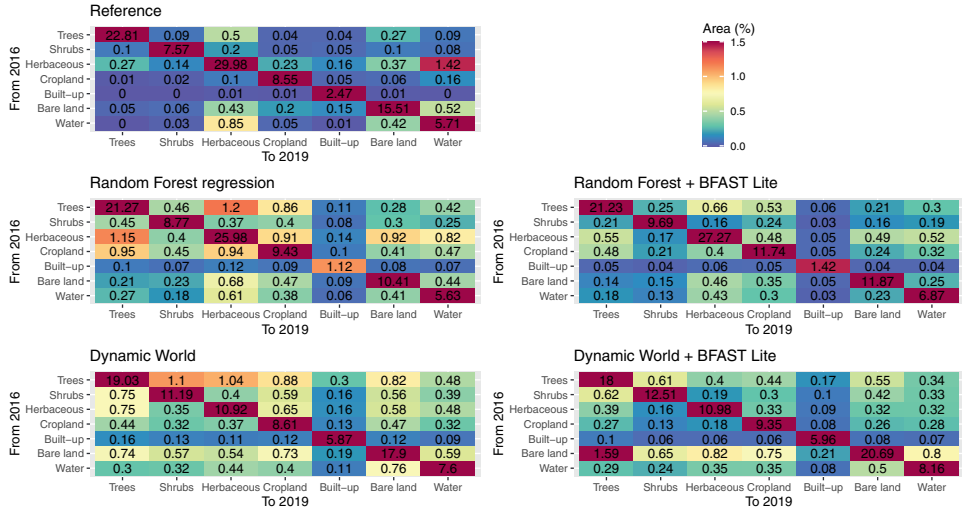


Figure 5.6: Change matrices from all land cover classes in 2016 to all land cover classes in 2019, in proportion of area covered by the reference data.

postprocessing causes further underestimation of change. And vice versa, postprocessing greatly decreases the underestimation of no change (e.g. stable bare land), but also increases the overestimation of no change (e.g. stable cropland).

5.4 Discussion

Mapping land cover fractions or probabilities for every satellite observation is a new and evolving field in remote sensing. The results of our study show the importance of using postprocessing to produce more consistent land cover maps throughout the time series. Even a simple postprocessing algorithm like LOESS significantly decreased the overestimation of change and resulted in more stable time series, when applied to either land cover fractions or class probabilities. This consistency across time is an important feature for many applications of land cover maps, including climate and biodiversity modelling, forest monitoring and land management.

BFAST Lite, either with the common approach of using NDVI as an input, or the novel approach of using dense land cover fractions or probabilities as an input, performed even better than LOESS (see Figure 5.4). The BFAST family of algorithms, including BFAST Lite, has a well-established ability to detect, and thus retain, large breaks in the input time series (Masiliūnas et al., 2021b; Verbesselt et al., 2010a; Verbesselt et al., 2012). In addition, it is able to remove noise from time series. The combination of these two

features is an important advantage over simpler models, although at the cost of increased computational complexity.

The novel approach of using dense land cover fractions or probabilities as input to BFAST Lite led to improvement in the estimation of land cover trends and variability in the predicted time series, compared to using NDVI as input. The main reason is that using land cover fractions or probabilities to model change, rather than discrete breakpoints in the time series, avoids the "stairstep" effect inherent in the NDVI-only approach (see Figure 5.3).

But in addition to the quantitative improvements, using dense land cover fractions as input to BFAST Lite also brings many qualitative improvements and possibilities. The base classifier, using machine learning, handles the selection of spectral indices, temporal and auxiliary features, and therefore can take a lot more information into account than just a single spectral index. Compared to the common approach of using a vegetation index time series for change detection, our method also removes the burden of selecting a suitable spectral index from the user. Furthermore, the output of BFAST Lite, when applied to dense land cover fractions, has richer semantic meaning than a spectral index. First, a change of 1% fraction of a land cover class is the same no matter whether the site contains a large fraction of the class or not. It avoids the common saturation effect problem of spectral indices. Second, if we have predictions for all land cover class fractions, then, together with an indication of a time series breakpoint, we also get an indication of which class changed into which other class, and at what magnitude. The user can then filter out locations where change of particular magnitude happened, e.g. filter out changes smaller than 30% per change event or per year, and can focus on areas with a particular land cover transition type, e.g. floods in natural areas. In contrast, a break in a spectral index is only an indication of a potential change in the spectrum, which is not necessarily nor directly linked to any particular land cover change and is not expressed in natural units. Third, the modelled time series enables the monitoring of both gradual and abrupt change in each of the land cover classes. BFAST Lite allows quantifying trends of each stable segment, which gives an opportunity to track long-term trends that would otherwise not be evident, in every land cover class, such as forest degradation or the increase in urban density. Lastly, this approach allows estimating the stability of each land cover class, by counting the number of breakpoints per class over a given period of time, and by calculating the time since the last breakpoint. In between breakpoints, stability can be estimated by analysing trends. Therefore, this approach allows quantifying land change metrics, which is one of the land cover change facets that is least studied, yet has the potential to revolutionise the land cover mapping field by providing new semantic classes, such as "greening urban" (Zhu et al., 2022).

In this study we used two base classifiers and several postprocessing algorithms, but there are many possible combinations. We performed a test with the BEAST change

detection algorithm applied on the Random Forest predictions, but the resulting time series was only slightly smoothed (see Figure 5.3), therefore still overestimating change and producing higher errors, even more than using LOESS (see Figure 5.4). This test shows that the effect of choosing an algorithm and its parameters on the output time series can be considerable, but also that the framework itself is easily adaptable to multiple change detection algorithms.

The choice of the algorithm depends not only on the accuracy of the resulting land cover maps, but also on the intended use case. BFAST Lite worked well for producing consistent time series, but it is not designed to detect changes close to the end of the time series. Therefore, to make the use of high frequency land cover maps as means to detect changes in near-real-time, different postprocessing algorithms would need to be used. For example, in this case BFAST Monitor (Verbesselt et al., 2012) could be a suitable postprocessing algorithm, applied in the same way that BFAST Lite was used in this study, i.e. using land cover fractions or probabilities as input. The output would be a smoother time series compared to the raw base classifier output, with land cover changes at the end of the time series indicated each time BFAST Monitor detects a break. The additional benefits of this combination would be largely similar to the ones outlined above about BFAST Lite. The approach could also be adapted from a dense land cover time series to work with yearly time series, by making use of a yearly change detection algorithm, such as LandTrendr (Kennedy et al., 2010). Likewise, this approach can be used with any base classifier and any land cover or land use classes, as long as they are expressed as continuous values, i.e. fractions or probabilities.

In this study, we analysed unsupervised remote sensing time series postprocesing methods. The advantage of unsupervised methods is the ease of application and automation, as they do not require any additional training data and are usable without the need of defining expert rules. In the field of land cover map postprocessing, most research has been done on supervised methods instead, most often Hidden Markov Models (Abercrombie and Friedl, 2016; Buchhorn et al., 2020; Liu et al., 2021a; Sulla-Menashe et al., 2019). These models can take the transition likelihood into account and learn it from the input data, and in most cases they successfully increase the map accuracy (Buchhorn et al., 2020; Liu et al., 2021a). However, building Markov models requires expert knowledge on the likelihood of land cover transitions, or extra training data for reliably estimating them. This is particularly problematic for dense time series, as it is difficult to obtain dense reference data due to the infrequent revisit time of fine spatial resolution sensor platforms. Another common method of postprocessing land cover maps is to use simple spatial and temporal filters based on predefined (expert) rules (Buchhorn et al., 2020; Malinowski et al., 2020; Souza et al., 2020). However, this method has a similar issue due to the difficulty of correctly defining the rules, as they may differ per region, time and land cover class, making it difficult to generalise these rules to larger areas.

5.5 Limitations and recommendations

The main limitation of postprocesing methods for land cover time series is the need to rely on the base classifier. Postprocesing methods can only adjust the output of the base classifier, but it will never be radically different (e.g. in Figure 5.3 (a) we can clearly distinguish two clusters that correspond to the two base classifiers), therefore errors such as misclassification cannot be easily corrected through postprocessing. Conversely, if the base classifier had perfect accuracy, there would be no need for any postprocessing. Therefore, while postprocessing helps to improve the consistency of the time series, the focus should be on improving the base classifier accuracy. One of the ways to improve the base classifier performance is to include auxiliary data as features (Masiliūnas et al., 2021a). Another way would be to make use of and combine existing land cover maps. For example, Dynamic World predictions could be used as features in the Random Forest regression model. A third option would be to make use of a deep learning base classifier that can take both space and time into account. A combination of these approaches is also possible.

One limitation of the reference data in this study is that it was collected yearly. Every reference site was only assessed once per year, typically around the peak vegetation season. Therefore, it was necessary to convert dense land cover time series into yearly time series for the purposes of statistical validation. Having a reference dataset with dense labels would allow a more in-depth analysis of the different postprocessing methods, such as how well they can predict the seasonal variability of land cover, or to determine the time lag between an abrupt event and its detection.

Our results showed that while postprocessing improves the consistency of the time series and improves change accuracy, the map accuracy slightly decreases (see Figure 5.4). Therefore, if map accuracy is the only important criterion, it is better to not perform any postprocessing, or to choose a supervised postprocessing method.

When applying an unsupervised postprocessing model, the results benefit from tuning the model parameters. It is recommended to tune the parameters for each dataset, as the model parameters used in this study may be specific to the global scope of the input data and may not be the best choice for local studies and when using different land cover class legends.

Making use of land cover fraction data as input into a postprocessing model is a useful way of assessing the land cover change target, namely, to automatically identify which class shifted to which other class and when. However, it remains difficult to assess the land cover change agent, that is, the underlying cause of the land cover change. For example, using the methods in this study we can detect areas that have changed (by a given fraction percentage) from trees to herbaceous cover, however, we cannot determine whether the change happened due to illegal deforestation, a natural forest fire, or regular forest management practices. To determine the change agent, additional information

from the ground is required. A potential way of obtaining it at a large scale is by using crowdsourced data, e.g. by automatic sourcing information from news reports and social media posts that relate to a detected land cover change event at a given location and time, or by using citizen science.

Our results showed that the postprocessing methods are able to reduce the amount of spurious transitions (see Figure 5.6), however, they do so in a way that is not class-specific. That means that if the base classifier was underestimating the change in a particular class, the postprocessing will make the underestimation worse. It would be possible to combine change detection postprocessing with a technique to adjust transitions based on their likelihood, using a transition matrix. A potential topic to investigate in follow-up studies is whether such a combination of postprocessing techniques, such as BFAST Lite and Hidden Markov Model (HMM), could lead to further accuracy improvements compared to the individual methods alone, and if so, in what order they should be applied.

In this study, we tested only temporal postprocessing methods. A method that combines both temporal and spatial information, such as ones suggested by Boucher et al. (2006) and Liu et al. (2021a), may result in higher accuracy, at the cost of greatly increased input data volume.

Given the benefits of having consistent dense land cover (fraction) maps, we recommend implementing postprocessing approaches in practise, as the tools and technology to do so are already available. Google is producing the Dynamic World product automatically on the Google Earth Engine platform, and the processing chain could be further extended to include derived postprocessed products. One way of releasing such postprocessed maps would be to apply a processing chain structure similar to ones used for processing satellite imagery, with different processing levels. For instance, a base classifier output without postprocessing would be level 0, output that is smoothed using BFAST Lite and BFAST Monitor could be level 1 (for historical and near-real-time land cover, respectively). Output derived from the postprocessing models, such as land cover transition targets and metrics (e.g. a map showing areas of decreasing and increasing trends in each land cover class), could be released as level 2 products. Products that are further derived from the targets and metrics, such as land cover change agents, could be level 3. An update to the level 0 maps, such as a newly classified image, could then trigger a rebuild of the other levels as well. Given the recent progress of implementing change detection methods in Google Earth Engine (e.g. Kennedy et al. (2018)), including BFAST family methods (e.g. Hamunyela et al. (2020)), it is already possible to create a workflow to produce level 1 and even level 2 maps on the platform. Doing so would provide a boost to the usability of the Dynamic World product. Even without official support from Google, it would be possible to implement postprocessing as a set of third-party processing scripts, similar to the approach in Reiche et al. (2021).

5.6 Conclusion

The recent developments in the field of land cover classification have enabled the creation of products providing dense time series of land cover. In turn, these dense time series provide an opportunity of more detailed assessment of land cover change, and also necessitate new methods for handling them.

In this study, we proposed a framework to use these time series of land cover as input into a change detection algorithm, such as BFAST Lite. We confirmed that postprocessing the time series obtained from a base classifier, even with simple methods, leads to an improvement in consistency. Among the tested algorithms, BFAST Lite, applied on dense land cover fractions, performed the best in terms of reducing the MAE of change, long-term trends and time series variability. It proved adaptable to handling both land cover fractions and probabilities, on both manually trained model output as well as existing operational land cover products. Furthermore, the output of BFAST Lite enables richer semantic information about land cover, such as tracking long-term land cover trends, filtering abrupt change events by a change threshold, and defining new land cover change classes such as "greening urban" or "degrading forest". The framework is adaptable to other change detection algorithms as well, such as ones specialised in near-real-time mapping of land cover change. The limitation of our proposed approach is that the map accuracy depends on the accuracy of the base classifier, and that postprocessing slightly decreases the static map accuracy in favour of increasing the change accuracy.

Our proposed framework is a step towards a processing chain for dense time series of land cover maps, that would provide users with more choice in land cover (change) products, such as different levels of processing appropriate for the needs of the users. At higher processing levels, the framework provides the opportunity to quantify and map the facets of land cover change metrics and targets, for (and among) all of the land cover classes. The technology already exists to implement such a processing chain in practice, which would even result in an improvement of existing operational land cover maps. Building upon this framework, in the future it will be possible to also identify land cover change actors, which would make land cover maps even more actionable.

Appendices

5.A Explanation of validation statistics with examples

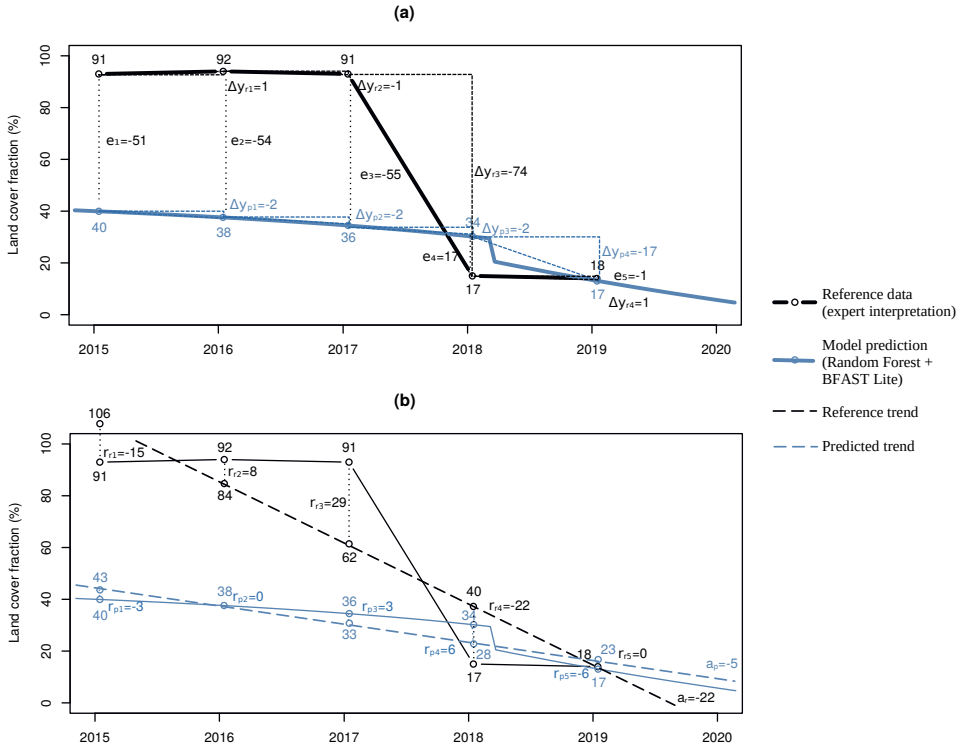


Figure 5.A.1: Example time series of a land cover fraction (grassland fraction in a *Eucalyptus* plantation in Uruguay). a) Calculation of map error from the differences (e_t) between predicted and reference land cover fractions, and change error from the differences between subsequent years between the predicted (Δy_{pt}) and reference (Δy_{rt}) land cover fractions. b) Calculation of trend errors from the reference (a_r) and predicted (a_p) trends, and variability errors from the differences between the regression lines and the predicted (r_{pt}) and reference (r_{rt}) fractions.

In the time series example in Figure 5.A.1(a), the map error statistics were calculated as $\text{RMSE}_t = \sqrt{(-51^2 + -54^2 + -55^2 + 17^2 + -1^2)/5} \approx 42.03$, $\text{MAE}_t = (|-51| + |-54| + |-55| + |17| + |-1|)/5 = 35.6$, and $\text{ME}_t = (-51 - 54 - 55 + 17 - 1|)/5 = -28.8$, indicating that the prediction is on average 28.8% lower compared to the reference values.

The RMSE of change for the example in Figure 5.A.1(a) would be calculated as $\text{RMSE}_c = \sqrt{((-2-1)^2 + (-2+1)^2 + (-2+74)^2 + (-17-1)^2)/4} \approx 37.14$.

The trend error in the example in Figure 5.A.1 (b) is simply the difference between the trends: $e_a = -5 - (-22) = 17$. When all the sample sites are pooled for estimating the error at the class level, MAE of trend is an indication of average absolute trend error and ME of trend is an indication of whether the predictions are overestimating or underestimating trends of the class.

The variability of the example predicted time series in Figure 5.A.1 (b) can be calculated as: $\text{RMSD}_p = \sqrt{((-3)^2 + 0 + 3^2 + 6^2 + (-6)^2)/5} \approx 4.24$. The example predicted time series is more stable (lower RMSD) than the reference suggests it should be: $\text{RMSD}_r = \sqrt{((-15)^2 + 8^2 + 29^2 + (-22)^2 + 0)/5} \approx 17.97$.

5.B Relative errors

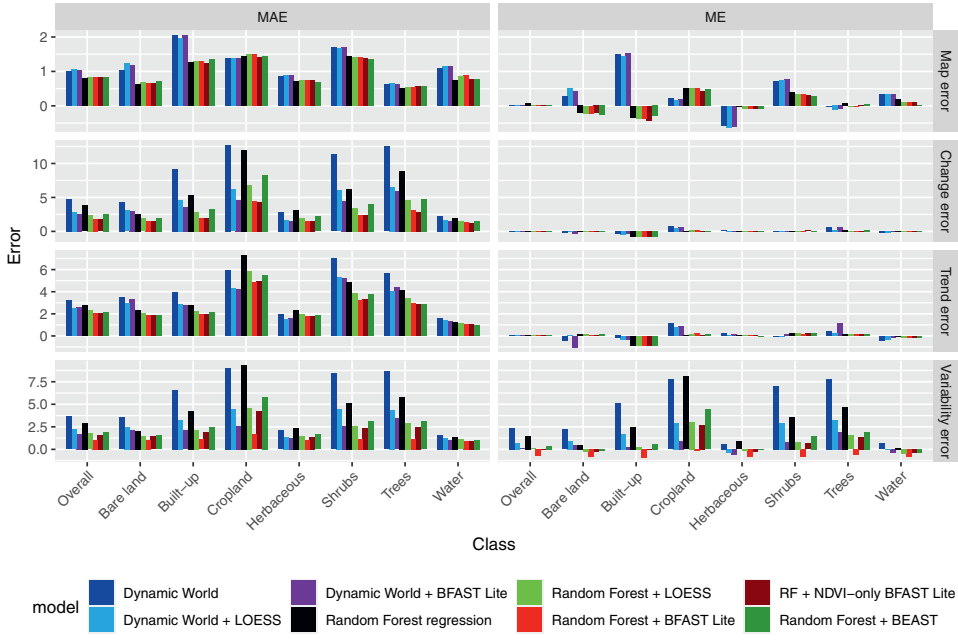


Figure 5.B.1: Relative validation statistics of the base classifiers and the postprocessing models, adjusted for the land cover class sample sizes. For absolute statistics, see Figure 5.4.

