The improvement of land use and land cover representation in regional climate models

Dissertation
with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Earth Sciences
at Universität Hamburg

submitted by Vanessa Reinhart

from Karlsruhe, Germany

Hamburg, 2021

Department of Earth Sciences

Date of Oral Defense:	07.02.2022
Reviewers:	Prof. Dr. Jürgen Böhner
	Dr. Diana Rechid
Members of the examination commission:	Prof. Dr. Johanna Baehr, Prof. Dr. Frank Lunkeit, Prof. Dr. Udo Schickhoff
Chair of the Subject Doctoral Committee Earth System Sciences:	Prof. Dr. Dirk Gajewski
Dean of Faculty MIN:	Prof. Dr. Heinrich Graener

Acknowledgements

Throughout the writing of this dissertation I have received a great deal of support and assistance from my colleagues and from my family and friends. First and foremost, I would like to thank my supervisors Diana Rechid and Jürgen Böhner for their scientific support during the PhD period. Both were highly approachable and present for me at any time during the conception and execution of my research in the last three and a half years. I would like to acknowledge my colleagues from the LANDMATE project Peter Hoffmann and Christina Asmus, who supported me through all the difficult phases of this past time. Special thanks goes to all my colleagues and friends at the University of Hamburg, who gave me not only the professional but also the emotional support I needed to complete this work. Last but not least, I want to thank my mother for her unconditional support throughout the years. I would not be where I am today without her.

Abstract

Land use and land cover change (LULCC) is one of the essential variables in climate modelling. The accurate representation of feedback mechanisms between the land surface and the atmosphere is crucial for Regional Climate Models (RCMs) to produce reliable information about the past, present and future climate. Several LULC products are available and used in RCM simulations. However, the LULC products used often lack the required spatial and temporal resolution and, most importantly, thorough validation. In order to overcome the short-comings of exiting LULC datasets this thesis introduces a newly developed, high-resolution plant functional type (PFT) map for Europe for the use in regional climate models, LANDMATE PFT 2015 Version 1.0. The map is generated translating high-resolution LULC data (ESA-CCI LC) into PFTs for the European continent. High-resolution climate data is used to differentiate the translation of LULC classes for 30 climate zones depending on temperature and precipitation data. Under consideration of the 30 climate zones, the LULC classes from the ESA-CCI LC map 2015 are translated into PFTs using a newly developed cross-walking procedure. The validation of the map is provided through comparison of LANDMATE PFT 2015 with an extensive ground truth database for the European continent. The newly developed validation workflow makes it possible to compare fractional LULC maps to a ground truth dataset, which is provided as point samples. The validation results show improvement of LANDMATE PFT in comparison to the ESA-CCI default PFT map and provide regional quality information to regional climate modellers. The workflow is tailored to the validation of fractional LULC maps and can be applied to already existing LULCC or PFT map products in order to tackle the challenge that arises from uncertainty in LULCC input in RCMs.

Contents

List of Tables

List of Figures

List of Abbreviations

CLC	CORINE Land Cover inventory
CWT	Cross-Walking Table
CWP	Cross-Walking Procedure
ESA-CCI	European Space Agency Climate Change Initiative
GT-SUR	LUCAS ground truth survey
LANDMATE	Modelling human LAND surface Modifications and its feedbacks on local and regional cliMATE
LUCAS	Land Use and Climte Across Scales
LULC	Land Use and Land Cover
LULCC	Land Use and Land Cover Change
PAC	Proportional Area Comparison
PFT	Plant Functional Type
RCM	Regional Climate Model

1. Introduction

The regional climate models (RCMs) used today require high quality and high resolution land use and land cover change (LULCC) data in order to generate realistic and reliable climate hindcast simulations and climate projections as a basis for climate change research. However, the availability of suitable LULCC input data is limited due to insufficient observations or observations in an unsuitable temporal or spatial resolution or data format.

In order to conduct RCM experiments focused on simulations that require realistic LULC conditions, the land use and land cover (LULC) product implemented into an RCM needs to be selected according to the respective research objective. However, for the majority of RCMs, one of the well-known LULC product is used for simulations (Bontemps et al., 2012a), regardless the present research objective.