Chapter 6

Synthesis

6.1 Main findings

The aim of this thesis was to advance the methods for global land cover fraction and change mapping, by improving both the accuracy as well as the consistency of change between global land cover maps. In this thesis, I investigated machine learning algorithms and tackled the zero-inflation problem to improve global land cover fraction mapping (Chapter 2), predicted yearly global land cover fractions and applied a Markov chain model to postprocess them to improve their temporal consistency (Chapter 3), developed the BFAST Lite algorithm to improve the speed and flexibility of the BFAST change detection algorithm (Chapter 4) and used it as a postprocessing algorithm to improve temporal consistency of temporally dense time series of global land cover predictions (Chapter 5). In all of the chapters, I used four datasets, created as part of the Copernicus Global Land Service Land Cover 100 m (CGLS-LC100) project, to train machine learning and postprocessing models and validate their output. Table 6.1 lists which of these datasets were used in which study and for which purpose.

Table 6.1: Datasets used for model training and validation in each study (see the end of section 6.1.1 for an overview of Xu et al. 2022). The 2015 training dataset (IIASA, Buchhorn et al., 2020) consists of over 150 000 sample sites; the 2015 validation dataset (WUR, Tsendbazar et al., 2019) consists of over 21 000 sample sites; the multitemporal (2015-2018) dataset (IIASA, Buchhorn et al., 2021) consists of over 30 000 sample sites, thus over 120 000 samples; and the multitemporal (2015-2019) validation dataset (WUR, Tsendbazar et al., 2021a) consists of over 30 000 sample sites, thus over 150 000 samples. Data is available on request from IIASA and WUR.

Study	IIASA 2015	IIASA 2015-2018	WUR 2015	WUR 2015-2019
Chapter 2	Training	-	Validation	-
Chapter 3	Training	Training	-	Validation
Chapter 4	-	Validation	-	-
Chapter 5	Training	-	-	Validation
Xu et al. (2022)	-	Cross-validation	-	Cross-validation

In the following subsections, the key answers to each of the research questions raised in Chapter 1 are given.

6.1.1 How well can global land cover fraction maps be produced given various regression methods and their input data?

This research question was answered in Chapter 2 and supplemented by my co-authored research done in Xu et al. (2022). Machine learning methods based on classification and regression trees (CARTs), such as random forest (RF), are widely used in the field of remote sensing for various tasks, including producing global land cover maps. I verified that these methods have advantages over other models also for producing global land

cover fraction maps, using their regression modes. The most important input into the regression models is remote sensing data, especially time series or features derived from satellite image time series. Adding a variety of data sources, such as auxiliary soil and climate data improves the model output, but most important data sources are those that are not redundant with others and when their observations correspond to the time of the predicted land cover.

In Chapter 2, I conducted a comparison between linear regression models, an SVM model, an MLP model and a number of CART models for global land cover fraction mapping. Results showed that CART models performed the best, with RF regression achieving the lowest root mean squared error (RMSE) and Cubist regression achieving the lowest mean absolute error (MAE). I further investigated if special handling of the data imbalance, namely the zero inflation issue inherent in land cover fraction data, would help improve the output land cover fraction maps. I proposed two novel multi-step models that combine classification and regression models to better handle the zero inflation issue. I also compared this approach with using a median vote for the RF ensemble, as well as a combination of both approaches. Results showed that both the multi-step approach and median voting reduce MAE by up to 1.5 percentage points by producing a more crisp output, i.e. with more pure 0% and 100% predictions, rather than tending towards predicting a mean value and resulting in more mixed-class pixels. However, these approaches also resulted in an increase in RMSE by up to 3.4 percentage points, due to a higher chance of complete misclassification of pure classes. The best was a combination of a three-step model and median voting, as median voting decreased MAE by 1.5 percentage points and the three-step model resulted in more accurate pure class predictions, resulting in a lower increase in RMSE (2.9 percentage points) compared to the other models and model combinations.

In Chapter 2 I found that features derived from satellite image time series were by far the most important for global land cover fraction mapping, as a model trained only on satellite imagery achieved 3 percentage points lower overall accuracy (OA) compared to a model that also included auxiliary terrain, climate and soil data (Figure 2.C.1). Nevertheless, these three sources of auxiliary data were useful enough that none of the features, except for the terrain aspect, could be removed from the model without the accuracy decreasing. Out of the auxiliary data, climate data derived from interpolated aggregated historical weather station data was particularly important, and it was the only source of data aside from satellite imagery that included a temporal component.

I further explored the effect of optical satellite image time series on RF predictions. This work has been published in Xu et al. (2022), which aimed to analyse the effects of optical imagery on land cover change detection. In contrast with Chapter 2, results in this study showed that including features from multiple satellite sensors in an RF model simultaneously only improves the prediction of global land cover change in areas with

frequent cloud coverage, and otherwise the added features may even hinder the model performance. Furthermore, Landsat 8 OLI data alone often led to better results compared to Sentinel-2, PROBA-V and their combination for the 100 m sample sites. The results confirm findings in other studies such as Ludwig et al. (2022) that it is important to consider omitting correlated features from RF land cover models.

6.1.2 How can the flexibility and scalability of change detection algorithms be improved for land cover change monitoring at the global scale?

This research question was addressed in Chapter 4 and supplemented by Chapter 5 and the research done in Xu et al. (2022). I successfully explored three main directions of improving the BFAST unsupervised change detection algorithm: optimising the speed, improving the parametrisation and enriching output statistics, and implemented these improvements in the R package `bfast`. I also explored combining the output of the unsupervised BFAST family algorithms, including my newly developed algorithm, with supervised machine learning techniques.

In Chapter 4, I introduced BFAST Lite, a variant of the original BFAST algorithm (Verbesselt et al., 2010a) that is up to six times faster and can natively handle missing data in the input time series. In addition, BFAST Lite has additional user-configurable parameters that have a large effect on the output of the algorithm, allowing users to decrease the commission error of the algorithm. Lastly, BFAST Lite includes new statistics about its fitted model: a goodness-of-fit indication (RSS and R^2) and several new more robust statistics of breakpoint magnitude. This extra output information can be used by the users to determine which breakpoints are of interest and which can be filtered out.

The improvements I introduced in Chapter 4 were not limited to the newly proposed change detection algorithm. The newly introduced C++ engine improved the speed of the entirety of the BFAST family of change detection algorithms. The improvements in speed are very important for applying all of these algorithms at scale, especially global scale. In addition, the original BFAST algorithm was supplemented with an approach that allows the algorithm to deal with a moderate amount of missing data in the time series without users having to interpolate the input time series.

In Chapter 5, I explored whether changing the input of BFAST Lite would result in improved output. The study showed that using a dense time series of land cover fractions instead of a vegetation index allows BFAST Lite to output more meaningful statistics. The breakpoint magnitudes of detected change are expressed in the units of change in fraction cover (percentages), and the model output allows tracking both abrupt and gradual change, as well as quantifying the trends of change for every individual land cover class. This added semantic meaning opens new possibilities for the monitoring, analysis and interpretation of global land cover change.

I further explored improving land cover change detection by combining unsupervised and supervised methods using a hybrid BFAST and RF approach in Xu et al. (2022). This study combined the output (including the new magnitude statistics) of various unsupervised BFAST algorithms and with varying parametrisation, using an RF model. The results showed that RF can act as a "smart filter", reducing the commission error from each individual unsupervised model. Combined with a balanced sampling strategy, the hybrid approach improved the F_1 score from 38.1 (best individual unsupervised model score; achieved using BFAST Lite) to 61.5 (best hybrid model score). Therefore, a hybrid approach is useful to decrease the commission errors of unsupervised change detection algorithms and leads to an improvement in the accuracy of global land cover change detection.

6.1.3 What postprocessing methods can be used to improve the consistency between frequently updated land cover fraction maps?

This research question was addressed in Chapters 3 and 5. I analysed two different methods of postprocessing global land cover fraction maps: using a Markov chain model on yearly land cover fraction predictions and using BFAST Lite on dense fraction predictions. Both of these methods managed to successfully improve the consistency of the time series of global land cover predictions and reduce the noise in the time series. A Markov chain can be applied to either yearly or temporally dense time series of land cover fractions, whereas BFAST Lite requires a dense time series. However, BFAST Lite not only smooths the output of a base classifier, but also provides additional valuable information, such as the timing of land cover change events, magnitude of change in fraction values, and quantified long-term trends. On the other hand, Markov models are partially supervised and extensible, therefore they can be customised for the problem at hand and can take expert knowledge into account.

In Chapter 3, I analysed three methods of postprocessing yearly time series of land cover fractions: a Markov chain, a recurrent RF and a temporal RF. Results showed that the Markov chain resulted in improvements in both the accuracy of yearly predictions (reducing RMSE from 23.7 to 22.4 in 2017, Table 3.4) and the accuracy of change between the years (reducing RMSE from 15.0 to 8.8 in 2015-2016, Table 3.5). In contrast, the RF models were in some cases behind even the base classifier model with no postprocessing: both models reduced the yearly map accuracy (RMSE increased from 23.7 to 25.2 with temporal RF and even 26.4 with recurrent RF in 2017) and recurrent RF reduced change accuracy (from 15.0 to 17.0 in 2015-2016). These results show that Markov models even with a relatively simple set of rules are able to reduce the variation in time series, whereas RF models do not perform well when directly predicting land cover fractions, especially when the training dataset has a different sampling scheme than the validation dataset.

In Chapter 5, I proposed the use of BFAST Lite as a postprocessing algorithm to improve the temporal consistency of dense land cover time series. Even though change detection algorithms are not originally designed as postprocessing algorithms, the use of land cover fractions as input time series instead of a VI has a number of advantages. BFAST Lite fits a piecewise linear regression model on the univariate input time series, which represents its idealised fit, without the noise caused by atmospheric or seasonal effects. This modelled time series retains both the gradual change and the abrupt change from the input time series, therefore not losing vital information, but improving temporal consistency. The results showed that BFAST Lite was adaptable to time series of both land cover fractions and land cover probabilities, and it reduced the change error from 4.0 to 1.9 MAE for my own RF land cover fraction predictions and from 5.3 to 2.9 MAE for the probabilities of the Google Dynamic World product. The BFAST Lite model also significantly reduced errors in trend and variability in the time series, more than any other tested method. However, unlike the Markov chain model, it did not improve the map accuracy. The results of this chapter show that BFAST Lite can successfully be used to improve the temporal consistency of global land cover fraction maps, while also adding additional semantic information based on the time series and the detected change in land cover.

6.2 Reflection and outlook

6.2.1 Outcomes of the research

The outcomes of this thesis are focused on fundamental research into technical improvements to global land cover mapping methods, which serve to increase the accuracy and robustness of global land cover maps and their time series. While the proposed improvements are incremental steps for enhancing global land cover products, they are already starting to have a measurable impact on the land cover community and the wider society.

The three-step method for handling the zero inflation issue of land cover fraction mapping (Chapter 2) has been implemented into the processing chain of CGLS-LC100 (Buchhorn et al., 2021). The use of the method I developed in a major global land cover project showcases the method's applicability. In turn, the resulting global land cover maps have been widely adopted by various user communities.

The BFAST Lite method introduced in Chapter 4 is free and open-source software, now published in the primary software repository of the R programming language, CRAN. Therefore, this method is easily accessible to all researchers in the world who are familiar with the R programming language. Work is being done to further expand the audience of this method. I am working together with the Department of Computer Science of the University of Copenhagen to produce a highly-optimised Python version of BFAST Lite (`https://github.com/mortvest/bfast-py`), which will further improve the speed of the

algorithm by leveraging graphics processing units (GPUs). In addition, the core breakpoint detection algorithm that BFAST Lite is based on is currently being implemented into Google Earth Engine. These further developments will make it even easier for users to make use of this new method at a global scale.

Users are also starting to adopt BFAST Lite. Notably, BFAST Lite was used by the CGLS-LC100 project to reduce spurious variation between yearly maps and to reduce the time needed for updating the yearly maps (Buchhorn et al., 2021). Ngadi Scarpetta et al. (2023) used BFAST Lite as a baseline to compare with their own variant of BFAST Monitor. Hong et al. (2022) used BFAST Lite in a novel application of detecting tipping points in (semi-)arid ecosystems. I am further contributing to the outreach by organising conference workshops, summer schools and educational activities demonstrating the application of BFAST Lite in practice.

Time series of land cover maps in general have a very high potential for impact in policymaking due to its power to uncover links between policy decisions and their impact on the ground. My co-author work in Chu et al. (2021) has shown that by using a time series of land cover maps, we can link particular events, such as the construction of a bridge, to its effects on infrastructure development and financial impact on various economic sectors. In addition, by having a long-term time series of land cover maps with all major classes represented, we can track the impact of long-term decisions, such as the land cover change happening over decades due to the implementation of a reforestation project (Chu et al., 2021). We can also track shifts in land cover, such as urban expansion resulting in loss of agricultural land, in turn getting compensated by the reclamation of land from the sea. Having accurate global land cover time series, especially of land cover fractions and/or at a fine spatial resolution, would allow for even higher-impact assessments of policy decisions across different scales, in a comparable way across the world. Global land cover time series are especially crucial for SDG monitoring and implementation, as all countries need to be comparable using uniform methods. As an extension, if we can model land cover change in future scenarios, such predictive maps can proactively influence policymaking by suggesting new policy choices and anticipating consequences of proposed legislation.

The land cover time series postprocessing methods proposed in Chapters 3 and 5 are currently under review and therefore have not been taken up by others yet. Nevertheless, in Chapter 5 I included an overview of why taking this method up in land cover map production would be highly beneficial for the users. According to the results of Chapter 5, even simple postprocessing techniques can significantly improve the stability of the land cover time series, which is an important step towards operational land cover change mapping. A product based on the results of postprocessing using a change detection algorithm, such as BFAST Lite, could include details about the time series of land cover, such as trends, number of land cover change events and their frequency, time and magnitude

of last change etc. Such a product would be highly valuable to the broad user community, and could serve as both a tool for improved policymaking and as a basis to build upon further.

6.2.2 Current advancements in global land cover fraction mapping and change detection

6.2.2.1 Advances in unsupervised and supervised change detection algorithms

The field of change detection has seen a number of new developments. These developments include improved versions of established unsupervised change detection algorithms, such as Stochastic Continuous Change Detection (S-CCD, Ye et al. (2021)) improving upon COntinuous monitoring of Land Disturbance (COLD, Zhu et al. (2020)), itself improving upon CCDC (Zhu and Woodcock, 2014). BFAST Lite, proposed in Chapter 4, is another example of an improved unsupervised change detection algorithm. New algorithms are also being developed. The scientific community has introduced, among others, EWMACD (Brooks et al., 2014), TSCCD (Yan et al., 2019), JUST (Ghaderpour and Vujadinovic, 2020) and BEAST (Zhao et al., 2019). Some change detection algorithms have been introduced by the industry, such as Prophet, a Facebook (Meta Platforms, Inc.) initiative (Taylor and Letham, 2018).

In addition to unsupervised methods, there is an increase in the availability of supervised change detection algorithms, many coming from the field of temporal deep learning, such as Gated Recurrent Units and 3D Convolutional Neural Networks. Deep learning has been an exceptionally quickly developing field that is promising for the improvement of global land cover mapping (Tuia et al., 2023). However, supervised change detection methods are less commonly used due to the added reference data requirements, and the difficulty of obtaining a robust set of land cover change reference data. In Chapter 3, I investigated the use of RF models as a supervised change detection method, however, the results were poor as the algorithms tended to overfit the training data. Therefore, supervised change detection methods are promising, but currently the reference data requirements of these methods prevent their wider use.

6.2.2.2 Advances in global land cover products and satellite imagery

The field of land cover mapping has been advancing rapidly in the previous years. It is evident from a number of new and ground-breaking global land cover products that have been produced. CCI LC (ESA, 2017) has introduced a yearly time series of moderate spatial resolution (300 m) global land cover spanning multiple decades. CGLS-LC100 (Buchhorn et al., 2020) introduced yearly-updated 100 m spatial resolution global land cover maps with land cover fraction layers, with an implementation of the methods introduced in Chapter 2. FROM-GLC10 (Gong et al., 2020) introduced a global land cover map at 10 m spatial resolution. WorldCover (Zanaga et al., 2021) showcased a yearly-updated

global land cover map at 10 m spatial resolution combining both optical (Sentinel-2) and radar (Sentinel-1) satellite imagery. Google Dynamic World (Brown et al., 2022) is a near-real-time system of land cover mapping of every incoming Sentinel-2 image, providing a dense time series of global land cover maps. In addition to these general-purpose land cover products, there have been a number of products produced that focus on a particular land cover class and its change over time, such as Global Forest Change (Hansen et al., 2013), Global Surface Water (Pekel et al., 2016), Global Human Settlement Layer (Corbane et al., 2019) and WorldCereal (Van Tricht et al., 2023).

The increase in the number and variety of global land cover maps has been enabled by the increase in publicly-available very fine spatial resolution imagery and the lengthening of satellite image time series. New satellite missions, including the entire Copernicus programme, has transformed the landscape of remote sensing by providing frequent, high quality, global and free imagery from its fleet of satellites. Sentinel-2 MSI has become the standard sensor providing data for land cover mapping projects, now with a long archive allowing time series analysis. PlanetScope satellites provide 3 m spatial resolution daily image time series, which, although commercial, is useful for collecting land cover reference data. Landsat 8 and Landsat 9 provide continuity with previous Landsat missions, which greatly benefits fields like change detection, where long time series is key for accurately detecting breaks in its time series. The combination of fine spatial resolution and long time series of Landsat 8 OLI imagery enabled the research done in Chapters 3 and 5.

6.2.2.3 Advances in land cover fraction products

Land cover fraction mapping has been advancing at a relatively slower pace, despite the continued relevance given the limitations of the discrete classification approach (Potapov et al., 2022). The primary recent land cover product with a thematically exhaustive set of land cover fractions has been CGLS-LC100. The other major development in land cover fraction mapping has been the continuation of the Vegetation Continuous Fields product, with fractions of only three classes: tree cover, short vegetation and bare ground, and focusing on the change between these classes (Song et al., 2018). The same research group later developed a hybrid discrete-fractional product with a few additional classes and tree height estimates (Hansen et al., 2022). The land cover fraction mapping field has generally tended to focus on mapping the fractions of a single land cover class, such as vegetation (Verger et al., 2023), tree cover (Pengra et al., 2015; Townshend, 2017), urban impervious surface (Kuang et al., 2021; Okujeni et al., 2018) or water (Li et al., 2019; Li et al., 2018; Liu et al., 2022). In addition, there have been multiple studies focusing on sub-pixel or super-resolution mapping, where spatial allocation of subpixels is attempted in addition to quantifying the fractions (Xu et al., 2014; Xu and Huang, 2014; Zhang et al., 2014), however, it continues to be a challenge and a product based on this line of research has yet to be developed. Therefore, this thesis is one of the few studies that are focusing on and continuing research into multi-class land cover fraction mapping.

6.2.3 Is there a future for land cover fraction mapping?

The relatively slow pace of developments in research and products providing land cover fractions is indicative of the improvements in spatial resolution of openly-available satellite image time series. As spatial resolution becomes finer, the proportion of pure pixels increases, bringing the problem closer to one solved by classification rather than regression.

Nevertheless, it is impossible to avoid mixed pixels altogether. No matter how fine the resolution, there will always be border effects. With more pure pixels, my findings on dealing with zero-inflated data in Chapter 2 will become even more relevant. The problem of discrete land cover classification underrepresenting rare and overrepresenting dominant land cover classes by area is inherent and cannot be adequately solved without using land cover fractions. Chapter 5 has exemplified how important it is to have a dense time series of continuous land cover for tracking gradual change, such as forest degradation. It would not be possible to track without either land cover fractions or their proxy, such as class probabilities.

In addition, at a different spatial resolution we can identify and map different features. The same classes at coarse resolution may not be relevant at fine resolution, and there are classes that can only be discerned at fine spatial resolution. For example, forest is a class that is defined at a moderate to coarse spatial resolution. At 10 m scale, the forest class loses meaning, as individual trees, rather than a forest, become discernible and therefore mappable. By extension, at 10 cm resolution we can tell apart, and therefore map, branches and even clumps of leaves. At 10 mm resolution we could map leaf types, branch orders and individual fruits. Conversely, given a map of trees and a set of rules and definitions, we can create a map of forests. By extension, given a map of branches and leaves, we can create a map of trees. At each level there are some aspects that are discernible and some that are not, and fraction mapping remains relevant for the aspects that are on the border of being discernible, no matter which spatial resolution we have.

The relevance of fractions depends not only on the spatial resolution, but also on the thematic resolution. If we define a unified "tree" class, then one may expect a forested area to appear homogeneous, i.e. as 100% trees. However, in reality, forests are not homogeneous; in between tree canopies there is shrub understorey and herbaceous glades. If we define separate "deciduous" and "evergreen" tree classes, then forests would appear even more heterogeneous, especially mixed forests. The more thematic detail we introduce, even at the same spatial scale, the more heterogeneity will be seen. The same holds for other classes as well: even a field of a monoculture crop is not homogeneous, with variations in the field that results from different soil background due to the variation in soil properties, treatments, and vegetation health. This variation is important to capture so that it can be precisely linked with other measurements in the field, such as evapotranspiration and

sun-induced fluorescence in fine-scale modelling studies (Vilà-Guerau de Arellano et al., 2020).

Another way of deriving fraction information is by calculating fractions at a coarser spatial resolution from a discrete classification at a finer spatial resolution. This method is useful when conducting scalability studies, e.g. scaling from canopy to landscape level (Vilà-Guerau de Arellano et al., 2020). In addition, several studies have been done on whether the costs of mapping at a very fine spatial resolution do not outweigh the benefits (Eid and Pebesma, 2023; Mîrț et al., 2022), and land cover fractions provide one means of reducing data without losing important information about the size of area occupied by each land cover class. A land cover fraction map at moderate (e.g. 300 m) spatial resolution, derived from a much finer spatial resolution map (e.g. 10 m), could be very useful for the climate modelling community that deals with global models (Zemp, 2022). For example, the global model for land-atmosphere interactions CLASS4GL (Wouters et al., 2019) makes use of the Vegetation Continuous Fields product to represent the fraction of vegetation based on MODIS data at 250 m spatial resolution, as well as other ancillary satellite imagery products at even coarser spatial resolution. When precise location is not important, a land cover fraction map can save space in memory and disk and speed up computations, while keeping the land cover class proportions highly accurate.

As Potapov et al. (2022) state, the limitations of thematic, discrete maps are known and cannot be solved without a continuous approach. Therefore, research into land cover fraction mapping will continue. The current land cover fraction products are planned to be maintained in the future by extending the time series of land cover fractions, such as in the successor project to CGLS-LC100, called LCFM (Joint Research Centre, 2023). But to truly unlock the potential of land cover fractions, more research and development is needed, in order to offer novel products that are easy for users to take up.

6.2.4 Remaining challenges

Even though my focus in this thesis has been on increasing the accuracy of global land cover fraction and global land cover fraction change products, major challenges remain to be overcome by the land cover mapping community in further improving global land cover and land cover change monitoring. The following subsections discuss these challenges and some of the potential future solutions.

6.2.4.1 Big data analysis computing infrastructures

A major limiting factor in the field of global mapping using satellite image time series is the availability of computing resources for big data analysis. Given the petabyte-size archives of modern satellite imagery, it is no longer feasible to perform computations, or even download the data, on own hardware. To achieve the best results, as shown in Chapter 2, we need to combine multiple sources of satellite image time series. The

key obstacle is that these satellites are owned by different institutions and therefore their imagery is physically stored in different locations across the world. Currently, only proprietary hyperscalers, i.e. highly scalable computing infrastructure across multiple data centres owned by companies, such as Google Cloud, Amazon Web Services and Microsoft Azure are able to mirror these enormous archives of satellite data, and only for a select number of most popular satellites. The platforms provided by these hyperscalers, such as Google Earth Engine and Microsoft Planetary Computer, are popular ways to tap into these archives among Earth observation scientists, including for land cover mapping (Parente et al., 2019a). However, given that the hyperscalers are commercial enterprises, researchers quickly run into limits to the amount of processing that can be done on these platforms, without incurring prohibitive costs. In addition, these platforms provide their own interfaces, making implementations incompatible between each other and all other computing infrastructure, which hinders reproducibility of science. At the moment, a suitable alternative to the proprietary hyperscalers is still lacking. The FAO SEPAL platform is a useful initiative that attempts to tackle the issue by providing links between Google Earth Engine and the Amazon Web Services hyperscaler infrastructure, but it is by definition limited by the hyperscaler policies and prescribed limits.

A part of the solution that is currently being developed in Europe is the European Open Science Cloud, a collaboration of European institutes and high performance computing providers that lowers the barrier to entry for scientists to perform computation tasks across existing European infrastructure (Directorate-General for Research and Innovation (European Commission), 2016). In addition, the EuroHPC joint initiative is a long-term investment into the creation and upgrade of a number of high performance computing data centres across Europe, which will increase the available computing power (Berberich et al., 2019). However, each service provider offers different interfaces and datasets to work with (if any), which leads to fragmentation and difficulties with data access, especially with regards to time series of satellite imagery. I have worked on the openEO project by the European Commission, that resulted in an open application programming interface that is a promising standard for eventually tackling the fragmentation issue. OpenEO provides a vendor-neutral interface that can be implemented on top of existing computing infrastructure, allowing users to develop scripts in their preferred programming language and submit them to run processing on a computing backend that holds data of interest (Pebesma et al., 2021; Schramm et al., 2021). The adoption of openEO in the Copernicus Data Space Ecosystem project is an important step forward that can eventually lead to better unification of computing infrastructures across Europe and eventually provide an alternative to the proprietary hyperscalers. I am also contributing to the Green Deal Data Space and its Community of Practice project of the European Commission, which aims to build a blueprint for the Green Deal Data Space. One of my aims in the project is to ensure that this new Europe-wide data sharing platform includes support for openEO, which would give it a development boost and make it more well-known and competitive. In turn,

openEO would be a very important boost for the production of global maps, especially broad land cover maps that benefit from a synergy between several sensors.

6.2.4.2 Needs for land cover reference data and features

Another important limiting factor for both global land cover mapping and land cover change detection is the availability of reference data. In this thesis I had access to the CGLS-LC100 reference data, which is made of four very extensive global datasets, which enabled me to carry out all of the research in this thesis (Table 6.1). However, despite its high quality due to data curation efforts and the involvement of local experts (Tsendbazar et al., 2021b), it nevertheless has several limitations. The major limitation was that the datasets had yearly timestamps for the change events, with very limited indications of when during the year the change took place. This limitation makes it difficult to use the datasets for validating methods based on dense time series of land cover. Temporally dense time series are important for tracking seasonal variability of land cover and for phenology studies, as well as near-real-time applications. The validation of dense time series of land cover is important for estimating its temporal accuracy. The only way to do the validation of dense time series of land cover with yearly datasets (as done in Chapter 5) is to composite the time series into yearly time steps, losing temporal precision and variability information. Similarly, using this data for model training introduces uncertainty of which day of the year (and therefore which season) a land cover observation represents.

To do more in-depth analysis with dense time series, a new reference dataset is needed. Firstly, it would record the exact time for which land cover is determined. That moment in time would generally be the timestamp of the high spatial resolution image that was available for expert interpreters to make the land cover assessment. If multiple images are available per year, then land cover should be annotated several times, with different timestamps. In addition, a hierarchical system for land cover classes could be used, e.g. a separate class for deciduous or evergreen trees could be very useful for added thematic detail, even if it is not always possible to determine the exact tree type. The added thematic detail could be useful in future works focused on harmonising reference datasets to obtain a larger training sample database. My proposed data collection approach is similar to the one used by Li et al. (2017), but the dataset would need to also have a focus on land cover change, which brings extra challenges. The number of recorded land cover change points in a given land cover dataset is typically very small, due to the relative rarity of land cover change events. Even though the datasets used in this thesis are large, the points that constituted land cover change amounted only to several percent of the whole dataset. Using change detection algorithms to stratify sampling for land cover change reference collection, as done in CGLS-LC100 (Buchhorn et al., 2021), results in a better balance between change and no change samples, but at the cost of potential bias getting introduced from the change detection algorithm. This bias leads to more samples collected from areas with clear abrupt change, under-representing gradual change. Therefore, more

research is needed into efficient sampling strategies for the collection of land cover change data, with one promising option being active learning (Settles, 2009). Lastly, the work in this thesis would not have been possible if the reference dataset did not include land cover fractions, or more specifically, a clustered design of a hundred 10 m subpixels per 100 m × 100 m sample site, from which fractions could be calculated. Therefore, future reference datasets should also include such means to obtain fraction information, even though the strategy could be different. For instance, in my work leading up to Chapter 2, I collected reference data by drawing polygons on very fine spatial resolution imagery, and calculated the area occupied by each polygon per sample site as a postprocessing step. Such approach avoids the underestimation of small classes that do not form continuous 10 m areas, such as in highly heterogeneous or sparse areas.

Another issue with data availability for global land cover mapping is the lack of certain training features for discerning between land cover classes, and especially their time series. One example is the lack of a time series of digital elevation models (DEMs). In Chapter 2 I found that terrain features from a DEM are important to distinguish several land cover classes, even though I only used a one-time DEM. For land cover change mapping, having a time series of DEMs, particularly canopy height derived from digital surface models and digital terrain models, would be especially beneficial in discerning between land cover classes that are defined by height, namely, shrubs from trees and shrubs from herbaceous vegetation. Elevation data is obtained by lidar or radar interferometry sensors or by stereophotogrammetry. The TerraSAR-X and TanDEM-X missions provide WorldDEM, a DEM that is updated, but only every five years, and it is not openly available. Only select regions are startinfig:static-errorg to be covered by annual elevation product updates, for instance, Antarctica, by making use of the lidar sensor on IceSat-2 (Shen et al., 2022). A number of countries are maintaining their own time series of DEMs, such as the Netherlands' Actueel Hoogtebestand Nederland (AHN), but the timing, sensors and methods differ between countries, and the updates are usually offered less frequently than once per year.

6.2.4.3 Performance of unsupervised change detection algorithms for land cover change

In Chapters 4 and 5, as well as in Xu et al. (2022), I explored ways to improve the accuracy of unsupervised change detection algorithms for use in land cover change analysis. Even though my research yielded improvements, there is still a lot of space for further improvement. Change detection algorithms tend to predict change too often, leading to high sensitivity (producer's accuracy of change) but low precision (user's accuracy of change), and therefore an F_1 score around 0.25 for general land cover change is the norm (Xu et al., 2022). New parametrisation in BFAST Lite detailed in Chapter 4 could tip the balance to make the algorithm a lot more conservative, although also increasing omission error, bringing the F_1 score up to 0.38. Using the hybrid approach with an RF algorithm in

Xu et al. (2022) tips the balance in favour of reducing commission errors as well, resulting in a greatly increased F_1 score of up to 0.62, but the sensitivity is nevertheless reduced to 0.53. The result is a trade-off between commission and omission errors, and given the diverse community of land cover users, both options have their own merits and uses in different communities. Nevertheless, more research is needed to improve change detection accuracy further, reducing both omission and commission errors.

Specifically, more research is needed to find ways to link breaks in time series with changes in land cover, which is not straight-forward, as breaks in a vegetation index time series do not necessarily correspond to land cover change and vice versa. In Chapter 5 I showcased the use of land cover fractions or probabilities as input into a change detection algorithm as a way to detect changes in land cover specifically, rather than merely vegetation dynamics. Further research is needed with different algorithms to determine which combination would work the best to overcome this domain shift problem. Alternatively, as I have also shown in Chapter 5, further improving the base classifier accuracy may already give a good boost also to the accuracy of change.

6.2.4.4 Accuracy of machine learning algorithms for land cover fraction prediction

Achieving a high base classifier accuracy of global land cover maps is also challenging. In Chapter 2, the highest accuracy achieved by the global land cover fraction models, including data from satellite image time series, climate modelling and terrain as well as soil mapping, was 72 % OA. Projects such as CGLS-LC100 and WorldCover achieve around 80% OA, by making extensive use of postprocessing, expert rules and feedback from users. The most difficult classes to accurately estimate are shrubs and built-up, followed by herbaceous cover. These are highly heterogeneous and at the same time fragmented classes. Shrubs are particularly difficult to quantify due to their definition that relies on vegetation height, which is currently difficult to determine using satellite image time series. Herbaceous cover can be easily confused with crop cover, especially for cases such as pasture, due to the land use rather than land cover nature of the cropland class. Built-up is both a highly heterogeneous and rare class, therefore it requires more training data to estimate well.

6.2.5 Revisiting land cover classes

The goal of land cover mapping is to characterise the biophysical cover at the surface of the ground, and satellite image time series provides the best source of information for this purpose. However, the fact that some classes can be estimated with high accuracy and some with low accuracy shows that there are gaps in information that base classifiers are trained on. For example, shrubs are defined mostly by the height of woody vegetation, and it is no surprise that this class is difficult to tell apart from others if the model does not have any access to a time series of elevation. Therefore, if we aim to improve

particular land cover classes, we need to investigate their definitions and try to fill these data gaps.

Alternatively, instead of attempting to produce maps according to predetermined classes, it may be useful to explore which land cover classes are discernible from the rest and which are not, given the currently available data. If the accuracy of a given class is low, it may be better to merge the class with another class rather than provide a highly uncertain map, as it may end up being used by users who assume that the class is mapped accurately. In some cases the class accuracy may be low only in particular geographical area due to the effects of training data availability, in which case the area should be masked out (Meyer and Pebesma, 2021). Even if the class is highly demanded by the users, the resulting mapped class needs to have high enough accuracy to actually be useful. Unsupervised classification and principal component analysis could help determine which classes can currently be separated and which cannot be. Even if there are classes that are well discernible from the rest, but are not of broad interest to current land cover users, e.g. particular type of urban infrastructure, it would be useful to separate it from more generic classes. The use cases for land cover maps are broad and many are novel, not initially foreseen by the map producers. The inclusion of extra classes would not be an issue for the users, especially if the classification system used is hierarchical and fractions are mapped, because the users can either choose a level of thematic detail to work with, or create their own combinations of classes to suit their needs. Increased separability (i.e. thematic purity) of classes could be helpful for base classifiers to perform better, as it could reduce the variability within the class, and these new classes may result in novel applications of land cover maps in the future.

The land cover mapping field has been moving towards the idea that the users, given flexible information in a land cover product, such as land cover fractions, can then create customised maps for their own use cases (Smets et al., 2020). Ultimately, the most flexible solution for mapping land cover in terms of class legend would be to map biophysical traits, rather than traditional classes. For instance, if we separately map fractions of woody and herbaceous vegetation, deciduous and evergreen canopies, leaf type and canopy height, we can reconstruct the class "open deciduous broadleaf forest" (UN LCCS class A12A3A11B2D1E2 as per Tsendbazar et al. 2021b) as areas with 15 to 70 % woody vegetation fraction, which is more than 50 % broadleaf and deciduous, and where the canopy height is over 5 m. The resulting class could also be offered as a fraction layer, i.e. "deciduous broadleaf forest", along with fractions of herbaceous and shrub cover that would provide information about the type of the open forest. Mapping of biophysical traits is an evolving field, with studies using both hyperspectral (Zhang et al., 2021) and multispectral (Ma et al., 2019) satellite imagery for deriving plant leaf and canopy traits. Another use of the approach of mapping biophysical traits would be to create downstream products, e.g., a number of traits could be used to define the amount of fuel each area contains, resulting in a potential fuel map; from this product, a fire risk map could be

further derived. The process of linking traits to products could be aided by explainable AI models, that could automatically select the relevant traits to use to derive the downstream model. The challenge of this approach is to make sure that the traits that need to be mapped have sufficient accuracy, because any errors in the base biophysical traits would propagate into all downstream applications. Some of the necessary traits may not be possible to discern with sufficient accuracy given the currently available data.

6.2.6 Next steps and new avenues in global land cover fraction mapping and map updating

6.2.6.1 Use of deep learning

To meet the challenges and user requirements for global land cover mapping, several avenues can already be explored. The first, highly promising avenue to explore is the application of deep learning for supervised land cover fraction mapping and change detection. Neural networks are natively capable of producing land cover fraction maps (Zhang and Foody, 2001), and the field of deep learning, based on neural networks, has experienced rapid development over the recent years, with new algorithms continuing to be developed (Tuia et al., 2023). Convolutional neural networks can take into account the spatial component, whereas recurrent neural networks can take into account the temporal component, and these models have been shown to perform well for land cover mapping (Parente et al., 2019b). Attention models have also performed well in land use classification, as they can combine both time and space dimensions (Masolele et al., 2021; Masolele et al., 2022). Deep learning has already been used to produce the Google Dynamic World product (Brown et al., 2022). Therefore, using deep learning for land cover fraction mapping is promising for improving the accuracy of the base land cover fraction predictions. The key limiting factors for the uptake of deep learning techniques are the increased demands for reference data, which is difficult and costly to collect; increased demands for computing power; difficulties in choosing the optimal model architecture and hyperparameters; limitations in the handling of time series data, such as missing values; lack of model interpretability; and a lack of user-friendly implementations. Nevertheless, researchers can evaluate deep learning algorithm suitability for land cover fraction mapping at a global scale by applying the algorithms on patches surrounding sample sites of available reference data.

6.2.6.2 Use of postprocessing algorithms

Another opportunity to improve global land cover fraction time series is to perform further research on temporal dynamics of land cover change. One option is to continue the research lines from Chapters 3 and 5 by combining Markov models with BFAST Lite postprocessing. Markov models can take into account transition probabilities, and a combination of training data and statistical data can be used to derive realistic transition probabilities for each land cover class. BFAST Lite can be used as an initial noise removal

step, before a Markov model. Or vice versa, it could be applied to time series postprocessed using a Markov model. Using a spatiotemporal Markov model, such as MRF, could further improve postprocessing results. In addition, exploring more complex Markov models is a promising line of research as well. HMMs, adapted to land cover fractions, could perform better than Markov chains explored in Chapter 3.

Temporal deep learning algorithms could also be promising for postprocessing purposes, as their ability to take the time dimension into account could result in better land cover change estimations. The limitation of these supervised algorithms is the need for a large reference dataset to train on, which is especially difficult to obtain for land cover change. As I found in Chapter 3, the reference data also needs to be unbiased and representative for the whole world, otherwise overfitting can lead to poor results.

6.2.6.3 New satellite sensors

Upcoming satellite missions have a potential to fill large gaps in the data needs for improving global land cover maps. Whereas Landsat 9 provides a useful reduction in the revisit time of the OLI sensor it shares with Landsat 8, the next generation Landsat Next mission is planned to be a big leap forward for land cover mapping. Landsat Next, to be launched in 2030, is planned to include a sensor that will cover all of the bands of OLI and Sentinel-2 MSI, plus ten new bands that could be helpful to discern between different land cover classes, and at a finer spatial resolution of 10 to 20 m. In addition, Landsat Next will be a constellation of three satellites, reducing the revisit time to 6 days. This new satellite mission promises to improve the timeliness of land cover change detection and improve the accuracy of global land cover products.

The ESA FLEX mission, currently scheduled for launch in 2025, will for the first time provide data on sun-induced chlorophyll fluorescence (SIF) at the global scale and with a frequent revisit time, with observations that are aligned with Sentinel-3 imagery. My work on SIF at the airborne scale (Wang et al., 2022b) and the unmanned aerial vehicle scale (Wang et al., 2021) has shown that this parameter is useful for discerning between vegetation types, including crop types, and therefore could help reduce the confusion between vegetation classes and broaden the number of classes that can be accurately mapped. SIF is also important for monitoring crop status, such as the effects of droughts (Wang et al., 2022a). FLEX will enable upscaling the monitoring of SIF and its derivatives to the globe, and therefore enable the tracking of global plant phenology and further applications such as evapotranspiration modelling (Vilà-Guerau de Arellano et al., 2020). SIF is a key parameter needed for mapping and monitoring of plant biophysical traits (Zhang et al., 2021), and therefore also a potential enabler of the next generation of land cover maps.

Time series of general-purpose hyperspectral data has a potential to give a boost to global land cover efforts as well. The ESA CHIME mission, scheduled for launch in 2028, will

provide open hyperspectral imagery at 30 m spatial resolution, which is also planned to be linked to the Sentinel-2 satellites. Data from both FLEX and CHIME will provide a new opportunity to make use of the full reflectance spectrum of land cover, therefore enabling machine learning and deep learning algorithms to discern land cover classes easier by making use of class-specific features in the reflectance spectrum. On the other hand, the amount of data that will be produced by these missions will prove an even larger big data analysis challenge than the currently available satellite imagery, therefore more research in reducing data amounts without reducing land cover map quality will be necessary in the near future.

There are upcoming improvements in satellite active remote sensing that can benefit land cover mapping as well. The Global Ecosystem Dynamics Investigation (GEDI) mission, featuring a lidar sensor mounted on the International Space Station (ISS), has recently been extended to continue in 2024. Originally planned for two years, it may now continue potentially until the decommissioning of the space station itself. The return of GEDI will provide an opportunity to track changes in tree height over a longer time series and with higher accuracy and precision. However, due to the dependence on the ISS orbit, it will not provide global coverage, limiting its use for global land cover mapping.