The importance of of high-quality LULCC representation in RCMs is investigated in multiple studies for the European continent (Davin et al., 2020; Perugini et al., 2017; Strandberg et al., 2019). It is shown that the impact and feedback effects of LULC and further LULCC on near-surface climate parameters, such as temperature and precipitation is non-negligible when investigating the climate system in a changing environment (de Noblet-Ducoudré et al., 2012). This impact is expected to be increasing when moving towards convective permitting scales of 3 km or less (Adinolfi et al., 2021; Tölle et al., 2021)

In the course of a user analysis in the context of the European Space Agency Climate Change Initiative (ESA-CCI, Bontemps et al., 2012b) a survey was conducted among the climate modelling community in order to identify the commonly used products in climate modelling (Bontemps et al., 2012a). According to the survey, mainly 12 LULC products are used among the community, most of which are only provided for one time step. Most used LULC products are the IGBP DISCover[1] by the United States Geological Service (USGS, Loveland et al. (2000)), the GLC2000 (Bartholome et al., 2005), the MODIS[2] products (Friedl et al., 2010; Sulla-Menashe et al., 2018) and the GlobCover database (Arino et al., 2012; Defourny et al., 2006). The versions of the products are not defined in the survey results so it is assumed that the products are used in all of their so far published versions. Beside the varying levels of LULCC accuracy and the differences in classification detail, several comprehensive comparative approaches show considerable differences between these products. Giri et al. (2005) and Herold et al. (2008) show large areas of disagreement between GLC2000 and MODIS over Europe. Herold et al. (2008) further include the IGBP DISCover dataset into the comparison while it was added to Giri et al. (2005) in a follow up approach (McCallum et al., 2006). The extreme disagreement between the IGBP DISCover dataset and GLC2000 is confirmed by both approaches.

GLC2000, MODIS (for the year 2005) and GlobCover (2004-2005) were compared

[1] International Geosphere-Biosphere Program Data and Information System
[2] Moderate Resolution Imaging Spectroradiometer

by Pérez-Hoyos et al. (2012). The fourth dataset in the comparison is the CORINE land cover inventory (CLC) for Europe (Büttner et al., 2004), which limited the region suitable for the comparison to the European continent, more precisely the 27 countries covered by CLC. The overall agreement between the assessed datasets is noticeably low. A major finding is that cropland is underestimated by GlobCover compared to the other datasets, which are in relatively good agreement regarding the overall cropland proportions. On the other hand, sparse vegetation is strongly underestimated by GlobCover. MODIS 2005 shows deviating proportions of mixed vegetation, which is due to the classification harmonization, where the MODIS LC class "Woody savannahs", that does not exist in the other datasets, is assigned to the mixed vegetation. Major to minor differences between the LULC class proportions are found for all other classes, where the smallest differences are found for needleleave vegetation. The low consistency between GLC2000, GlobCover and MODIS is confirmed by (Hua et al., 2018), where the focus is on multiyear consistency. The agreement between MODIS an GlobCover increases from GlobCover 2005 to GlobCover 2009 but is still below 65 %. Reasons for the strong disagreement between the datasets identified by multiple comparative studies are different methods of data acquisition, classification and different spatial resolution. Nevertheless, all datasets are used simultaneously within the climate modelling community and the dataset differences cause uncertainty in simulations. The impact of these uncertainties in RCM experiments arising from differences in the used LULC products, is not negligible.

Sertel et al. (2010) created and used an updated LULC map and compared the

RCM simulation results to the results using the default GLCC[3] map, where the focus of the newly created map was set on the precise representation of urban expansion in the Marmara Region in Turkey. While the impact on precipitation could not be assessed due to dry weather conditions in the simulation period, temperature representation was improved noticeably. A positive effect of more accurate LULC representation on the simulation of wind estimates was shown by Santos-Alamillos et al. (2015) and De Meij et al. (2014). The comparison of two RCM simulations, one with the GLCC dataset, and one with CLC showed, that the CLC based simulation is advantageous regarding the simulation of wind estimates over Southern Spain. The positive effect of the implementation of updated LC (CLC and MODIS) into a high-resolution RCM was also confirmed over two pilot regions in Austria Schicker et al. (2016). A study over the complex terrain of the Eastern Pyrenees showed, that with the implementation of CLC the entire RCM performance improved on average compared to the simulation using the model specific default LULC data Jiménez-Esteve et al. (2018).