The upcoming ESA Biomass mission, planned to launch in 2024, includes a P-band SAR instrument that is capable of radar interferometry, therefore it can both provide more insight into the biomass of vegetation and potentially lead to a more frequently updated global DEM. P-band SAR is also important for land cover mapping due to its ability to penetrate clouds and leaves, therefore improving coverage over rainforests and other cloudy areas. It is particularly suited for near-real-time change detection, as the wait for a cloud-free optical observation is eliminated. Unfortunately, the ESA Biomass mission will not have global coverage either, as P-band usage is restricted by the International Telecommunication Union due to interference concerns (Carreiras et al., 2017).

The Sentinel-1 mission, featuring an L-band SAR instrument, has been important for tracking change in land cover over clouded areas and for improving the timing in near-real-time tracking applications. It has been one of the two key sources of data for the WorldCover maps. Given the anomaly that inhibited the use of the Sentinel-1B satellite, it will be very important for the replacement Sentinel-1C and Sentinel-1D satellites to be deployed. Both replacement satellites are currently scheduled for launch later in 2024. Their added coverage will reduce the revisit time and improve near-real-time change detection timing of the Sentinel-1 mission.

6.2.7 A future for global land cover fraction mapping and monitoring

The field of land cover mapping is well on its way of meeting the user requirements of its diverse communities. The requirements set by the climate community (Zemp, 2022) have been largely met at the “breakthrough” level, i.e. having a yearly 300 m land cover

map time series covering a decade, and updated every year, at around 80% accuracy at monotemporal level. Even the finer spatial resolution (30 m or less) breakthrough requirements are now being met. This thesis has contributed towards improving the accuracy and stability of global land cover maps, for continuous variables in the form of land cover fractions. As discussed in Chapter 2, the climate community is not only a major user of land cover maps, but climate data itself is also an important auxiliary input that helps improve land cover maps. The recent developments in the climate community, especially the Destination Earth initiative (Nativi et al., 2021), are promising for a further improvement of climate data, that could in turn be useful for other broad communities, including land cover mapping. Destination Earth is also a showcase in a novel way of carrying out big data analysis, and its results could further improve the landscape of big data analysis needed for the production of global land cover maps.

To achieve the "goal" level of the user requirements, global land cover maps still need to be improved. In particular, land cover change and its temporal stability remain challenges that require further improvement to meet the user needs. With the recent increase in available satellite data, planned future satellite missions, higher availability and improvements to auxiliary data sources, accumulation and improvement of reference data for model training and validation, improved algorithms, rise of deep learning methods and an increase in big data platforms, the field of global land cover mapping is moving towards further improvements on the way to meeting the users' end goals.

Existing land cover and land cover change mapping algorithms will continue to be improved and new algorithms will be developed. A user-oriented mindset is key to the success of improved algorithms, however. Algorithms that are simple to apply by a broad user community will also have a bigger impact. Making them free and open-source is an important part of their future uptake: such algorithms can be inspected, verified and improved by the user community. The same is true for entire processing chains, and global land cover processing chains should also be openly available to the user community, both to apply them and to inspect them. That is also a driving objective of the Open Earth Monitor Cyberinfrastructure project that I am also contributing to: to provide tools for producing global maps that would be open, transparent, and user-friendly. User-oriented tools are key for the creation of user-oriented and flexible global land cover maps, and further into the future, of flexible estimation of biophysical parameters, from which these maps can be built in a transparent way, for the benefit of many communities across different disciplines.

This thesis has contributed towards making global land cover and its change more multi-faceted (Zhu et al., 2022) by incorporating not only location and time, but also target land cover fractions and change metrics. Identifying land cover change drivers continues to be elusive; however, it is a goal that future studies can build upon. A global product that identifies change drivers will have a large impact on policymaking, as change drivers can

be acted upon, and thus quantifying their impact will lead to better-informed decisions. Monitoring land cover and its change, in detail and uniformly across the world, is key for assessing the progress we make towards achieving the Sustainable Development Goals.

Acronyms

AIC Akaike's Information Criterion. 71, 82

BEAST Bayesian estimator of abrupt change, seasonality & trend. 98, 120

BFAST Breaks For Additive Season and Trend. 4, 8, 65, 68, 69, 70, 71, 72, 73, 74, 75, 78, 79, 80, 81, 82, 83, 84, 85, 90, 94, 95, 96, 97, 98, 102, 103, 104, 105, 106, 108, 109, 114, 116, 117, 118, 119, 120, 126, 129

BIC Bayesian Information Criterion. 71, 73, 80, 82

CART classification and regression tree. 17, 35, 115

CCDC Continuous Change Detection and Classification. 4, 69, 90, 120

CCI LC ESA Climate Change Initiative Land Cover. 3, 6, 11, 120

CGLS-LC100 Copernicus Global Land Service Land Cover 100 m. 4, 5, 6, 7, 10, 11, 12, 13, 14, 15, 48, 70, 76, 89, 91, 94, 118, 119, 120, 121, 123, 125, 127

DEM digital elevation model. 126, 131

ESA European Space Agency. 2, 5, 6, 130, 131

EVI Enhanced Vegetation Index. 15, 93

EWMACD Exponentially Weighted Moving Average Change Detection. 69, 120

FNC fuzzy nearest centroid. 36

FROM-GLC10 Finer Resolution Observation and Monitoring Global Land Cover. 11, 21, 26, 28, 120

GDAL the Geospatial Data Abstraction Library. 16

GEDI Global Ecosystem Dynamics Investigation. 131

GFCC Global Forest Cover Change. 21, 26, 28

GHSL Global Human Settlement Layer. 21, 26, 28

GLM general linear regression model. 34, 36

GPP gross primary productivity. 30

GSW Global Surface Water. 21, 26, 28

HMM Hidden Markov Model. 5, 7, 46, 47, 65, 130

ICM Iterated Conditional Modes. 55

IIASA the International Institute for Applied Systems Analysis. 13, 14, 48, 76, 91, 92, 114

ISS International Space Station. 131

JUST Jumps Upon Spectrum and Trend. 69, 120

LCCS Land Cover Classification Scheme. 13, 49, 128

LCFM Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring service. 4, 123

LOESS locally estimated scatterplot smoothing. 15, 50, 90, 91, 95, 97, 98, 102, 103, 104, 106

LWZ information criterion of Liu, Wu and Zidek. 71, 73, 78, 79, 80, 82, 84, 95

MAE mean absolute error. 19, 20, 21, 22, 24, 26, 27, 28, 29, 31, 34, 35, 36, 37, 39, 40, 55, 56, 57, 58, 60, 99, 109, 111, 115, 118

ME mean error. 19, 20, 26, 42, 99, 111

MLP multi-layer perceptron. 12, 18, 36, 115

MLR multinomial logistic regression. 34, 36

MODIS Moderate Resolution Imaging Spectroradiometer. 3, 5, 68, 123

MRF Markov Random Field. 5, 47, 65, 130

MSI MultiSpectral Instrument. 31, 94, 121, 130

NASA National Aeronautics and Space Administration. 2

NBR Normalised Burn Ratio. 93

NDMI Normalised Difference Moisture Index. 15, 39, 40

NDSI Normalised Difference Snow Index. 93

NDVI Normalised Difference Vegetation Index. 15, 16, 39, 40, 50, 90, 91, 93, 95, 96, 97, 98, 102, 103, 104, 105

NIR near infra-red. 50

NIR_v near infra-red of vegetation. 16, 39, 40, 75

NN neural network. 12, 34, 36

NSE the Nash–Sutcliffe model efficiency coefficient. 20

OA overall accuracy. 20, 22, 24, 31, 36, 56, 60, 127

OLI Operational Land Imager. 50, 91, 92, 93, 96, 116, 121, 130

OLS ordinary least squares. 20, 99

OLS-MOSUM ordinary least squares residual moving sum. 71, 72, 73, 76, 78, 80

OSAVI Optimised Soil-Adjusted Vegetation Index. 16

PA producer accuracy. 20

PLS partial least squares. 34, 36

RF random forest. 12, 18, 21, 22, 23, 24, 26, 27, 28, 29, 31, 35, 36, 37, 39, 40, 42, 44, 45, 46, 47, 48, 50, 51, 52, 53, 55, 57, 59, 60, 61, 62, 63, 64, 65, 66, 95, 115, 116, 117, 118, 120, 126

RMAE relative mean absolute error. 19, 37

RME relative mean error. 19

RMSD root mean squared deviation. 29, 95, 98, 111

RMSE root mean squared error. 18, 19, 20, 21, 22, 24, 26, 27, 28, 29, 35, 36, 37, 40, 42, 55, 56, 57, 58, 59, 99, 111, 115, 117

RRMSE relative root mean squared error. 19, 37

RSS residual sum of squares. 71, 82, 99, 116

SAR synthetic aperture radar. 30, 131

SCM subpixel confusion-uncertainty matrix. 20, 24, 27, 31, 35, 55, 56, 60

SDG Sustainable Development Goal. 2, 3, 11, 119

SIF sun-induced chlorophyll fluorescence. 30, 130

SMA spectral mixture analysis. 12

STL Seasonal decomposition of Time series by LOESS. 71, 72, 81

SVM support vector machine. 12, 34, 36, 45, 115

SWIR shortwave infra-red. 50

TCg Tasselled Cap (greenness). 93

TPI Terrain Position Index. 16

TSCCD Time-Series Classification approach based on Change Detection. 69, 120

UA user accuracy. 20

USGS United States Geological Survey. 2, 50

UTM Universal Transverse Mercator. 49

VI vegetation index. 4, 5, 16, 17, 31, 36, 50, 118

WUR Wageningen University & Research. 13, 14, 48, 91, 114

References

Abercrombie, S. P. and M. A. Friedl (2016). "Improving the Consistency of Multitemporal Land Cover Maps Using a Hidden Markov Model". *IEEE Transactions on Geoscience and Remote Sensing* 54.2, 703–713. DOI: 10.1109/TGRS.2015.2463689.

Adams, J. B., D. E. Sabol, V. Kapos, R. Almeida Filho, D. A. Roberts, M. O. Smith, and A. R. Gillespie (1995). "Classification of multispectral images based on fractions of endmembers: Application to land-cover change in the Brazilian Amazon". *Remote sensing of Environment* 52.2, 137–154. DOI: 10.1016/0034-4257(94)00098-8.

Allaire, J. and F. Chollet (2018). *keras: R Interface to 'Keras'*. R package version 2.2.4. URL: https://CRAN.R-project.org/package=keras.

Arino, O., D. Gross, F. Ranera, M. Leroy, P. Bicheron, C. Brockman, P. Defourny, C. Vancutsem, F. Achard, L. Durieux, L. Bourg, J. Latham, A. D. Gregorio, R. Witt, M. Herold, J. Sambale, S. Plummer, and J. L. Weber (2007). "GlobCover: ESA service for global land cover from MERIS". In: *2007 IEEE International Geoscience and Remote Sensing Symposium*, 2412–2415. DOI: 10.1109/IGARSS.2007.4423328.

Asner, G. P., M. Keller, R. Pereira, and J. C. Zweede (2002). "Remote sensing of selective logging in Amazonia". *Remote Sensing of Environment* 80.3, 483–496. DOI: 10.1016/S0034-4257(01)00326-1.

Awty-Carroll, K., P. Bunting, A. Hardy, and G. Bell (2019). "An Evaluation and Comparison of Four Dense Time Series Change Detection Methods Using Simulated Data". *Remote Sensing* 11.23, 2779. DOI: 10.3390/rs11232779.

Badgley, G., C. B. Field, and J. A. Berry (2017). "Canopy near-infrared reflectance and terrestrial photosynthesis". *Science Advances* 3.e1602244. DOI: 10.1126/sciadv.1602244.

Bai, J. and P. Perron (2003). "Computation and analysis of multiple structural change models". *Journal of Applied Econometrics* 18.1, 1–22. DOI: 10.1002/jae.659.

Baranowski, E., N. Thevs, A. Khalil, A. Baibagyssov, M. Iklassov, R. Salmurzauli, S. Nurtazin, and V. Beckmann (2020). "Pastoral Farming in the Ili Delta, Kazakhstan, under Decreasing Water Inflow: An Economic Assessment". *Agriculture* 10.7, 281. DOI: 10.3390/agriculture10070281.

Bartholomé, E. and A. S. Belward (2005). "GLC2000: a new approach to global land cover mapping from Earth observation data". *International Journal of Remote Sensing* 26.9, 1959–1977. DOI: 10.1080/01431160412331291297.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). "Fitting Linear Mixed-Effects Models Using lme4". *Journal of Statistical Software* 67, 1–48. DOI: 10.18637/jss.v067.i01.

Belgiu, M. and L. Drăguţ (2016). "Random forest in remote sensing: A review of applications and future directions". *ISPRS Journal of Photogrammetry and Remote Sensing* 114, 24–31. DOI: 10.1016/j.isprsjprs.2016.01.011.

Berberich, F., J. Liebmann, J.-P. Nominé, O. Pineda, P. Segers, and V. Teodor (2019). "European HPC Landscape". In: *2019 15th International Conference on eScience (eScience)*, 471–478. DOI: 10.1109/eScience.2019.00062.

Bernardino, P. N., W. D. Keersmaecker, R. Fensholt, J. Verbesselt, B. Somers, and S. Horion (2020). "Global-scale characterization of turning points in arid and semi-arid ecosystem functioning". *Global Ecology and Biogeography* 29.7, 1230–1245. DOI: 10.1111/geb.13099.

Besag, J. (1986). "On the Statistical Analysis of Dirty Pictures". *Journal of the Royal Statistical Society: Series B (Methodological)* 48.3, 259–279. DOI: 10.1111/j.2517-6161.1986.tb01412.x.

Bogaert, P., C. Lamarche, and P. Defourny (2022). "Hidden Markov Models for Annual Land Cover Mapping—Increasing Temporal Consistency and Completeness". *IEEE Transactions on Geoscience and Remote Sensing* 60, 1–14. DOI: 10.1109/TGRS.2021.3123738.

Bontemps, S., M. Herold, L. Kooistra, A. van Groenestijn, A. Hartley, O. Arino, I. Moreau, and P. Defourny (2012). "Revisiting land cover observation to address the needs of the climate modeling community". *Biogeosciences* 9.6, 2145–2157. DOI: 10.5194/bg-9-2145-2012.

Boucher, A., K. Seto, and A. Journel (2006). "A Novel Method for Mapping Land Cover Changes: Incorporating Time and Space With Geostatistics". *IEEE Transactions on Geoscience and Remote Sensing* 44.11, 3427–3435. DOI: 10.1109/TGRS.2006.879113.

Breiman, L. (2001). "Random Forests". *Machine Learning* 45.1, 5–32. DOI: 10.1023/A:1010933404324.

Broadbent, E. N., D. J. Zarin, G. P. Asner, M. Peña-Claros, A. Cooper, and R. Littell (2006). "Recovery of forest structure and spectral properties after selective logging in lowland Bolivia". *Ecological Applications* 16.3, 1148–1163. DOI: 10.1890/1051-0761(2006)016[1148:ROFSAS]2.0.CO;2.

Brooks, E. B., R. H. Wynne, V. A. Thomas, C. E. Blinn, and J. W. Coulston (2014). "On-the-Fly Massively Multitemporal Change Detection Using Statistical Quality Control

Charts and Landsat Data". *IEEE Transactions on Geoscience and Remote Sensing* 52.6. DOI: 10.1109/TGRS.2013.2272545.

Brown, C. F., S. P. Brumby, B. Guzder-Williams, T. Birch, S. B. Hyde, J. Mazzariello, W. Czerwinski, V. J. Pasquarella, R. Haertel, S. Ilyushchenko, K. Schwehr, M. Weisse, F. Stolle, C. Hanson, O. Guinan, R. Moore, and A. M. Tait (2022). "Dynamic World, Near real-time global 10 m land use land cover mapping". *Scientific Data* 9.1, 251. DOI: 10.1038/s41597-022-01307-4.

Buchhorn, M., L. Bertels, B. Smets, M. Lesiv, and N.-E. Tsendbazar (2019a). *Moderate dynamic land cover 100 m algorithm theoretical basis document*. Version: 2.0. DOI: 10.5281/zenodo.3606446.

Buchhorn, M., L. Bertels, B. Smets, B. D. Roo, M. Lesiv, N.-E. Tsendbazar, D. Masiliūnas, and L. Li (2021). *Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Algorithm Theoretical Basis Document*. Tech. rep. Version Dataset v3.0, doc issue 3.4. DOI: 10.5281/zenodo.4723924.

Buchhorn, M., M. Lesiv, N.-E. Tsendbazar, M. Herold, L. Bertels, and B. Smets (2020). "Copernicus Global Land Cover Layers—Collection 2". *Remote Sensing* 12.6, 1044. DOI: 10.3390/rs12061044.

Buchhorn, M., B. Smets, L. Bertels, M. Lesiv, N.-E. Tsendbazar, M. Herold, and S. Fritz (2019b). *Copernicus Global Land Service: Land Cover 100m: epoch 2015: Globe*. Version: V2.0.2. Zenodo. DOI: 10.5281/zenodo.3243509.

Burger, R. and D. Masiliūnas (2024). *Code for land cover fraction mapping using Markov chains*. DOI: 10.5281/zenodo.10572684.

Burrell, A. L., J. P. Evans, and Y. Liu (2017). "Detecting dryland degradation using Time Series Segmentation and Residual Trend analysis (TSS-RESTREND)". *Remote Sensing of Environment* 197.1, 43–57. DOI: 10.1016/j.rse.2017.05.018.

Cai, S., D. Liu, D. Sulla-Menashe, and M. A. Friedl (2014). "Enhancing MODIS land cover product with a spatial–temporal modeling algorithm". *Remote Sensing of Environment* 147, 243–255. DOI: 10.1016/j.rse.2014.03.012.

Cannon, R. L., J. V. Dave, J. C. Bezdek, and M. M. Trivedi (1986). "Segmentation of a Thematic Mapper Image Using the Fuzzy c-Means Clusterng Algorthm". *IEEE Transactions on Geoscience and Remote Sensing* GE-24.3, 400–408. DOI: 10.1109/TGRS.1986.289598.

Carreiras, J. M. B., S. Quegan, T. Le Toan, D. Ho Tong Minh, S. S. Saatchi, N. Carvalhais, M. Reichstein, and K. Scipal (2017). "Coverage of high biomass forests by the ESA BIOMASS mission under defense restrictions". *Remote Sensing of Environment* 196, 154–162. DOI: 10.1016/j.rse.2017.05.003.

Carter, S. and M. Herold (2019). *Specifications of land cover datasets for SDG indicator monitoring*. Tech. rep. Global Observation for Forest Cover and Land Dynamics (GOFC-GOLD).

Cauchy, M. A. (1847). "Méthode générale pour la résolution des systemes d'équations simultanées". *Comptes Rendus Hebd. Seances Acad. Sci.* 25, 536–538.

Chen, B., B. Xu, and P. Gong (2021). "Mapping essential urban land use categories (EULUC) using geospatial big data: Progress, challenges, and opportunities". *Big Earth Data* 5.3, 410–441. DOI: 10.1080/20964471.2021.1939243.

Chen, J., J. Chen, A. Liao, X. Cao, L. Chen, X. Chen, C. He, G. Han, S. Peng, M. Lu, W. Zhang, X. Tong, and J. Mills (2015). "Global land cover mapping at 30m resolution: A POK-based operational approach". *ISPRS Journal of Photogrammetry and Remote Sensing*. Global Land Cover Mapping and Monitoring 103, 7–27. DOI: 10.1016/j.isprsjprs.2014.09.002.

Chu, L., Y. Zou, D. Masiliūnas, T. Blaschke, and J. Verbesselt (2021). "Assessing the impact of bridge construction on the land use/cover and socio-economic indicator time series: A case study of Hangzhou Bay Bridge". *GIScience & Remote Sensing* 58.2, 199–216. DOI: 10.1080/15481603.2020.1868212.

Claverie, M., J. Ju, J. G. Masek, J. L. Dungan, E. F. Vermote, J.-C. Roger, S. V. Skakun, and C. Justice (2018). "The Harmonized Landsat and Sentinel-2 surface reflectance data set". *Remote Sensing of Environment* 219, 145–161. DOI: 10.1016/j.rse.2018.09.002.

Cleveland, R. B., W. S. Cleveland, J. E. McRae, and I. J. Terpenning (1990). "STL: A seasonal-trend decomposition". *Journal of Official Statistics* 6.1, 3–73.

Cleveland, W. S. (1979). "Robust Locally Weighted Regression and Smoothing Scatterplots". *Journal of the American Statistical Association* 74.368, 829–836. DOI: 10.1080/01621459.1979.10481038.

Cleveland, W. S., E. Grosse, and W. M. Shyu (1992). "Local Regression Models". In: *Statistical Models in S*. Routledge, 68.

Cohen, W. B., S. P. Healey, Z. Yang, S. V. Stehman, C. K. Brewer, E. B. Brooks, N. Gorelick, C. Huang, M. J. Hughes, R. E. Kennedy, T. R. Loveland, G. G. Moisen, T. A. Schroeder, J. E. Vogelmann, C. E. Woodcock, L. Yang, and Z. Zhu (2017). "How similar are forest disturbance maps derived from different Landsat time series algorithms?" *Forests* 8.4, 98. DOI: 10.3390/f8040098.

Colditz, R. R., M. Schmidt, C. Conrad, M. C. Hansen, and S. Dech (2011). "Land cover classification with coarse spatial resolution data to derive continuous and discrete maps for complex regions". *Remote Sensing of Environment* 115.12, 3264–3275. DOI: 10.1016/j.rse.2011.07.010.

Congalton, R. G., J. Gu, K. Yadav, P. Thenkabail, and M. Ozdogan (2014). "Global Land Cover Mapping: A Review and Uncertainty Analysis". *Remote Sensing* 6.12, 12070–12093. DOI: 10.3390/rs61212070.

Corbane, C., A. Florczyk, M. Pesaresi, P. Politis, and V. Syrris (2018). *GHS-BUILT R2018A - GHS built-up grid, derived from Landsat, multitemporal (1975-1990-2000-2014)*. DOI: 10.2905/jrc-ghsl-10007.

Corbane, C., M. Pesaresi, T. Kemper, P. Politis, A. J. Florczyk, V. Syrris, M. Melchiorri, F. Sabo, and P. Soille (2019). "Automated global delineation of human settlements from 40 years of Landsat satellite data archives". *Big Earth Data* 3.2, 140–169. DOI: 10.1080/20964471.2019.1625528.

Coyle, J. R., N. S. Hejazi, I. Malenica, and O. Sofrygin (2020). *sl3: Pipelines for Machine Learning and Super Learning*. R package version 1.3.7. DOI: 10.5281/zenodo.1342293. URL: https://github.com/tlverse/sl3.

Dierckx, W., S. Sterckx, I. Benhadj, S. Livens, G. Duhoux, T. V. Achteren, M. Francois, K. Mellab, and G. Saint (2014). "PROBA-V mission for global vegetation monitoring: standard products and image quality". *International Journal of Remote Sensing* 35.7, 2589–2614. DOI: 10.1080/01431161.2014.883097.

Directorate-General for Research and Innovation (European Commission) (2016). *Realising the European open science cloud: first report and recommendations of the Commission high level expert group on the European open science cloud*. Luxembourg: Publications Office of the European Union. DOI: 10.2777/940154.

Dozat, T. (2016). "Incorporating Nesterov Momentum into Adam". In: *International Conference on Learning Representations*. Puerto Rico.

Dreyfus, S. E. (1990). "Artificial neural networks, back propagation, and the Kelley-Bryson gradient procedure". *Journal of Guidance, Control, and Dynamics* 13.5, 926–928. DOI: 10.2514/3.25422.

Dutrieux, L. P., C. C. Jakovac, S. H. Latifah, and L. Kooistra (2016). "Reconstructing land use history from Landsat time-series: Case study of a swidden agriculture system in Brazil". *International Journal of Applied Earth Observation and Geoinformation* 47, 112–124. DOI: 10.1016/j.jag.2015.11.018.

Dutrieux, L. P., J. Verbesselt, L. Kooistra, and M. Herold (2015). "Monitoring forest cover loss using multiple data streams, a case study of a tropical dry forest in Bolivia". *ISPRS Journal of Photogrammetry and Remote Sensing*. Multitemporal remote sensing data analysis 107, 112–125. DOI: 10.1016/j.isprsjprs.2015.03.015.

Dwivedi, R., A. Kumar, S. K. Ghosh, and P. S. Roy (2012). "Optimisation of Fuzzy Based Soft Classifiers for Remote Sensing Data". *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 39B3, 385–390. DOI: 10.5194/isprsarchives-XXXIX-B3-385-2012.

Eid, Y. and E. Pebesma (2023). "Is maximizing spatial resolution worth the computational cost?" In: EGU23-14915. Conference Name: EGU23. DOI: 10.5194/egusphere-egu23-14915.

ESA (2017). *Land Cover CCI Product User Guide Version 2*. Tech. rep. European Space Agency.

Fang, X., Q. Zhu, L. Ren, H. Chen, K. Wang, and C. Peng (2018). "Large-scale detection of vegetation dynamics and their potential drivers using MODIS images and BFAST: A case study in Quebec, Canada". *Remote Sensing of Environment* 206, 391–402. DOI: 10.1016/j.rse.2017.11.017.

Fick, S. E. and R. J. Hijmans (2017). "WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas". *International Journal of Climatology* 37.12, 4302–4315. DOI: 10.1002/joc.5086.

Foody, G. M. (1996). "Approaches for the production and evaluation of fuzzy land cover classifications from remotely-sensed data". *International Journal of Remote Sensing* 17.7, 1317–1340. DOI: 10.1080/01431169608948706.

– (1997). "Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network". *Neural Computing & Applications* 5.4, 238–247. DOI: 10.1007/BF01424229.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang (2010). "MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets". *Remote Sensing of Environment* 114.1, 168–182. DOI: 10.1016/j.rse.2009.08.016.

Friedman, J. H., T. Hastie, and R. Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent". *Journal of Statistical Software* 33.1, 1–22. DOI: 10.18637/jss.v033.i01.

Gagniuc, P. A. (2017). *Markov Chains: From Theory to Implementation and Experimentation*. Wiley & Sons.

Gao, J. and B. C. O'Neill (2020). "Mapping global urban land for the 21st century with data-driven simulations and Shared Socioeconomic Pathways". *Nature Communications* 11.1, 2302. DOI: 10.1038/s41467-020-15788-7.

GDAL/OGR contributors (2020). *GDAL/OGR Geospatial Data Abstraction software Library*. Open Source Geospatial Foundation. URL: https://gdal.org.

Gessner, U., M. Machwitz, C. Conrad, and S. Dech (2013). "Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles". *Remote Sensing of Environment* 129, 90–102. DOI: 10.1016/j.rse.2012.10.026.

Ghaderpour, E. and T. Vujadinovic (2020). "Change Detection within Remotely Sensed Satellite Image Time Series via Spectral Analysis". *Remote Sensing* 12.23, 4001. DOI: 10.3390/rs12234001.

Gong, P., X. Li, J. Wang, Y. Bai, B. Chen, T. Hu, X. Liu, B. Xu, J. Yang, W. Zhang, and Y. Zhou (2020). "Annual maps of global artificial impervious area (GAIA) between 1985 and 2018". *Remote Sensing of Environment* 236, 111510. DOI: 10.1016/j.rse.2019.111510.

Gong, P., H. Liu, M. Zhang, C. Li, J. Wang, H. Huang, N. Clinton, L. Ji, W. Li, Y. Bai, B. Chen, B. Xu, Z. Zhu, C. Yuan, H. Ping Suen, J. Guo, N. Xu, W. Li, Y. Zhao, J. Yang, C. Yu, X. Wang, H. Fu, L. Yu, I. Dronova, F. Hui, X. Cheng, X. Shi, F. Xiao, Q. Liu, and L. Song (2019). "Stable classification with limited sample: transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017". *Science Bulletin* 64.6, 370–373. DOI: 10.1016/j.scib.2019.03.002.

Gong, W., S. Fang, G. Yang, and M. Ge (2017). "Using a Hidden Markov Model for Improving the Spatial-Temporal Consistency of Time Series Land Cover Classification". *ISPRS International Journal of Geo-Information* 6.10, 292. DOI: 10.3390/ijgi6100292.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone". *Remote Sensing of Environment* 202, 18–27. DOI: 10.1016/j.rse.2017.06.031.

Hafen, R. (2016). *stlplus: Enhanced Seasonal Decomposition of Time Series by Loess*. R package version 0.5.1. URL: https://CRAN.R-project.org/package=stlplus.

Hall, D. K., G. A. Riggs, and V. V. Salomonson (1995). "Development of methods for mapping global snow cover using moderate resolution imaging spectroradiometer data". *Remote Sensing of Environment* 54.2, 127–140. DOI: 10.1016/0034-4257(95)00137-P.

Hampel, F., E. Ronchetti, P. Rousseeuw, and W. Stahel (2005). "Introduction and Motivation". In: *Robust Statistics*. John Wiley and Sons, Ltd. Chap. 1, 1–77. DOI: https://doi.org/10.1002/9781118186435.ch1.

Hamunyela, E., S. Rosca, A. Mîrț, E. Engle, M. Herold, F. Gieseke, and J. Verbesselt (2020). "Implementation of BFASTmonitor Algorithm on Google Earth Engine to Support Large-Area and Sub-Annual Change Monitoring Using Earth Observation Data". *Remote Sensing* 12.18, 2953. DOI: 10.3390/rs12182953.

Hamunyela, E., J. Verbesselt, and M. Herold (2016). "Using spatial context to improve early detection of deforestation from Landsat time series". *Remote Sensing of Environment* 172, 126–138. DOI: 10.1016/j.rse.2015.11.006.

Hansen, M. C., R. S. DeFries, J. R. G. Townshend, M. Carroll, C. Dimiceli, and R. A. Sohlberg (2003). "Global Percent Tree Cover at a Spatial Resolution of 500 Meters: First Results of the MODIS Vegetation Continuous Fields Algorithm". *Earth Interactions* 7.10, 1–15. DOI: 10.1175/1087-3562(2003)007<0001:GPTCAA>2.0.CO;2.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend (2013). "High-Resolution Global Maps of 21st-Century Forest Cover Change". *Science* 342.6160, 850–853. DOI: 10.1126/science.1244693.

Hansen, M. C., A. Egorov, D. P. Roy, P. Potapov, J. Ju, S. Turubanova, I. Kommareddy, and T. R. Loveland (2011). "Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project". *Remote Sensing Letters* 2.4, 279–288. DOI: 10.1080/01431161.2010.519002.

Hansen, M. C., P. V. Potapov, A. H. Pickens, A. Tyukavina, A. Hernandez-Serna, V. Zalles, S. Turubanova, I. Kommareddy, S. V. Stehman, X.-P. Song, and A. Kommareddy (2022). "Global land use extent and dispersion within natural land cover using Landsat data". *Environmental Research Letters* 17.3, 034050. DOI: 10.1088/1748-9326/ac46ec.

Hengl, T., J. M. d. Jesus, G. B. M. Heuvelink, M. R. Gonzalez, M. Kilibarda, A. Blagotić, W. Shangguan, M. N. Wright, X. Geng, B. Bauer-Marschallinger, M. A. Guevara, R. Vargas, R. A. MacMillan, N. H. Batjes, J. G. B. Leenaars, E. Ribeiro, I. Wheeler, S. Mantel, and B. Kempen (2017). "SoilGrids250m: Global gridded soil information based on machine learning". *PLOS ONE* 12.2, e0169748. DOI: 10.1371/journal.pone.0169748.

Hengl, T., D. J. J. Walvoort, A. Brown, and D. G. Rossiter (2004). "A double continuous approach to visualization and analysis of categorical maps". *International Journal of Geographical Information Science* 18.2, 183–202. DOI: 10.1080/13658810310001620924.

Herold, M., L. See, N.-E. Tsendbazar, and S. Fritz (2016). "Towards an Integrated Global Land Cover Monitoring and Mapping System". *Remote Sensing* 8.12, 1036. DOI: 10.3390/rs8121036.

Hijmans, R. J., S. Phillips, J. Leathwick, and J. Elith (2017). *dismo: Species Distribution Modeling*. R package version 1.1-4. URL: https://CRAN.R-project.org/package=dismo.

Hill, S. D. and J. C. Spall (2019). "Stationarity and Convergence of the Metropolis-Hastings Algorithm: Insights into Theoretical Aspects". *IEEE Control Systems Magazine* 39.1, 56–67. DOI: 10.1109/MCS.2018.2876959.

Hobbs, S. (2003). *Linear mixture modelling solution methods for satellite remote sensing*. Report. ISBN: 9781871564839. College of Aeronautics, Cranfield University.

Hong, X., F. Huang, H. Zhang, and P. Wang (2022). "Characterizing the Turning Points in Ecosystem Functioning and Their Linkages to Drought and Human Activities over the Arid and Semi-Arid Regions of Northern China". *Remote Sensing* 14.21, 5396. DOI: 10.3390/rs14215396.

Hothorn, T., F. Bretz, and P. Westfall (2008). "Simultaneous Inference in General Parametric Models". *Biometrical Journal* 50.3, 346–363. DOI: 10.1002/bimj.200810425.

Huete, A., C. Justice, and W. Van Leeuwen (1999). *MODIS vegetation index (MOD13): algorithm theoretical basis document*. Version 3.

Iacono, M., D. Levinson, A. El-Geneidy, and R. Wasfi (2015). "A Markov Chain Model of Land Use Change". *TeMA - Journal of Land Use, Mobility and Environment* 8.3, 263–276. DOI: 10.6092/1970-9870/2985.

Jakubauskas, M. E., D. R. Legates, and J. H. Kastens (2001). "Harmonic analysis of time-series AVHRR NDVI data". *Photogrammetric Engineering & Remote Sensing* 67.4, 461–470.

Jia, K., S. Liang, X. Wei, L. Zhang, Y. Yao, and S. Gao (2014). "Automatic land-cover update approach integrating iterative training sample selection and a Markov Random Field model". *Remote Sensing Letters* 5.2, 148–156. DOI: 10.1080/2150704X.2014.889862.

Joint Research Centre (2023). "Part 2: Technical Specifications". In: *Framework Service Contract Copernicus Global Land Cover and Tropical Forest Mapping and Monitoring ('LCFM')*. Brussels: European Commission.

De Jong, R., J. Verbesselt, A. Zeileis, and M. E. Schaepman (2013). "Shifts in Global Vegetation Activity Trends". *Remote Sensing* 5.3, 1117–1133. DOI: 10.3390/rs5031117.

Karalas, K., G. Tsagkatakis, M. Zervakis, and P. Tsakalides (2016). "Land classification using remotely sensed data: Going multilabel". *IEEE Transactions on Geoscience and Remote Sensing* 54.6, 3548–3563. DOI: 10.1109/TGRS.2016.2520203.

Kasetkasem, T., M. K. Arora, and P. K. Varshney (2005). "Super-resolution land cover mapping using a Markov random field based approach". *Remote Sensing of Environment* 96.3, 302–314. DOI: 10.1016/j.rse.2005.02.006.

Kauth, R. J. and G. S. Thomas (1976). "The Tasselled Cap – A Graphic Description of the Spectral-Temporal Development of Agricultural Crops as Seen by LANDSAT". In: *Symposium on Machine Processing of Remotely Sensed Data*. Vol. 159. IEEE Catalog No. 76CH1103-1 MPRSD. Indiana: The Institute of Electrical and Electronics Engineers, Inc.

Keller, J. M., M. R. Gray, and J. A. Givens (1985). "A fuzzy K-nearest neighbor algorithm". *IEEE Transactions on Systems, Man, and Cybernetics* SMC-15.4, 580–585. DOI: 10.1109/TSMC.1985.6313426.

Kennedy, R. E., Z. Yang, and W. B. Cohen (2010). "Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. LandTrendr — Temporal segmentation algorithms". *Remote Sensing of Environment* 114.12, 2897–2910. DOI: 10.1016/j.rse.2010.07.008.

Kennedy, R. E., Z. Yang, N. Gorelick, J. Braaten, L. Cavalcante, W. B. Cohen, and S. Healey (2018). "Implementation of the LandTrendr Algorithm on Google Earth Engine". *Remote Sensing* 10.5, 691. DOI: 10.3390/rs10050691.

Key, C. H., N. Benson, D. Ohlen, S. M. Howard, and Z. Zhu (2002). "The normalized burn ratio and relationships to burn severity: ecology, remote sensing and implementation". In: *Ninth biennial remote sensing applications conference, Apr 8–12, San Diego, CA*.

Kriegler, F. J., W. A. Malila, R. F. Nalepka, and W. Richardson (1969). "Preprocessing Transformations and Their Effects on Multispectral Recognition". In: *Sixth International Symposium on Remote Sensing of Environment*. Vol. II. Michigan: Environmental Research Institute of Michigan, 97.

Kuang, W., Y. Hou, Y. Dou, D. Lu, and S. Yang (2021). "Mapping Global Urban Impervious Surface and Green Space Fractions Using Google Earth Engine". *Remote Sensing* 13.20, 4187. DOI: 10.3390/rs13204187.

Kuhn, M. and R. Quinlan (2020). *Cubist: Rule- And Instance-Based Regression Modeling*. R package version 0.2.3. URL: https://CRAN.R-project.org/package=Cubist.

Van der Laan, M. J., E. C. Polley, and A. E. Hubbard (2007). "Super Learner". *Statistical Applications in Genetics and Molecular Biology* 6.1. DOI: 10.2202/1544-6115.1309.

Li, C., P. Gong, J. Wang, Z. Zhu, G. S. Biging, C. Yuan, T. Hu, H. Zhang, Q. Wang, X. Li, X. Liu, Y. Xu, J. Guo, C. Liu, K. O. Hackman, M. Zhang, Y. Cheng, L. Yu, J. Yang, H. Huang, and N. Clinton (2017). "The first all-season sample set for mapping global land cover with Landsat-8 data". *Science Bulletin* 62.7, 508–515. DOI: 10.1016/j.scib.2017.03.011.

Li, L., A. Skidmore, A. Vrieling, and T. Wang (2019). "A new dense 18-year time series of surface water fraction estimates from MODIS for the Mediterranean region". *Hydrology and Earth System Sciences* 23.7, 3037–3056. DOI: 10.5194/hess-23-3037-2019.

Li, L., A. Vrieling, A. Skidmore, T. Wang, and E. Turak (2018). "Monitoring the dynamics of surface water fraction from MODIS time series in a Mediterranean environment". *International Journal of Applied Earth Observation and Geoinformation* 66, 135–145. DOI: 10.1016/j.jag.2017.11.007.

Li, M., S. Zang, B. Zhang, S. Li, and C. Wu (2014). "A Review of Remote Sensing Image Classification Techniques: the Role of Spatio-contextual Information". *European Journal of Remote Sensing* 47.1, 389–411. DOI: 10.5721/EuJRS20144723.

Liu, C., W. Song, C. Lu, and J. Xia (2021a). "Spatial-Temporal Hidden Markov Model for Land Cover Classification Using Multitemporal Satellite Images". *IEEE Access* 9, 76493–76502. DOI: 10.1109/ACCESS.2021.3080926.

Liu, D., K. Jia, H. Jiang, M. Xia, G. Tao, B. Wang, Z. Chen, B. Yuan, and J. Li (2021b). "Fractional Vegetation Cover Estimation Algorithm for FY-3B Reflectance Data Based on Random Forest Regression Method". *Remote Sensing* 13.11, 2165. DOI: 10.3390/rs13112165.

Liu, J., S. Wu, and J. V. Zidek (1997). "On segmented multivariate regression". *Statistica Sinica* 7.2, 497–525.

Liu, Y., R. Liu, and R. Shang (2022). "GLOBMAP SWF: a global annual surface water cover frequency dataset during 2000–2020". *Earth System Science Data* 14.10, 4505–4523. DOI: 10.5194/essd-14-4505-2022.

Lizarazo, I. (2012). "Quantitative land cover change analysis using fuzzy segmentation". *International Journal of Applied Earth Observation and Geoinformation*. Special Issue on Geographic Object-based Image Analysis: GEOBIA 15, 16–27. DOI: 10.1016/j.jag.2011.05.012.

Ludwig, M., J. Bahlmann, E. Pebesma, and H. Meyer (2022). "Developing transferable spatial prediction models: a case study of satellite based landcover mapping". *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 43, 135–141.

Ma, X., M. D. Mahecha, M. Migliavacca, F. van der Plas, R. Benavides, S. Ratcliffe, J. Kattge, R. Richter, T. Musavi, L. Baeten, I. Barnoaiea, F. J. Bohn, O. Bouriaud, F. Bussotti, A. Coppi, T. Domisch, A. Huth, B. Jaroszewicz, J. Joswig, D. E. Pabon-Moreno, D. Papale, F. Selvi, G. V. Laurin, F. Valladares, M. Reichstein, and C. Wirth (2019). "Inferring plant functional diversity from space: the potential of Sentinel-2". *Remote Sensing of Environment* 233, 111368. DOI: 10.1016/j.rse.2019.111368.

Magnússon, R. Í., J. Limpens, D. Kleijn, K. van Huissteden, T. C. Maximov, S. Lobry, and M. M. P. D. Heijmans (2021). "Shrub decline and expansion of wetland vegetation revealed by very high resolution land cover change detection in the Siberian lowland tundra". *Science of The Total Environment* 782, 146877. DOI: 10.1016/j.scitotenv.2021.146877.