In order to overcome the uncertainty caused by using different LULC products with different classifications and levels of accuracy, a new product is needed, that is customized to the specific user needs of the RCM community. The European Space Agency Climate Change Initiative (ESA-CCI) tackled the challenge, that arises from LULC uncertainty caused by LULC products in RCMs and released a remote sensing based land cover time series (ESA-CCI LC, ESA, 2017) on a continuous, global 300 m grid. The time series was created following the investigation on the requirements of climate modellers to a LULCC product (Bontemps et al., 2012a). Summarized, the

[3]Global Land Cover Characterization, Loveland et al., 2000

requirements are focused on (1) a high-resolution (desirably 300 m or coarser) product that represents intra-annual or even monthly dynamics regarding phenology and LULCC, (2) a product with a LULC classification that is translatable to the Plant Functional Type (PFT) concept used by many RCMs and (3) transparent quality information. The ESA-CCI LC dataset covers the period from 1992 to 2020 in annual time steps. The product is based on sensors from different satellite missions as shown in table 1.1

Table 1.1: Satellite missions involved in the production of ESA-CCI LC according to ESA, 2017

Time period	Satellite product
Baseline Production 2003-2012	MERIS FR/RR [4] global SR [5] composites
1992-1999	Baseline 10-year global map; AVHRR [6] global SR composites for back-dating baseline
1999-2013	Baseline 10-year global map; SPOT-VGT [7] global SR composites for up and back-dating the baseline; PROBA-V [8] global SR composites at 300 m
2013-2015	Baseline 10-year global map; PROBA-V global SR composites at 1 km for years 2014 and 2015 for updating the baseline; PROBA-V time series at 300 m
Since 2016	Sentinel-3 OLCI and SLSTR [9] 7-day composites

[4]MEdium Resolution Imaging Spectrometer Full Resolution/Reduced Resolution (E. A. P. ESA, 2002)
[5]Surface Reflectance
[6]Advanced Very-High-Resolution Radiometer (Hastings et al., 1992)
[7]SPOT Vegetation satellite program (Maisongrande et al., 2004)
[8]Project for On-Board Autonomy - Vegetation (Dierckx et al., 2014)
[9]Ocean and Land Colour Instrument (OLCI) and Sea and Land Surface Temperature Radiometer (SLSTR) (Donlon et al., 2012)

The ESA-CCI LC is quality checked, which includes the comparison to a global LC ground truth database with validation through regional experts. The global, overall accuracy is over 75 % for the 2015 map (ESA, 2017), while the regional accuracy of the dataset is assessed by independent researchers (Karvonen et al., 2018; Koubodana et al., 2019; Majasalmi et al., 2021; Samasse et al., 2018; Vilar et al., 2019; Yang et al., 2017). However, the investigation of regional accuracy over Europe has received little attention so far (Reinhart et al., 2021a). The classification of ESA-CCI is based on the United Nations Land Cover Classification protocol (UN-LCCS, Di Gregorio (2005)) and includes various mixed classes without specification of the respective LULC type proportions within a grid cell. However, the survey conducted by Bontemps et al. (2012a) regarding user needs within the RCM community revealed the translatability of the LULC classification into RCM specific classes or PFTs as a main issue.

For the translation of the ESA-CCI LC classes into PFTs, the ESA-CCI provides a dedicated user tool. The tool is set up for the user to facilitate processing of the ESA-CCI LC maps, such as translating the maps into the desired PFTs or resampling the maps into the desired resolution. Within the user tool, the ESA-CCI provides a default PFT translation (Poulter et al., 2015) where the ESA-CCI LC map classes are translated into 10 PFTs. An urban PFT was added shortly after publication to separate the urban areas, that were translated mainly to bare soil before.

The translation from LULC classes into PFTs is not trivial considering the definitions of the mixed classes, where only a percentage range of each LULC type proportion is given. When individual proportions of PFT types are over- or underestimated, the total PFT ratio is necessarily changed. The differences in global trends shown by ESA-CCI

PFT in comparison to other available LULC maps are shown by W. Li et al. (2018), where the ESA-CCI PFT time series is compared to other available long term LULC time series. Further, differences between other available PFT datasets are shown by Hoffmann et al. (2021). The global comparison between the MODIS PFT and the ESA-CCI LC PFT time series shows clear differences not only in the proportions but also in the global LULC change trends, especially in Europe. Hartley et al. (2017) elaborate on the impact of uncertainty in PFT maps for RCMs. However, no quality information is given for the two PFT time series (ESA-CCI PFT and MODIS PFT), that could verify the LULC status or the regional trends.

Consequently, within the creation of a new PFT time series, quality information needs to be considered. An accuracy assessment workflow for determining the quality of a LULC product or time series needs to be developed and follow certain standards but nevertheless, needs to be adjusted to the respective LULC product and to the purpose of the respective assessment Foody (2015), Olofsson et al. (2014), and Stehman et al. (2019). Summarized, the assessment workflow needs to be carried out with as little as possible modification of the assessed dataset and with as much detail as possible regarding the classification. The workflow needs to be provided transparently to the user in order to ensure total clarity of the resulting accuracy measures in order to determine strengths and weaknesses not only of the assessed dataset but also of the performed assessment.

This cumulative dissertation is conceptualized to fill the research gaps that are identified regarding the use of LULCC products in RCMs, which are formulated as follows:

- The four LULCC products mainly used in RCMs are partly outdated and not valdiated, yet used simultaneously among each other. There is a need for compaprison of and investigation of differences between these commonly used products.
- In between the LULCC products are considerable differences in LULC proportions and classifications that lead necessarily to uncertainty in between simulations. It needs to be investigated how these commonly used products contribute to inter-model variability and uncertainty.
- Little to no attention is paid by the RCM community to validation of LULCC products used in RCMs. A validation workflow needs to be developed that provides quality information for LULCC products used in regional climate modelling.
- The translation of LULC classes into RCM specific LULC classification necessarily changes the LULC class proportions and needs to be addressed as an uncertainty within RCM simulations. The modifications applied due to implementation into RCM specific LULC classifications need to be investigated and documented thoroughly.

Section 2.1 gives an overview of the four main research questions of this dissertation. The research design is elaborated in chapter 2.2 where the steps taken to answer the research questions within the associated publications are shortly summarized. Chapter 4 connects the individual publications through an overarching discussion of the methods and the results. Overall conclusions and an outlook to future tasks that necessarily have to follow up to this dissertation are given in chapter 5.

2. Research approach

This cumulative dissertation is incorporated within the LANDMATE (Modelling human LAND surface Modifications and its feedbacks on local and regional cliMATE)[10] project, which was launched in 2018. LANDMATE is one of the seven projects within the Helmholtz Institute for Climate Service Science" (HICSS)[11], a cooperation between the Climate Service Center Germany (GERICS) and the Universität Hamburg.

The LANDMATE project is dedicated to improve the representation of LULCC in regional climate models and to to include land management practices, such as irrigation into RCMs. Further, in the course of the participation in the WCRP[12] CORDEX[13] FPS LUCAS[14], the aim of LANDMATE is the quantification of the signal of LULCC on regional climate in Europe. Within this framework of LANDMATE, the overall goals of this thesis are focused on creating the prerequisite for the FPS LUCAS through development and provision of validated, high-resolution LULCC input data.