Mahmood, R., R. A. Pielke Sr., K. G. Hubbard, D. Niyogi, P. A. Dirmeyer, C. McAlpine, A. M. Carleton, R. Hale, S. Gameda, A. Beltrán-Przekurat, B. Baker, R. McNider, D. R. Legates, M. Shepherd, J. Du, P. D. Blanken, O. W. Frauenfeld, U. Nair, and S. Fall (2014). "Land cover changes and their biogeophysical effects on climate". *International Journal of Climatology* 34.4, 929–953. DOI: 10.1002/joc.3736.

Makridakis, S. G., S. C. Wheelwright, and R. J. Hyndman (1997). *Forecasting: Methods and Applications*. 3rd ed. New York: Wiley.

Malinowski, R., S. Lewiński, M. Rybicki, E. Gromny, M. Jenerowicz, M. Krupiński, A. Nowakowski, C. Wojtkowski, M. Krupiński, E. Krätzschmar, and P. Schauer (2020). "Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery". *Remote Sensing* 12.21, 3523. DOI: 10.3390/rs12213523.

Markov, A. (1907). "Extension of the Limit Theorems of Probability Theory to a Sum of Variables Connected in a Chain". In: *The Notes of the Imperial Academy of Sciences of St. Petersburg VIII Series*. Vol. XXII. St. Petersburg: Physio-Mathematical College, 9.

Masek, J., E. Vermote, N. Saleous, R. Wolfe, F. Hall, K. Huemmrich, F. Gao, J. Kutler, and T.-K. Lim (2006). "A Landsat surface reflectance dataset for North America, 1990-2000".

IEEE Geoscience and Remote Sensing Letters 3.1, 68–72. DOI: 10.1109/LGRS.2005.857030.

Masiliūnas, D. (2020). *Code for global land cover fraction mapping on PROBA-V MEP*. Version: v1.1. Zenodo. DOI: 10.5281/zenodo.4281202.

– (2023). *Code for exploring postprocessing effects on high frequency global land cover maps for change assessment*. DOI: 10.5281/zenodo.10409915.

Masiliūnas, D., R. Burger, N.-E. Tsendbazar, and D. Marcos (2024). "Improving global land cover fraction change detection using a Markov chain model". Under review.

Masiliūnas, D., N.-E. Tsendbazar, M. Herold, M. Lesiv, M. Buchhorn, and J. Verbesselt (2021a). "Global land characterisation using land cover fractions at 100 m resolution". *Remote Sensing of Environment* 259, 112409. DOI: 10.1016/j.rse.2021.112409.

Masiliūnas, D., N.-E. Tsendbazar, M. Herold, D. Marcos, and J. Verbesselt (2023). "Post-processing high frequency global land cover maps for change assessment". Under revision.

Masiliūnas, D., N.-E. Tsendbazar, M. Herold, and J. Verbesselt (2021b). "BFAST Lite: A Lightweight Break Detection Method for Time Series Analysis". *Remote Sensing* 13.16, 3308. DOI: 10.3390/rs13163308.

Masiliūnas, D., J. Verbesselt, A. Zeileis, and M. Appel (2021c). *bfast*. Version 1.6.1. DOI: 10.5281/zenodo.5018513.

Masolele, R. N., V. De Sy, M. Herold, D. Marcos, J. Verbesselt, F. Gieseke, A. G. Mullissa, and C. Martius (2021). "Spatial and temporal deep learning methods for deriving land-use following deforestation: A pan-tropical case study using Landsat time series". *Remote Sensing of Environment* 264, 112600. DOI: 10.1016/j.rse.2021.112600.

Masolele, R. N., V. De Sy, D. Marcos, J. Verbesselt, F. Gieseke, K. A. Mulatu, Y. Moges, H. Sebrala, C. Martius, and M. Herold (2022). "Using high-resolution imagery and deep learning to classify land-use following deforestation: a case study in Ethiopia". *GIScience & Remote Sensing* 59.1, 1446–1472. DOI: 10.1080/15481603.2022.2115619.

Mersmann, O. (2019). *microbenchmark: Accurate Timing Functions*. R package version 1.4-7. URL: https://CRAN.R-project.org/package=microbenchmark.

Mevik, B.-H., R. Wehrens, and K. H. Liland (2016). *pls: Partial Least Squares and Principal Component Regression*. R package version 2.6-0. URL: https://CRAN.R-project.org/package=pls.

Meyer, H. and E. Pebesma (2021). "Predicting into unknown space? Estimating the area of applicability of spatial prediction models". *Methods in Ecology and Evolution* 12.9, 1620–1633. DOI: 10.1111/2041-210X.13650.

Mîrț, A., J. Reiche, J. Verbesselt, and M. Herold (2022). "A Downsampling Method Addressing the Modifiable Areal Unit Problem in Remote Sensing". *Remote Sensing* 14.21, 5538. DOI: 10.3390/rs14215538.

Montesano, P. M., R. Nelson, G. Sun, H. Margolis, A. Kerber, and K. J. Ranson (2009). "MODIS tree cover validation for the circumpolar taiga–tundra transition zone". *Remote Sensing of Environment* 113.10, 2130–2141. DOI: 10.1016/j.rse.2009.05.021.

Murillo-Sandoval, P. J., T. Hilker, M. A. Krawchuk, and J. van den Hoek (2018). "Detecting and Attributing Drivers of Forest Disturbance in the Colombian Andes Using Landsat Time-Series". *Forests* 9.5, 269. DOI: 10.3390/f9050269.

Mutanga, O., E. Adam, and M. A. Cho (2012). "High density biomass estimation for wetland vegetation using WorldView-2 imagery and random forest regression algorithm". *International Journal of Applied Earth Observation and Geoinformation* 18, 399–406. DOI: 10.1016/j.jag.2012.03.012.

NASA, METI, AIST, Japan Spacesystems, and U.S./Japan ASTER Science Team (2019). *ASTER Global Digital Elevation Model V003*. NASA EOSDIS Land Processes DAAC. DOI: 10.5067/ASTER/ASTGTM.003.

Nash, J. E. and J. V. Sutcliffe (1970). "River flow forecasting through conceptual models part I — A discussion of principles". *Journal of Hydrology* 10.3, 282–290. DOI: 10.1016/0022-1694(70)90255-6.

Nativi, S., P. Mazzetti, and M. Craglia (2021). "Digital Ecosystems for Developing Digital Twins of the Earth: The Destination Earth Case". *Remote Sensing* 13.11, 2119. DOI: 10.3390/rs13112119.

Neter, J., M. Kutner, and C. Nachtsheim (1996). *Applied linear statistical models*. 4th ed. McGraw-Hill/Irwin.

Ngadi Scarpetta, Y., V. Lebourgeois, A.-E. Laques, M. Dieye, J. Bourgoin, and A. Bégué (2023). "BFASTm-L2, an unsupervised LULCC detection based on seasonal change detection – An application to large-scale land acquisitions in Senegal". *International Journal of Applied Earth Observation and Geoinformation* 121, 103379. DOI: 10.1016/j.jag.2023.103379.

Okeke, F. and A. Karnieli (2006). "Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm". *Ecological Informatics* 1.1, 117–124. DOI: 10.1016/j.ecoinf.2005.10.006.

Okujeni, A., F. Canters, S. D. Cooper, J. Degerickx, U. Heiden, P. Hostert, F. Priem, D. A. Roberts, B. Somers, and S. van der Linden (2018). "Generalizing machine learning regression models using multi-site spectral libraries for mapping vegetation-impervious-soil fractions across multiple cities". *Remote Sensing of Environment* 216, 482–496. DOI: 10.1016/j.rse.2018.07.011.

Van Oort, P. A. J. (2005). "Improving land cover change estimates by accounting for classification errors". *International Journal of Remote Sensing* 26.14, 3009–3024. DOI: 10.1080/01431160500057848.

Padilla, M., S. V. Stehman, J. Litago, and E. Chuvieco (2014). "Assessing the Temporal Stability of the Accuracy of a Time Series of Burned Area Products". *Remote Sensing* 6.3, 2050–2068. DOI: 10.3390/rs6032050.

Parente, L., V. Mesquita, F. Miziara, L. Baumann, and L. Ferreira (2019a). "Assessing the pasturelands and livestock dynamics in Brazil, from 1985 to 2017: A novel approach based on high spatial resolution imagery and Google Earth Engine cloud computing". *Remote Sensing of Environment* 232, 111301. DOI: 10.1016/j.rse.2019.111301.

Parente, L., E. Taquary, A. P. Silva, C. Souza, and L. Ferreira (2019b). "Next Generation Mapping: Combining Deep Learning, Cloud Computing, and Big Remote Sensing Data". *Remote Sensing* 11.23, 2881. DOI: 10.3390/rs11232881.

Pearson, K. (1895). "Notes on Regression and Inheritance in the Case of Two Parents". In: *Proceedings of the Royal Society of London*. Vol. 58, 240–242.

Pebesma, E., P. Griffiths, C. Briese, A. Jacob, A. Skerlevaj, J. Dries, G. Camara, and M. Mohr (2021). "Analyzing large-scale Earth Observation data repositories made simple with OpenEO Platform". In: EGU21-9602. Conference Name: EGU21. DOI: 10.5194/egusphere-egu21-9602.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward (2016). "High-resolution mapping of global surface water and its long-term changes". *Nature* 540.7633, 418–422. DOI: 10.1038/nature20584.

Pengra, B., J. Long, D. Dahal, S. V. Stehman, and T. R. Loveland (2015). "A global reference database from very high resolution commercial satellite data and methodology for application to Landsat derived 30m continuous field tree cover data". *Remote Sensing of Environment* 165, 234–248. DOI: 10.1016/j.rse.2015.01.018.

Pereira, H. M., S. Ferrier, M. Walters, G. N. Geller, R. H. G. Jongman, R. J. Scholes, M. W. Bruford, N. Brummitt, S. H. M. Butchart, A. C. Cardoso, N. C. Coops, E. Dulloo, D. P. Faith, J. Freyhof, R. D. Gregory, C. Heip, R. Höft, G. Hurtt, W. Jetz, D. S. Karp, M. A. McGeoch, D. Obura, Y. Onoda, N. Pettorelli, B. Reyers, R. Sayre, J. P. W. Scharlemann, S. N. Stuart, E. Turak, M. Walpole, and M. Wegmann (2013). "Essential Biodiversity Variables". *Science* 339.6117. Publisher: American Association for the Advancement of Science, 277–278. DOI: 10.1126/science.1229931.

Pielke Sr., R. A., A. Pitman, D. Niyogi, R. Mahmood, C. McAlpine, F. Hossain, K. K. Goldewijk, U. Nair, R. Betts, S. Fall, M. Reichstein, P. Kabat, and N. de Noblet (2011). "Land use/land cover changes and climate: modeling analysis and observational evidence". *WIREs Climate Change* 2.6, 828–850. DOI: 10.1002/wcc.144.

Potapov, P., M. C. Hansen, A. Pickens, A. Hernandez-Serna, A. Tyukavina, S. Turubanova, V. Zalles, X. Li, A. Khan, F. Stolle, N. Harris, X.-P. Song, A. Baggett, I. Kommareddy, and A. Kommareddy (2022). "The Global 2000-2020 Land Cover and Land Use Change Dataset Derived From the Landsat Archive: First Results". *Frontiers in Remote Sensing* 3, 856903. DOI: 10.3389/frsen.2022.856903.

Quinlan, J. R. (1992). "Learning with continuous classes". In: *AI'92*. Singapore: World Scientific. DOI: 10.1142/9789814536271.

R Core Team (2021). *A language and environment for statistical computing.*

Reiche, J., A. Mullissa, B. Slagter, Y. Gou, N.-E. Tsendbazar, C. Odongo-Braun, A. Vollrath, M. J. Weisse, F. Stolle, A. Pickens, G. Donchyts, N. Clinton, N. Gorelick, and M. Herold (2021). "Forest disturbance alerts for the Congo Basin using Sentinel-1". *Environmental Research Letters* 16.2, 024005. DOI: 10.1088/1748-9326/abd0a8.

Van Rossum, G. and F. L. Drake (2009). *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace.

Sales, M. H. R., S. de Bruin, C. Souza, and M. Herold (2022). "Land Use and Land Cover Area Estimates From Class Membership Probability of a Random Forest Classification". *IEEE Transactions on Geoscience and Remote Sensing* 60, 1–11. DOI: 10.1109/TGRS.2021.3080083.

Saunier, S., B. Pflug, I. M. Lobos, B. Franch, J. Louis, R. de los Reyes, V. Debaecker, E. G. Cadau, V. Boccia, F. Gascon, and S. Kocaman (2022). "Sen2Like: Paving the Way towards Harmonization and Fusion of Optical Data". *Remote Sensing* 14.16, 3855. DOI: 10.3390/rs14163855.

Savitzky, A. and M. J. E. Golay (1964). "Smoothing and Differentiation of Data by Simplified Least Squares Procedures." *Analytical Chemistry* 36.8, 1627–1639. DOI: 10.1021/ac60214a047.

Schramm, M., E. Pebesma, M. Milenković, L. Foresta, J. Dries, A. Jacob, W. Wagner, M. Mohr, M. Neteler, M. Kadunc, T. Miksa, P. Kempeneers, J. Verbesselt, B. Gößwein, C. Navacchi, S. Lippens, and J. Reiche (2021). "The openEO API–Harmonising the Use of Earth Observation Cloud Services Using Virtual Data Cube Functionalities". *Remote Sensing* 13.6, 1125. DOI: 10.3390/rs13061125.

Schrodt, F., J. J. Bailey, W. D. Kissling, K. F. Rijsdijk, A. C. Seijmonsbergen, D. van Ree, J. Hjort, R. S. Lawley, C. N. Williams, M. G. Anderson, P. Beier, P. van Beukering, D. S. Boyd, J. Brilha, L. Carcavilla, K. M. Dahlin, J. C. Gill, J. E. Gordon, M. Gray, M. Grundy, M. L. Hunter, J. J. Lawler, M. Monge-Ganuzas, K. R. Royse, I. Stewart, S. Record, W. Turner, P. L. Zarnetske, and R. Field (2019). "To advance sustainable stewardship, we must document not only biodiversity but geodiversity". *Proceedings of the National Academy of Sciences* 116.33, 16155–16158. DOI: 10.1073/pnas.1911799116.

Schroeder, R., K. C. McDonald, B. D. Chapman, K. Jensen, E. Podest, Z. D. Tessler, T. J. Bohn, and R. Zimmermann (2015). "Development and Evaluation of a Multi-Year Fractional Surface Water Data Set Derived from Active/Passive Microwave Remote Sensing Data". *Remote Sensing* 7.12, 16688–16732. DOI: 10.3390/rs71215843.

See, L., D. Schepaschenko, M. Lesiv, I. McCallum, S. Fritz, A. Comber, C. Perger, C. Schill, Y. Zhao, V. Maus, M. A. Siraj, F. Albrecht, A. Cipriani, M. Vakolyuk, A. Garcia, A. H. Rabia, K. Singha, A. A. Marcarini, T. Kattenborn, R. Hazarika, M. Schepaschenko,

M. van der Velde, F. Kraxner, and M. Obersteiner (2015). "Building a hybrid land cover map with crowdsourcing and geographically weighted regression". *ISPRS Journal of Photogrammetry and Remote Sensing*. Global Land Cover Mapping and Monitoring 103, 48–56. DOI: 10.1016/j.isprsjprs.2014.06.016.

Seo, B., C. Bogner, T. Koellner, and B. Reineking (2016). "Mapping Fractional Land Use and Land Cover in a Monsoon Region: The Effects of Data Processing Options". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9.9, 3941–3956. DOI: 10.1109/JSTARS.2016.2544802.

Seto, K. C., B. Güneralp, and L. R. Hutyra (2012). "Global forecasts of urban expansion to 2030 and direct impacts on biodiversity and carbon pools". *Proceedings of the National Academy of Sciences* 109.40, 16083–16088. DOI: 10.1073/pnas.1211658109.

Settles, B. (2009). *Active Learning Literature Survey*. Tech. rep. 1648. University of Wisconsin–Madison.

Sexton, J. O., X.-P. Song, M. Feng, P. Noojipady, A. Anand, C. Huang, D.-H. Kim, K. M. Collins, S. Channan, C. DiMiceli, and J. R. Townshend (2013). "Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error". *International Journal of Digital Earth* 6.5, 427–448. DOI: 10.1080/17538947.2013.786146.

Shao, Q. and N. A. Campbell (2002). "Applications: Modelling trends in groundwater levels by segmented regression with constraints". *Australian & New Zealand Journal of Statistics* 44.2, 129–141. DOI: 10.1111/1467-842X.00216.

Sharma, C. S., M. D. Behera, A. Mishra, and S. N. Panda (2011). "Assessing Flood Induced Land-Cover Changes Using Remote Sensing and Fuzzy Approach in Eastern Gujarat (India)". *Water Resources Management* 25.13, 3219. DOI: 10.1007/s11269-011-9853-7.

Sharma, R., U. Nehren, S. A. Rahman, M. Meyer, B. Rimal, G. Aria Seta, and H. Baral (2018). "Modeling Land Use and Land Cover Changes and Their Effects on Biodiversity in Central Kalimantan, Indonesia". *Land* 7.2, 57. DOI: 10.3390/land7020057.

Shen, X., C.-Q. Ke, Y. Fan, and L. Drolma (2022). "A new digital elevation model (DEM) dataset of the entire Antarctic continent derived from ICESat-2". *Earth System Science Data* 14.7, 3075–3089. DOI: 10.5194/essd-14-3075-2022.

Shimabukuro, Y. and J. Smith (1991). "The least-squares mixing models to generate fraction images derived from remote sensing multispectral data". *IEEE Transactions on Geoscience and Remote Sensing* 29.1, 16–20. DOI: 10.1109/36.103288.

Shimizu, K. and H. Saito (2021). "Country-wide mapping of harvest areas and post-harvest forest recovery using Landsat time series data in Japan". *International Journal of Applied Earth Observation and Geoinformation* 104, 102555. DOI: 10.1016/j.jag.2021.102555.

Silván-Cárdenas, J. L. and L. Wang (2008). "Sub-pixel confusion–uncertainty matrix for assessing soft classifications". *Remote Sensing of Environment* 112.3, 1081–1095. DOI: 10.1016/j.rse.2007.07.017.

Smets, B., N. Souverijns, G. Jattrain, M. Buchhorn, A. Moiret, A. V. Quang, M. Lesiv, and N.-E. Tsendbazar (2020). "Elastic Mapping Through the Copernicus Global Land Cover Layers". In: *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 4251–4254. DOI: 10.1109/IGARSS39084.2020.9324013.

Solberg, A., T. Taxt, and A. Jain (1996). "A Markov random field model for classification of multisource satellite imagery". *IEEE Transactions on Geoscience and Remote Sensing* 34.1, 100–113. DOI: 10.1109/36.481897.

Somers, B., G. P. Asner, L. Tits, and P. Coppin (2011). "Endmember variability in Spectral Mixture Analysis: A review". *Remote Sensing of Environment* 115.7, 1603–1616. DOI: 10.1016/j.rse.2011.03.003.

Song, X.-P., M. C. Hansen, S. V. Stehman, P. V. Potapov, A. Tyukavina, E. F. Vermote, and J. R. Townshend (2018). "Global land change from 1982 to 2016". *Nature* 560.7720, 639–643. DOI: 10.1038/s41586-018-0411-9.

Souverijns, N., M. Buchhorn, S. Horion, R. Fensholt, H. Verbeeck, J. Verbesselt, M. Herold, N.-E. Tsendbazar, P. N. Bernardino, B. Somers, and R. van de Kerchove (2020). "Thirty Years of Land Cover and Fraction Cover Changes over the Sudano-Sahel Using Landsat Time Series". *Remote Sensing* 12.22, 3817. DOI: 10.3390/rs12223817.

Souza, C. M., J. Z. Shimbo, M. R. Rosa, L. L. Parente, A. A. Alencar, B. F. T. Rudorff, H. Hasenack, M. Matsumoto, L. G. Ferreira, P. W. M. Souza-Filho, S. W. de Oliveira, W. F. Rocha, A. V. Fonseca, C. B. Marques, C. G. Diniz, D. Costa, D. Monteiro, E. R. Rosa, E. Vélez-Martin, E. J. Weber, F. E. B. Lenti, F. F. Paternost, F. G. C. Pareyn, J. V. Siqueira, J. L. Viera, L. C. F. Neto, M. M. Saraiva, M. H. Sales, M. P. G. Salgado, R. Vasconcelos, S. Galano, V. V. Mesquita, and T. Azevedo (2020). "Reconstructing Three Decades of Land Use and Land Cover Changes in Brazilian Biomes with Landsat Archive and Earth Engine". *Remote Sensing* 12.17, 2735. DOI: 10.3390/rs12172735.