In order to address these overall goals, the following main objectives were defined:

[10] https://www.hicss-hamburg.de/projects/landmate/index.php.en
[11] https://www.gerics.de/science/hicss/index.php.de
[12] World Climate Research Programme, https://www.wcrp-climate.org/
[13] COordinated Regional Downscaling EXperiment, https://cordex.org/
[14] Flagship Pilot Study Land Use and land Cover Across Scales, https://ms.hereon.de/cordex$_{f}ps_{l}ucas/index.php.en$

(1) the investigation of differences between LULC datasets used in RCMs focused on available continuous, high-resolution time series. The collection of available LULC products prepared objective (2), the conceptualization and production of a new PFT dataset that is tailored to the needs of regional climate modellers in Europe. (3) Special attention is paid to validation of LULC products by developing a tailored accuracy assessment procedure for the newly developed dataset.

2.1 Research questions

The initial research questions are formulated as follows according to the research gaps identified in chapter 1.

1. **Which datasets are suitable to produce a consistent historical time series of a LULCC dataset for RCMs over Europe?**

   The in-depth investigation on currently used LULC data in RCMs elaborated in chapter 2 shows that the currently used data is to a large extent not suitable for the challenges and tasks of future RCM experiments. Following that finding, currently available LULC products are collected and checked for temporal and spatial resolution and the overall quality of LULC representation in particular.

2. **How can the hot spots of LULCC in Europe be validated?**

   The first step revealed that the ESA-CCI LC dataset released by ESA in 2017 is most suitable for the upcoming analysis and the task to produce a high-quality and high-resolution LULCC time series for RCMs. However, to that point a quality assessment of ESA-CCI LC focused on and tailored to the European continent was missing. As a necessity, the hot spots of LULCC in Europe that are suggested by the ESA-CCI LC for Europe need to be reviewed and verified.

3. **Which methods are suitable to produce a consistent historical, high-quality and high-resolution time series of a LULCC dataset for Europe?**

   In the previous quality assessment of ESA-CCI LC over Eastern Europe, the dataset is considered to be suitable to represent the hot spots of LULCC over Europe accurately enough for ESA-CCI LC to be the baseline product for a LULCC time series for RCMs. The baseline product needs to be produced under the premise that it has to be flexibly implementable into multiple RCM families but at the same time does not forfeit LULC information and therefore, quality. Another requirement to a newly developed product is the fine scale, regional applicability and accuracy which can only be achieved by the consideration of specific regional information beyond the pure LULC information derived from the ESA-CCI LC dataset.

4. **What is a suitable method to determine the quality of the newly developed PFT dataset?**

   The translation of LULC classes and the merge of multiple additional datasets necessarily results in a change of quality of the new product in comparison to the baseline product. Therefore, it is inevitable to assess the quality of the final product respectively. Challenges, such as the fractional structure of the dataset, need to be tackled in the quality assessment design in a way that produces useful quality information for climate modellers, which are the target user group.

2.2 Research design

In the first work of this cumulative dissertation, proportional area comparison (PAC) is used to validate the regional accuracy of ESA-CCI LC over Eastern Europe (Reinhart et al., 2021a). This first study is designed to give insights on the reliability of the LULCC suggested by ESA-CCI LC in order to evaluate the suitability of ESA-CCI LC as baseline product for the development of a new LULC product. According to Liu et al. (2018), ESA-CCI suggests a LULCC of 5.36 % of the land area in Europe corresponding to 0.53 Mio km^2, where 42 % of the LULC transitions are forest to cropland or vice versa. The largest proportion of these most common transitions occurs in Eastern Europe, which is why the area was selected as pilot region for the comprehensive accuracy assessment. Challenges in accuracy assessment of LULC products elaborated in Reinhart et al., 2021a should be overcome using the PAC, a method that can be applied to compare two LULC products with different structure and resolution without spatial modification. Although the advantages of the approach did not become evident in the comparison over Eastern Europe, a PAC over Portugal by Fonte et al. (2020) shows that the structure and different resolution of the reference and the assessed map have non-negligible influence on the accuracy measures.

The comparison with CLC reveals the strengths and weaknesses of ESA-CCI LC in the pilot region. The main LULC types (forest and cropland) are represented very well while the lesser represented types, such as urban and shrubland show minor uncertainties compared to CLC. Following this assessment, the ESA-CCI LC was selected to be the basemap for the upcoming task.

Following the preparing investigation on suitable input data within the framework of the thesis, a PFT dataset was developed for the use in RCMs. The newly developed LANDMATE PFT dataset[15] is a high-resolution PFT map for Europe. The PFT map is available in two spatial resolutions (0.1° and 0.018°) for the year 2015. The PFT distribution is based on the ESA-CCI LC map for 2015. The translation from ESA-CCI LC classes to PFTs is based on the translation from Wilhelm et al. (2014), where the GLOBCOVER2000 dataset, which shares a similar LULC classification with ESA-CCI LC, is translated into the 16 REMO-iMOVE PFTs. During the development of LANDMATE PFT 2015, the existing translation is thoroughly revised and adjusted to the ESA-CCI LC classes as elaborated in the corresponding publication (Reinhart et al., 2021b). The revision of the translations was done closely accompanied by extensive comparison of the product with validated, high-quality and high-resolution LULC maps and the inclusion of expert knowledge. The use of climate data in the form of Holdridge Life Zones (HLZs) makes it possible to adjust the PFT fraction proportions according to different climate zones and with that, to account for the variety of vegetation communities along the latitudinal range.

The comprehensive development process of LANDMATE PFT 2015 should ensure a very high quality of the map. Nevertheless, validation needs to be done in order to provide the required quality information to the user. For the LANDMATE PFT dataset for 2015, which is of fractional structure, validation with ground truth data is even more challenging than for a product, that contains only one LULC type per grid cell. Recent approaches tackled the challenge by comparing PFT products or PFT distributions manufactured for the respective purpose to other existing LULC datasets (Hoffmann

[15] LANDMATE PFT land cover dataset for Europe 2015 (Version 1.0) available at https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=LM_PFT_LandCov_EUR2015_v1.0

et al., 2021; W. Li et al., 2018; Liang et al., 2015; Törmä et al., 2015). A comparison of a PFT product to real ground truth data has not been published to this time. Reinhart et al. (2021b) present a novel approach for the validation of fractional, gridded data, where one unit of a grid cell contains more than one LULC type that is furthermore not precisely located within the respective grid cell. Using the proposed approach it is ensured, that the best practice recommendations for LULC accuracy assessments are followed thoroughly and that the accuracy measures are comprehensible and reproducible. The approach makes use of the most comprehensive ground truth database for Europe and produces a transparent, spatial overview over the strengths and weaknesses of the LANDMATE PFT dataset for Europe for 2015.

2.3 Novelty of the approach

The entire approach followed within this cumulative PhD thesis is designed to tackle the uncertainty that arises from the use of LULC products in RCMs. Latest findings from remote sensing based research on LULC map production and validation are transferred to benefit the improvement of LULC representation in RCMs. While in remote sensing research, validation techniques for LULCC products are advanced and constantly improved, the techniques are not used for products used in (or manufactured for) RCM studies. Therefore, the products that are in use today in the climate modelling community are insufficiently validated and to a large extent outdated. Yet, the maps are used simultaneously within the RCM community.

Comprehensive validation of LULC input for RCMs with independent reference data following the good practice recommendations given by the remote sensing community is put in the center of attention of the present approach in order to address the various uncertainties that arise from the respective LULC input within an RCM simulation. The newly developed high-resolution LANDMATE PFT 2015 map provides a product that fills the gaps that were identified in chapter 1. Due to the transparency of the production process provided by Reinhart et al. (2021b), translatability of LANDMATE PFT into model-specific classifications is ensured. Further, the urgent need for transparent quality information is met. The accuracy assessment design that is developed for the validation of fractional, gridded map products produces reliable quality information in a transparent, comprehensible way to the user community.

With this approach we are the first to apply validation with ground truth data to a PFT map. The method used is easily adjustable to other fractional, gridded PFT maps as it is done within chapter 4.2 to the ESA-CCI PFT map. Through the validation, the uncertainty in PFT distribution that arises form the CWP is identified and the CWP can be modified, if necessary. The validation workflow can further be applied for the validation of other products outside Europe, assuming that ground truth data is available.