Spearman, C. (1904). "Rank's correlation". *American Journal of Psychology* 15.88, 17–31.

Stavrakoudis, D. G., J. B. Theocharis, and G. C. Zalidis (2011). "A Boosted Genetic Fuzzy Classifier for land cover classification of remote sensing imagery". *ISPRS Journal of Photogrammetry and Remote Sensing* 66.4, 529–544. DOI: 10.1016/j.isprsjprs.2011.01.010.

Steinwart, I. and P. Thomann (2017). *liquidSVM: A Fast and Versatile SVM package*. arXiv: 1702.06899 [stat.ML].

Sulla-Menashe, D., J. M. Gray, S. P. Abercrombie, and M. A. Friedl (2019). "Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product". *Remote Sensing of Environment* 222, 183–194. DOI: 10.1016/j.rse.2018.12.013.

Suykens, J. and J. Vandewalle (1999). "Least Squares Support Vector Machine Classifiers". *Neural Processing Letters* 9.3, 293–300. DOI: 10.1023/A:1018628609742.

Taylor, S. J. and B. Letham (2018). "Forecasting at Scale". *The American Statistician* 72.1, 37–45. DOI: 10.1080/00031305.2017.1380080.

Theil, H. (1969). "A Multinomial Extension of the Linear Logit Model". *International Economic Review* 10.3, 251–259. DOI: 10.2307/2525642.

Tibshirani, R. (1996). "Regression Shrinkage and Selection Via the Lasso". *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1, 267–288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

Townshend, J. R. G. (1999). *MODIS Enhanced Land Cover and Land Cover Change Product Algorithm Theoretical Basis Documents (ATBD) Version 2.0*. Tech. rep. Maryland: University of Maryland.

– (2017). *Global Forest Cover Change (GFCC) Tree Cover Multi-Year Global 30 m V003*. DOI: 10.5067/MEASURES/GFCC/GFCC30TC.003.

Van Tricht, K., J. Degerickx, S. Gilliams, D. Zanaga, M. Savinaud, M. Battude, R. Buguet de Chargère, G. Dubreule, A. Grosu, J. Brombacher, H. Pelgrum, M. Lesiv, J. C. L. Bayas, S. Karanam, S. Fritz, I. Becker-Reshef, B. Franch, B. M. Bononad, J. Cintas, H. Boogaard, A. K. Pratihast, L. Kucera, and Z. Szantoi (2023). *ESA WorldCereal 10 m 2021 v100*. DOI: 10.5281/zenodo.7875105.

Trivedi, M. M. and J. C. Bezdek (1986). "Low-Level Segmentation of Aerial Images with Fuzzy Clustering". *IEEE Transactions on Systems, Man, and Cybernetics* 16.4, 589–598. DOI: 10.1109/TSMC.1986.289264.

Tsendbazar, N.-E., M. Herold, S. de Bruin, M. Lesiv, S. Fritz, R. van de Kerchove, M. Buchhorn, M. Duerauer, Z. Szantoi, and J.-F. Pekel (2018). "Developing and applying a multi-purpose land cover validation dataset for Africa". *Remote Sensing of Environment* 219, 298–309. DOI: 10.1016/j.rse.2018.10.025.

Tsendbazar, N.-E., M. Herold, L. Li, A. Tarko, S. de Bruin, D. Masiliūnas, M. Lesiv, S. Fritz, M. Buchhorn, B. Smets, R. van de Kerchove, and M. Duerauer (2021a). "Towards operational validation of annual global land cover maps". *Remote Sensing of Environment* 266, 112686. DOI: 10.1016/j.rse.2021.112686.

Tsendbazar, N.-E., M. Herold, M. Lesiv, and S. Fritz (2019). *Copernicus Global Land Service: Land Cover 100m: version 2 Globe 2015: Validation Report*. Tech. rep. Copernicus Global Land Service. DOI: 10.5281/zenodo.3606438.

Tsendbazar, N.-E., A. Tarko, L. Li, M. Herold, M. Lesiv, S. Fritz, and V. Maus (2021b). *Copernicus Global Land Service: Land Cover 100m: version 3 Globe 2015-2019: Validation Report*. Tech. rep. Zenodo. DOI: 10.5281/zenodo.4723975.

Tuia, D., K. Schindler, B. Demir, G. Camps-Valls, X. X. Zhu, M. Kochupillai, S. Džeroski, J. N. van Rijn, H. H. Hoos, F. Del Frate, M. Datcu, J.-A. Quiané-Ruiz, V. Markl, B. L.

Saux, and R. Schneider (2023). *Artificial intelligence to advance Earth observation: a perspective*. arXiv:2305.08413 [cs, eess, stat].

Tyukavina, A., M. C. Hansen, P. Potapov, D. Parker, C. Okpa, S. V. Stehman, I. Kommareddy, and S. Turubanova (2018). "Congo Basin forest loss dominated by increasing smallholder clearing". *Science Advances* 4.11, eaat2993. DOI: 10.1126/sciadv.aat2993.

U.S. Geological Survey (2023). *Landsat 8-9 Collection 2 Level 2 Science Product Guide*. Version 5.0.

Uma Shankar, B., S. K. Meher, and A. Ghosh (2011). "Wavelet-fuzzy hybridization: Feature-extraction and land-cover classification of remote sensing images". *Applied Soft Computing* 11.3, 2999–3011. DOI: 10.1016/j.asoc.2010.11.024.

Venables, W. N. and B. D. Ripley (2002). *Modern Applied Statistics with S*. Fourth edition. ISBN 0-387-95457-0. New York: Springer.

Verbesselt, J., R. Hyndman, G. Newnham, and D. Culvenor (2010a). "Detecting trend and seasonal changes in satellite image time series". *Remote Sensing of Environment* 114.1, 106–115. DOI: 10.1016/j.rse.2009.08.014.

Verbesselt, J., R. Hyndman, A. Zeileis, and D. Culvenor (2010b). "Phenological change detection while accounting for abrupt and gradual trends in satellite image time series". *Remote Sensing of Environment* 114.12, 2970–2980. DOI: 10.1016/j.rse.2010.08.003.

Verbesselt, J., A. Zeileis, and M. Herold (2012). "Near real-time disturbance detection using satellite image time series". *Remote Sensing of Environment* 123, 98–108. DOI: 10.1016/j.rse.2012.02.022.

Verger, A., J. Sánchez-Zapero, M. Weiss, A. Descals, F. Camacho, R. Lacaze, and F. Baret (2023). "GEOV2: Improved smoothed and gap filled time series of LAI, FAPAR and FCover 1 km Copernicus Global Land products". *International Journal of Applied Earth Observation and Geoinformation* 123, 103479. DOI: 10.1016/j.jag.2023.103479.

Vilà-Guerau de Arellano, J., P. Ney, O. Hartogensis, H. de Boer, K. van Diepen, D. Emin, G. de Groot, A. Klosterhalfen, M. Langensiepen, M. Matveeva, G. Miranda-García, A. F. Moene, U. Rascher, T. Röckmann, G. Adnew, N. Brüggemann, Y. Rothfuss, and A. Graf (2020). "CloudRoots: integration of advanced instrumental techniques and process modelling of sub-hourly and sub-kilometre land–atmosphere interactions". *Biogeosciences* 17.17, 4375–4404. DOI: 10.5194/bg-17-4375-2020.

Walton, J. T. (2008). "Subpixel Urban Land Cover Estimation: Comparing Cubist, Random Forests and Support Vector Regression". *Photogrammetric Engineering & Remote Sensing* 74.10, 1213–1222. DOI: 10.14358/PERS.74.10.1213.

Wang, C., J. Qi, and M. Cochrane (2005). "Assessment of Tropical Forest Degradation with Canopy Fractional Cover from Landsat ETM+ and IKONOS Imagery". *Earth Interactions* 9.22, 1–18. DOI: 10.1175/EI133.1.

Wang, J., Y. Zhao, C. Li, L. Yu, D. Liu, and P. Gong (2015). "Mapping global land cover in 2001 and 2010 with spatial-temporal consistency at 250m resolution". *ISPRS Journal of Photogrammetry and Remote Sensing*. Global Land Cover Mapping and Monitoring 103, 38–47. DOI: 10.1016/j.isprsjprs.2014.03.007.

Wang, L., X. Huang, C. Zheng, and Y. Zhang (2017). "A Markov random field integrating spectral dissimilarity and class co-occurrence dependency for remote sensing image classification optimization". *ISPRS Journal of Photogrammetry and Remote Sensing* 128, 223–239. DOI: 10.1016/j.isprsjprs.2017.03.020.

Wang, N., J. G. P. W. Clevers, S. Wieneke, H. Bartholomeus, and L. Kooistra (2022a). "Potential of UAV-based sun-induced chlorophyll fluorescence to detect water stress in sugar beet". *Agricultural and Forest Meteorology* 323, 109033. DOI: 10.1016/j.agrformet.2022.109033.

Wang, N., B. Siegmann, U. Rascher, J. G. Clevers, O. Muller, H. Bartholomeus, J. Bendig, D. Masiliūnas, R. Pude, and L. Kooistra (2022b). "Comparison of a UAV-and an airborne-based system to acquire far-red sun-induced chlorophyll fluorescence measurements over structurally different crops". *Agricultural and Forest Meteorology* 323, 109081.

Wang, N., J. Suomalainen, H. Bartholomeus, L. Kooistra, D. Masiliūnas, and J. G. Clevers (2021). "Diurnal variation of sun-induced chlorophyll fluorescence of agricultural crops observed from a point-based spectrometer on a UAV". *International Journal of Applied Earth Observation and Geoinformation* 96, 102276.

Watts, L. M. and S. W. Laffan (2014). "Effectiveness of the BFAST algorithm for detecting vegetation response patterns in a semi-arid region". *Remote Sensing of Environment* 154, 234–245. DOI: 10.1016/j.rse.2014.08.023.

Wehmann, A. and D. Liu (2015). "A spatial–temporal contextual Markovian kernel method for multi-temporal land cover mapping". *ISPRS Journal of Photogrammetry and Remote Sensing* 107, 77–89. DOI: 10.1016/j.isprsjprs.2015.04.009.

Wold, S., M. Sjöström, and L. Eriksson (2001). "PLS-regression: a basic tool of chemometrics". *Chemometrics and Intelligent Laboratory Systems* 58.2, 109–130. DOI: 10.1016/S0169-7439(01)00155-1.

Wolters, E., W. Dierckx, M.-D. Iordache, and E. Swinnen (2016). *PROBA-V Products User Manual v2.0*. Version v2.0. VITO N.V. 98 pp.

Wouters, H., I. Y. Petrova, C. C. van Heerwaarden, J. Vilà-Guerau de Arellano, A. J. Teuling, V. Meulenberg, J. A. Santanello, and D. G. Miralles (2019). "Atmospheric boundary layer dynamics from balloon soundings worldwide: CLASS4GL v1.0". *Geoscientific Model Development* 12.5, 2139–2153. DOI: 10.5194/gmd-12-2139-2019.

Wright, M. N. and A. Ziegler (2017). "ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R". *Journal of Statistical Software* 77, 1–17. DOI: 10.18637/jss.v077.i01.

Xie, S., L. Liu, X. Zhang, and J. Yang (2022). "Mapping the annual dynamics of land cover in Beijing from 2001 to 2020 using Landsat dense time series stack". *ISPRS Journal of Photogrammetry and Remote Sensing* 185, 201–218. DOI: 10.1016/j.isprsjprs.2022.01.014.

Xu, L., M. Herold, N.-E. Tsendbazar, D. Masiliūnas, L. Li, M. Lesiv, S. Fritz, and J. Verbesselt (2022). "Time series analysis for global land cover change monitoring: A comparison across sensors". *Remote Sensing of Environment* 271, 112905. DOI: 10.1016/j.rse.2022.112905.

Xu, X., Y. Zhong, and L. Zhang (2014). "Adaptive Subpixel Mapping Based on a Multiagent System for Remote-Sensing Imagery". *IEEE Transactions on Geoscience and Remote Sensing* 52.2, 787–804. DOI: 10.1109/TGRS.2013.2244095.

Xu, Y. and B. Huang (2014). "A Spatio-Temporal Pixel-Swapping Algorithm for Subpixel Land Cover Mapping". *IEEE Geoscience and Remote Sensing Letters* 11.2, 474–478. DOI: 10.1109/LGRS.2013.2268153.

Yan, J., L. Wang, W. Song, Y. Chen, X. Chen, and Z. Deng (2019). "A time-series classification approach based on change detection for rapid land cover mapping". *ISPRS Journal of Photogrammetry and Remote Sensing* 158, 249–262. DOI: 10.1016/j.isprsjprs.2019.10.003.

Yang, G., S. Fang, Y. Dian, and C. Bi (2016). "Improving Seasonal Land Cover Maps of Poyang Lake Area in China by Taking into Account Logical Transitions". *ISPRS International Journal of Geo-Information* 5.9, 165. DOI: 10.3390/ijgi5090165.

Yang, G., S. Fang, W. Gong, Y. Zhao, and M. Ge (2021). "Evaluating the reliability of time series land cover maps by exploiting the hidden Markov model". *Stochastic Environmental Research and Risk Assessment* 35.4, 881–892. DOI: 10.1007/s00477-020-01915-9.

Yang, J., P. J. Weisberg, and N. A. Bristow (2012). "Landsat remote sensing approaches for monitoring long-term tree cover dynamics in semi-arid woodlands: Comparison of vegetation indices and spectral mixture analysis". *Remote Sensing of Environment* 119, 62–71. DOI: 10.1016/j.rse.2011.12.004.

Ye, S., J. Rogan, Z. Zhu, and J. R. Eastman (2021). "A near-real-time approach for monitoring forest disturbance using Landsat time series: stochastic continuous change detection". *Remote Sensing of Environment* 252, 112167. DOI: 10.1016/j.rse.2020.112167.

Zanaga, D., R. van de Kerchove, W. de Keersmaecker, N. Souverijns, C. Brockmann, R. Quast, J. Wevers, A. Grosu, A. Paccini, S. Vergnaud, O. Cartus, M. Santoro, S. Fritz, I. Georgieva, M. Lesiv, S. Carter, M. Herold, L. Li, N.-E. Tsendbazar, F. Ramoino, and O. Arino (2021). *ESA WorldCover 10 m 2020 v100*. DOI: 10.5281/zenodo.5571936.

Zeileis, A., C. Kleiber, W. Krämer, and K. Hornik (2003). "Testing and dating of structural changes in practice". *Computational Statistics & Data Analysis* 44.1, 109–123. DOI: 10.1016/S0167-9473(03)00030-6.

Zemp, M. (2022). *GCOS 2022 Implementation Plan*. Geneva: World Meteorological Organization. DOI: 10.5167/UZH-224271.

Zhang, J. and G. M. Foody (2001). "Fully-fuzzy supervised classification of sub-urban land cover from remotely sensed imagery: Statistical and artificial neural network approaches". *International Journal of Remote Sensing* 22.4, 615–628. DOI: 10.1080/01431160050505883.

Zhang, Y., Y. Du, F. Ling, S. Fang, and X. Li (2014). "Example-Based Super-Resolution Land Cover Mapping Using Support Vector Regression". *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7.4, 1271–1283. DOI: 10.1109/JSTARS.2014.2305652.

Zhang, Y., M. Migliavacca, J. Penuelas, and W. Ju (2021). "Advances in hyperspectral remote sensing of vegetation traits and functions". *Remote Sensing of Environment* 252, 112121. DOI: 10.1016/j.rse.2020.112121.

Zhao, J., Y. Le, H. Liu, H. Huang, J. Wang, and P. Gong (2021). "Towards an open and synergistic framework for mapping global land cover". *PeerJ*. DOI: 10.7717/peerj.11877.

Zhao, K., M. A. Wulder, T. Hu, R. Bright, Q. Wu, H. Qin, Y. Li, E. Toman, B. Mallick, X. Zhang, and M. Brown (2019). "Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm". *Remote Sensing of Environment* 232, 111181. DOI: 10.1016/j.rse.2019.04.034.

Zhou, X. and A. del Valle (2020). "Range Based Confusion Matrix for Imbalanced Time Series Classification". In: *2020 6th Conference on Data Science and Machine Learning Applications (CDMA)*. Riyadh, Saudi Arabia: IEEE, 1–6. DOI: 10.1109/CDMA47397.2020.00006.

Zhu, Z., S. Qiu, and S. Ye (2022). "Remote sensing of land change: A multifaceted perspective". *Remote Sensing of Environment* 282, 113266. DOI: 10.1016/j.rse.2022.113266.

Zhu, Z. and C. E. Woodcock (2014). "Continuous change detection and classification of land cover using all available Landsat data". *Remote Sensing of Environment* 144, 152–171. DOI: 10.1016/j.rse.2014.01.011.

Zhu, Z., J. Zhang, Z. Yang, A. H. Aljaddani, W. B. Cohen, S. Qiu, and C. Zhou (2020). "Continuous monitoring of land disturbance based on Landsat time series". *Remote Sensing of Environment* 238, 111116. DOI: 10.1016/j.rse.2019.03.009.

Acknowledgements

"Is your PhD going to be a rat race, a steeplechase, or a smooth ride?" was a question repeatedly raised by Claudius van de Vijver and Lennart Suselbeek at the PE&RC graduate school retreats. Even though most PhD candidates around me experienced a mid-PhD dip, I kept positive and happy throughout the whole process. It was only possible for me to enjoy the smooth ride, if somewhat longer than I initially expected, thanks to the support of my supervisors, friends and family. Words are not enough to express my gratitude to everyone who has supported and encouraged me on the PhD journey throughout all these years. Nevertheless, I shall try to do so in broad strokes in this section; but much more is still left to be expressed.

First and foremost I would like to thank my promotor Martin Herold and co-promotors Jan Verbesselt and Nandika Tsendbazar for providing me with the opportunity to carry out the PhD and for supporting me every step of the way. Throughout the journey you have become not only mentors, but also friends. The care and support you have provided to me was key to my happiness throughout both my MSc and PhD journeys, and you will always be my role models for thesis supervision.

Martin, your creativity is phenomenal and a never-ending source of new ideas. Whenever I tried an idea and it did not work, you would always come up with five new ideas to try. I am lucky to have had a promotor so closely involved in both my manuscripts and projects, and it has been an honour to work with you. Even though you are now in GFZ in Potsdam, I hope we will continue to work together in the future.

Jan, I will never forget that day at the end of the Geoscripting course in 2016 when you sat with the students to have lunch, and asked me what thesis topic I would be interested in. The result was not only an MSc topic, but also this thesis, further development of your BFAST package and your Geoscripting course, and much more. I am honoured to have become a successor to several of your projects and to follow your footsteps. Your ability to obtain research grants has been legendary, and it makes perfect sense for you to have moved to Brussels to further work with the science funding organisation BELSPO. Brussels is not very far, therefore I am sure we will have many more opportunities to meet each other face to face.

Nandika, you have always been the most accessible supervisor for me, and not just by virtue of being the only one remaining in Wageningen. When I started my MSc studies, I would have never imagined that I would get invited by my supervisor for a dumpling party, during which we would happily chat together with, among other friends and colleagues, a chairholder. Not only have you improved my scientific life with your extensive knowledge of land cover mapping and validation, but also my social life, and I'm glad to have a friend (and almost neighbour!) like you.

Special thanks goes to my paranymphs, Suzy Rębisz and 萬泉興 (Quanxing Wan). We have known each other from much earlier than the start of my PhD track, going back to the MGI days. You have always been very supportive, helpful and friendly, always happy to share your knowledge and think along. We all share a sense of admiration for each other and hold each each other in high esteem, which has resulted in an enduring friendship. Your involvement in making the defence a success is essential, and I am very grateful for your willingness to spend time and energy on it. 泉興 (Quanxing), your passion for travelling and photography is a great combination and is very inspiring; I still remember how we were discussing the university-organised field trip to Iceland back in the MGI days and how enthusiastic you were for the opportunity to go to a Nordic country, with many more trips following that one. I wish you a great finish of your own PhD journey! Suzy, your kindness and idealism are second to none, and your interest in philosophy and psychology makes for excellent conversation. I will always remember the Cutting Edge Ecology course at the NIOO and the conference in Orion where you received the famous "cube microphone" to ask a question. I wish you all the best in finding a way to realise yourself and your dreams.

I would also like to thank all four of my opponents for taking the time to read and evaluate my thesis, and for their willingness to travel to Wageningen specifically for my defence. Each of you has a very interesting research profile, linking to different activities I have done over the PhD track, and I would like to work together with each one of you in the future. I also specifically thank you, the reader, for taking your time to read my thesis; I hope it is insightful for you!

I would like to wholeheartedly thank all my friends for the support I have received throughout the PhD journey. Thanks to you, I have learned a lot, not only about science but also about life. First and foremost I would like to thank my circle-zero friends 丁劍奇 (Jianqi Ding) and 王娜 (Na Wang). You have been with me from the very beginning of my PhD journey, supported me throughout, enabled my spiritual growth and made my life a lot more colourful. I wish you all the best in the far-away lands that you are in now and will be in the future. I'd like to thank my friends and colleagues who attended the (sometimes ad-hoc) 中餐會 (lunch gatherings) for the company and keeping me up-to-date with recent developments, including 許盼盼 (Panpan Xu), 胡小璐 (Xiaolu Hu), 李琳琳 (Linlin Li), 田韞鈺 (Yunyu Tian), 諶娜 (Na Chen), 李楊 (Yang Li), 許麗麗 (Lili Xu), 張萌 (Meng Zhang), 鄒雨廷 (Yuting Zou),

Xin Li 李鑫, Shanshan Yang 楊珊珊, Yanjun Song 宋彦君 and Karimon. I thank Yi Xiong 熊苡 and Yingxia Liu 劉迎夏 for igniting my interest in Chinese culture 中華文明. Thanks to my former neighbours Jinfeng Wu 吳金鳳, Anyi Wang 王安憶, Ante Ivančić and Yibo Jiang for making Bornsesteeg corridor 11b a more social place.

I thank all my colleagues and coauthors, you have formed a great network for me and I am proud to also call you my friends. Among others, I would like to thank Kirsten, Ioannis, Devis, Lammert, Harm, Johannes Reiche, Jan, Niki, Arnold, Sytze, Ron van Lammeren, John, Arend, Alexander Klippel, Aldo, Willy, Corné, Frida, Judith, Gerlinde, Antoinette, Truus, Benjamin Brede, Mathieu, Sarah, Andrei, Sabina, Daniela, Ximena, Kalkidan, Natalia, Danaë, Arnan, Wanda, Anne, Meng Lu, Karina, Alemu, Marian, Arun, Diego, Benjamin Kellenberger, Marc, Sylvain, Agnieszka, Shivangi, Alex, Ilan, Ron van Bree, Michiel, Dilli, Aike, Hiske, Hilmy, Tianyi, Milutin, Alvaro, José, João, Peter Roosjen, Gustavo, Marston, Magdalena, Chenglong, Hasib, Jens, Umar, Yaowu, Simon, Patric, Robert, Yaqing, Adugna, Sietse, Bart, Johannes Balling, Dirk, Maciej, Samantha Martin, Jiayan, Sebastian, Jess, Braden, Lukasz, Tom Hardy, Myke, Elina, Elise, Gijs, Marcello, Erika Speelman, Federico, Lisa, Astrid, Erika Romeijn, Deborah, Anne-Juul, Maryam, Mirjam, Jascha, Paolo, Jelmar, Gildas, Yang Chen, Jalal, Gonzalo, Ovidiu, Qijun, Paulina, Arjen, Johan, Marcio, Xuemeng, Martijn Witjes, Carmelo, Tom Hengl, Gerard, Jeroen, Matthias Mohr, Marius, Matthias Schramm, Peter Zellner, Alexander Jacob, Christian, Benjamin Schumacher, Chara, Raymond, Kor, Derek, Nevena, Sebastien, Marcel, Bruno, Steffen, Myroslava, Yang Li and everyone else whom I may have forgotten to mention.

I would also like to thank my friends from MGI and GIMA, including Eline, Christina, Laura, Robbert-Jan, Madeleine, Lieven, Roeland, Geetika, Merijn, William, Jasper, Tim, Maria Herrera, Careli, Samantha Krawczyk, Jorn, Maartje, Jos, Nicolas, Maria Christy, Elke, Liliana, Bas, Daan, Thanasis, Sophie, David, Julian, Matthew, Rodrigo, Yneke, Amy, Louise, Ping, Yingbao, Yiran, Yue, Yuting Zou, Jinwen, Thijs, Gerjon, Sytze, Xinzhi, Chenyue, Xuerui, Yuhan, Rong, Paula, Shoyo, Martijn van Sluijs, Kim, Niamh, Zina, Zoi, Friedrich, Jacotte, Yelle, Shuying, Muxu, Jiaxin, Yi An, Sven-Arne, Rob, August, Liam, Stijn, Ivan, Nik, Gabrielė and many others. I hope to see many of you again in the future in MGI anniversary and alumni events!

I would also like to extend my gratitude to all the other friends of mine who have supported me during the PhD journey, including Elena, Greta, Merle, Yi Li, Meixiu, An(na), Jin, Mengting, Yanjie, Keli, Chunfeng, Huifang, Hui, Qingbo, Wannida, Frances, Ying, Qi, Pengyao, Yunya, Damian, Lina, Dainis, Yutaka, Thibault, Weiyi, Caixia, Jiaqi, Ke, Luc, Simona, Yuting Tai, Congcong, Zhizhen, Solen and many others. Special thanks to Claudius and Lennart from PE&RC, as well as the rest of the PE&RC staff, for all the guidance and support, and the opportunity to meet many of the aforementioned people in social and educational events! Through you I have learned in practice that a doctorate 博士 is inseparable from universal love 博愛.

Special thanks goes to my family for being there throughout my whole life and supporting me during the PhD, with all its twists and turns. Thanks to my brother Aidas and his fiancée Lei, whose footsteps I have followed by arriving in the Netherlands; how wonderful that we can often see each other live, and in the future perhaps even more often! Aidas, thank you for being an inexhaustible source of jokes as well as deep insight, and for all your unwavering support throughout. Lei, I still remember the day we all met in Maastricht; already then I thought that Aidas was very lucky to know you! I'm very happy for you two and wish all the best in the future. I also thank my mother Loreta, who has always been very kind and supportive, for all the wisdom shared, and for all the strength needed when both children end up leaving far away. Creating and maintaining a company while also raising children is no small feat! I wish you a much more relaxed time in the future, with more time for yourself and less stress.

Last but certainly not least, a deepest thank you goes to my wife, 楚麗霞 (Lixia Chu), and also her family, my family-in-law. 楚楚 (Lixia), I will always remember the day we met in Milan, and how you graciously hosted me for almost a year in Salzburg. You have sacrificed a lot to follow me all the way here to Wageningen! Your support throughout the thesis process, the urging to spend more time on it and protecting my time from other distractions have significantly improved the quality of my work. Your love and support are phenomenal, always reciprocated and our happiness together will be everlasting.

About the author

Dainius Masiliūnas was born on April 7, 1993 in Vilnius, Lithuania. Between 2006 and 2011 he studied in Vilnius Lyceum, the top high-school in Lithuania, in a class specialising in biochemistry. In 2011, he started a BSc in Ecology at Vilnius University. He graduated with a *cum laude* distinction in 2015, after defending his thesis on the correlation between the abundance and distribution of Lithuanian bird species.

His GIS-focused thesis and his background in ecology prompted him to further pursue education in a field combining both GIS and ecology. Namely, he moved to the Netherlands to pursue the Master of Geo-Information Science (MGI) in Wageningen University & Research. As part of the MSc programme, he defended a major thesis entitled "Fuzzy land cover classification method assessment using PROBA-V satellite data" and a minor thesis entitled "Evaluating the potential of Sentinel-2 and Landsat image time series for detecting selective logging in the Amazon". He graduated from MGI with a *cum laude* distinction in 2017.

After graduating, Dainius was offered to continue the research based on his major thesis in a PhD track in the Laboratory of Geo-information Science and Remote Sensing (GRS) of Wageningen University. As part of the PhD track, Dainius was involved in the Horizon 2020 project "openEO", Copernicus Global Land Operations project "Vegetation and Energy" on moderate land cover, 100 m and COST action "SENSECO". His PhD focused on predicting global land cover fractions and change detection in land cover time series, using PROBA-V, MODIS, Landsat and Sentinel-2 satellite imagery.

In 2020, Dainius started a lecturer position at the GRS chair group, focusing on coordinating the "Geoscripting" course as part of the MGI programme, but continuing to pursue his PhD part-time. In 2022 and 2023, he was involved in the successful acquisition of three more research projects: SURF "Research Cloud Community Engagement", Digital Europe "The Green Deal Data Space Foundation and its Community of Practice" (GREAT) and Horizon Europe "Open Earth Monitor Cyberinfrastructure".

After finishing his PhD, Dainius is going to continue his work in the GRS chair group with a permanent lecturer position, with longer-term prospects of entering the tenure track.

Peer-reviewed Journal Publications

Masiliūnas, **D.**, N.-E. Tsendbazar, M. Herold, M. Lesiv, M. Buchhorn, and J. Verbesselt (2021a). "Global land characterisation using land cover fractions at 100 m resolution". *Remote Sensing of Environment* 259, 112409. DOI: 10.1016/j.rse.2021.112409.

Masiliūnas, **D.**, R. Burger, N.-E. Tsendbazar, and D. Marcos (2024). "Improving global land cover fraction change detection using a Markov chain model". Under review.

Masiliūnas, **D.**, N.-E. Tsendbazar, M. Herold, and J. Verbesselt (2021b). "BFAST Lite: A Lightweight Break Detection Method for Time Series Analysis". *Remote Sensing* 13.16, 3308. DOI: 10.3390/rs13163308.

Masiliūnas, **D.**, N.-E. Tsendbazar, M. Herold, D. Marcos, and J. Verbesselt (2023a). "Postprocessing high frequency global land cover maps for change assessment". Under revision.

Xu, L., M. Herold, N.-E. Tsendbazar, **D. Masiliūnas**, L. Li, M. Lesiv, S. Fritz, and J. Verbesselt (2022). "Time series analysis for global land cover change monitoring: A comparison across sensors". *Remote Sensing of Environment* 271, 112905. DOI: 10.1016/j.rse.2022.112905.

Chu, L., Y. Zou, **D. Masiliūnas**, T. Blaschke, and J. Verbesselt (2021). "Assessing the impact of bridge construction on the land use/cover and socio-economic indicator time series: A case study of Hangzhou Bay Bridge". *GIScience & Remote Sensing* 58.2, 199–216. DOI: 10.1080/15481603.2020.1868212.

Slomp, A., **D. Masiliūnas**, and N.-E. Tsendbazar (2024). "Improving the robustness of global land cover fraction change detection using temporal deep learning". In preparation.

Li, Y., M. A. Wulder, Z. Zhu, J. Verbesselt, **D. Masiliūnas**, Y. Liu, G. Bohrer, Y. Cai, Y. Zhou, Z. Ding, and K. Zhao (2024). "Detecting breakpoints in multispectral time series – a multivariate algorithm". In review.

Tsendbazar, N., M. Herold, L. Li, A. Tarko, S. de Bruin, **D. Masiliunas**, M. Lesiv, S. Fritz, M. Buchhorn, B. Smets, R. Van De Kerchove, and M. Duerauer (2021). "Towards operational validation of annual global land cover maps". *Remote Sensing of Environment* 266, 112686. DOI: 10.1016/j.rse.2021.112686.

Wang, N., J. Suomalainen, H. Bartholomeus, L. Kooistra, **D. Masiliūnas**, and J. G. Clevers (2021). "Diurnal variation of sun-induced chlorophyll fluorescence of agricultural crops observed from a point-based spectrometer on a UAV". *International Journal of Applied Earth Observation and Geoinformation* 96, 102276.

Wang, N., B. Siegmann, U. Rascher, J. G. Clevers, O. Muller, H. Bartholomeus, J. Bendig, **D. Masiliūnas**, R. Pude, and L. Kooistra (2022). "Comparison of a UAV- and an airborne-based system to acquire far-red sun-induced chlorophyll fluorescence measurements over structurally different crops". *Agricultural and Forest Meteorology* 323, 109081.

Chu, L., J. Nelen, **D. Masiliūnas**, C. Hein, and C. Lofi (2024). "Relationships between geo-spatial features and COVID-19 hospitalisations revealed by machine learning models and SHAP values". Under revision.

Liu, Y., G. B. Heuvelink, Z. Bai, P. He, X. Xu, J. Ma, and **D. Masiliūnas** (2020). "Space-time statistical analysis and modelling of nitrogen use efficiency indicators at provincial scale in China". *European Journal of Agronomy* 115. Publisher: Elsevier, 126032.

Luan, H., W. Gao, S. Huang, J. Tang, M. Li, H. Zhang, X. Chen, and **D. Masiliūnas** (2020a). "Substitution of manure for chemical fertilizer affects soil microbial community diversity, structure and function in greenhouse vegetable production systems". *PLoS One* 15.2, e0214041.

– (2020b). "Organic amendment increases soil respiration in a greenhouse vegetable production system through decreasing soil organic carbon recalcitrance and increasing carbon-degrading microbial activity". *Journal of Soils and Sediments* 20.7, 2877–2892. DOI: 10.1007/s11368-020-02625-z.

Luan, H., W. Gao, J. Tang, R. Li, M. Li, H. Zhang, X. Chen, **D. Masiliūnas**, and S. Huang (2020c). "Aggregate-associated changes in nutrient properties, microbial community and functions in a greenhouse vegetable field based on an eight-year fertilization experiment of China". *Journal of Integrative Agriculture* 19.10. Publisher: Elsevier, 2530–2548.

Other Scientific Publications

Masiliūnas, **D.**, D. Marcos, N.-E. Tsendbazar, M. Herold, and J. Verbesselt (2023b). "Making dense land cover products more consistent and meaningful using BFAST Lite postprocessing". In: *AGILE 2023*. Delft, the Netherlands.

– (2022a). "High frequency land cover classification method for supporting global monitoring". In: *AGILE 2022*. Vilnius, Lithuania. DOI: 10.5281/zenodo.8310036.

– (2022b). "BFAST Lite for detecting land cover change using time series of estimated land cover fractions". In: *ESA Living Planet Symposium 2022*. Bonn, Germany.

Masiliūnas, **D.**, N.-E. Tsendbazar, M. Herold, M. Lesiv, and J. Verbesselt (2019). "Change detection in satellite image time series for continuous land cover map updating". In: *ESA Living Planet Symposium 2019*. Milan, Italy.

Masiliūnas, **D.**, N.-E. Tsendbazar, and J. Verbesselt (2018). "Comparison of machine learning algorithms for fuzzy land cover classification". In: *PROBA-V Symposium 2018*. Ostend, Belgium.

Buchhorn, M., L. Bertels, R. van de Kerchove, B. Smets, M. Herold, N.-E. Tsendbazar, **D. Masiliūnas**, S. Fritz, M. Lesiv, and M. Duerauer (2018). "The yearly 100m Global Land Cover Layer from Copernicus Global Land". In: *PROBA-V Symposium 2018*. Ostend, Belgium.

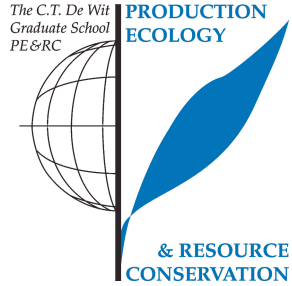
Buchhorn, M., L. Bertels, B. Smets, B. De Roo, M. Lesiv, N.-E. Tsendbazar, **D. Masiliūnas**, and L. Li (2020). *Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015-2019: Algorithm Theoretical Basis Document*. Tech. rep. Publisher: Zenodo.

Buchhorn, M., M. Lesiv, N.-E. Tsendbazar, B. Smets, L. Bertels, R. Van De Kerchove, M. Herold, **D. Masiliūnas**, and S. Fritz (2019). "Copernicus Global Land Cover Service-Elastic, Operational Land Cover Mapping at Global Scale using Time Series Analysis". In: *AGU Fall Meeting Abstracts*. Vol. 2019, B23C–02.

Chu, L., **D. Masiliūnas**, A. Crivellari, and C. Lofi (2022). "Response of air quality to Covid-19 lockdown policies from Sentinel-5P TROPOMI sensor". In: *Living Planet Symposium 2022*.

PE&RC Training and Education Statement

With the training and education activities listed below the PhD candidate has complied with the requirements set by the C.T. de Wit Graduate School for Production Ecology and Resource Conservation (PE&RC) which comprises of a minimum total of 32 ECTS (= 22 weeks of activities)



Review/project proposal (9 ECTS)

- Large-scale land cover change monitoring from dense time series of satellite data; remote sensing thematic group meeting
- Large-scale land cover change monitoring from dense time series of satellite data

Post-graduate courses (4 ECTS)

- Geocomputation using free and open source software; PE&RC (2019)
- Summer school; OpenGeoHub (2020)

Laboratory training and working visits (4.5 ECTS)

- OpenEO second intense collaboration meeting; University of Münster (WWU) (2018)
- Copernicus Global Land Operations face-to-face meeting; VITO (2018)
- OpenEO Hackathon; TU Wien (2018)
- CGLOPS Workshop; VITO (2018)
- SURF Research Cloud community engagement hackathon; Utrecht University (2022)
- Green Deal Data Space and its community of practice meeting; SURF (2023)
- Open Earth Monitor Cyberinfrastructure OpenEO workshop; University of Münster (WWU) (2023)

Invited review of journal manuscripts (4 ECTS)

- Ecological Informatics: CropPhenology: an R package for extracting crop phenology from time-series remotely sensed vegetation index imagery (2017/2018)
- International Journal of Remote Sensing: retrieving phenological parameters of tea crop through time series analysis of SENTINEL-2 satellite images: a case study in north-east Black Sea region of Turkey (2021)
- Remote Sensing of Environment: graph-based monthly block-level urban change detection using Sentinel-2 time series (2021)
- Remote Sensing of Environment: leveraging past information and machine learning to accelerate land disturbance monitoring (2023)

Competence, skills and career-oriented activities (7.92 ECTS)

- Project and time management; WGS (2018)
- Supervising BSc and MSc thesis students; Education support (2018)
- HPC Basic; WUR (2018)
- Scientific writing; Wageningen in'to Languages (2018)
- PhD Workshop carousel; WGS (2019)
- Supervising gifted students; Education support (2019)
- (Re)designing a course; Education support (2020)
- Start to teach; Education support (2020)
- Assessment: written exams – blended course; Education support (2020)
- Teaching in the intercultural classroom; Education support (2021)
- Assessment: graded assignments; Education support (2021)
- Scientific integrity; WGS (2021)
- University teaching; Education support (2021)

PE&RC Annual meetings, seminars and PE&RC weekend/retreat (2.1 ECTS)

- PE&RC Weekend for first years (2018)
- PE&RC Midterm weekend (2019)
- PE&RC Last year retreat (2022)

Discussion groups/local seminars or scientific meetings (10.6 ECTS)

- Discussion group on agriculture – climate – forests – food (2017)
- R Users meeting (2017–2022)
- Time series focus group meeting (2018–2019)

International symposia, workshops and conferences (24 ECTS)

- Proba-V Symposium (2018)
- NCG Symposium (2018)
- ESA Living Planet Symposium (2019)
- FOSS4G (2019)
- ESA Phi week (2019)
- International symposium of Digital Earth (2021)
- IGARSS (2021)
- Open Data Science Europe (2021)
- ESA Living Planet Symposium (2022)
- AGILE 2022
- AGILE 2023

Lecturing/supervision of practicals/tutorials (21.9 ECTS)

- Geoscripting (2018–2023)
- Introduction to LaTeX (2019–2023)
- Advanced Earth Observation (2020–2022)
- Machine learning (2020–2022)
- Deep learning (2020–2023)

BSc/MSc thesis supervision (24 ECTS)

- Distinguishing natural versus plantation forests in South Africa using Sentinel-2 imagery
- Quantifying the effect of surrounding forest cover on the recovery rate of secondary forests using Landsat timeseries
- Tracing the footprints of human activities from the dynamics of land use change in Hangzhou Bay
- Detecting and characterising water dynamics using MODIS time-series imagery
- Comparing multi-temporal unsupervised land cover change detection algorithms on a global scale
- Analysing the relation between land use and subsidence in the Randstad in the Netherlands
- Improving global land cover fraction change detection using a Markov chain model
- Improving the robustness of global land cover fraction change detection using temporal deep learning

The research described in this thesis was financially supported by the Horizon 2020 project "openEO - a common, open source interface between Earth Observation data infrastructures and front-end applications" (grant no. 776242), Copernicus Global Land Operations "Vegetation and Energy" ("CGLOPS-1") Framework Service Contract (grant no. 199494), Digital Europe project "The Green Deal Data Space Foundation and its Community of Practice (GREAT)" (project no. 101083927) and Horizon Europe project "Open Earth Monitor Cyberinfrastructure" (grant no. 101059548).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover design by Dainius Masiliūnas and DALL-E 2

Printed by ProefschriftMaken on FSC-certified recycled paper