3. Publication overview

3.1 Publication I

Title	Comparison of ESA climate change initiative land cover to CORINE land cover over Eastern Europe and the Baltic States from a regional climate modeling perspective
Authors	Vanessa Reinhart, Cidália C. Fonte, Peter Hoffmann, Benjamin Bechtel, Diana Rechid, Jürgen Böhner
Journal	International Journal of Applied Earth Observation and Geoinformation
DOI	https://doi.org/10.1016/j.jag.2020.102221
Submitted	April 2020
Published	September 2020
Personal contribution	I raised the research interest, conceptualized the research approach, wrote the original draft and applied the methodology adopted from the research group of C.C. Fonte to the dataset of interest. I prepared the presentation and interpretation of results. Diana Rechid and Jürgen Böhner. supervised the study and revised the manuscript.

Abstract

High-quality land use and land cover (LULC) information is of crucial importance for the performance of regional climate models (RCMs), in particular at high spatial resolutions down to convection permitting scales below 4 km. Several satellite-based high-resolution products are currently available for implementation into RCMs. One of the most recent products is the European Space Agency Climate Change Initiative Land Cover (ESA CCI LC) dataset. While the ESA CCI LC has been assessed globally, an evaluation against regional, independent LULC datasets is necessary to

identify LULC inaccuracies in the respective region of interest and to give regional climate modelers estimates for the uncertainty in the land use forcing. In the present work the ESA CCI LC dataset is compared to the COoRdination and INformation on the Environment (CORINE) Land Cover (CLC). Agreement between the datasets is assessed by proportional area comparison (PAC). The resulting agreement measures are compared to the results of a majority approach (MA) to explore possible differences between the methods. Three timesteps of ESA CCI LC matching the timesteps of CLC are assessed to take a change in agreement over time into account. In addition to the quantification of agreement, spatial patterns of possible issues with ESA CCI LC are identified through utilization of geospatial information systems (GIS). Using the PAC, the agreement of ESA CCI LC with CLC is found to be˜76 % for the research area (RA). Although the agreement decreases slightly using the PAC, no substantial differences in agreement measures were found compared to the results of the MA. Dominant LULC categories agriculture and forest show an agreement of over 80 % with CLC. A few major issues were found for grassland, wetlands, settlements, and water bodies in the RA of which some might influence RCM performance if the dataset is implemented without adjustment. We highly recommend to apply the PAC to other regions in Europe and further globally to investigate if the found issues are also found elsewhere. The use of more independent regional and specified datasets for validation but also for possible improvement of the ESA CCI LC dataset is suggested.

3.2 Publication II

Title LANDMATE PFT land cover dataset for Europe 2015 (Version 1.0)

Authors Vanessa Reinhart, Peter Hoffmann, Diana Rechid

Publisher World Data Center for Climate (WDCC) at DKRZ

DOI https://doi.org/10.26050/WDCC/LM_PFT_LandCov_EUR2015_v1.0

Submitted April 2021

Published August 2021

Personal contribution I and Peter Hoffmann developed the cross-walking procedure and the corresponding cross-walking tables to generate LANDMATE PFT 2015. I revised the inclusion of climate data within the workflow. I prepared the LANDMATE PFT 2015 map for publication within the WDCC DKRZ database. I co-wrote the supplementary data description document together with Peter Hoffmann.

Summary

The LANDMATE PFT dataset provides a land cover map for Europe for the year 2015 in 0.1° (10km) and 0.018° (2km) resolution. The dataset is based on land cover data of the ESA Climate Change Initiative (ESA-CCI, native resolution: 300m) which is translated into 16 plant functional types (PFTs) and non-vegetated classes employing the cross-walking procedure introduced by Reinhart et al. (2021). The translation is done under consideration of the Holdridge Life Zones (HLZs), a system, that classifies

land areas based on bioclimatic properties. Through the HLZs, regional distinction of the individual PFT distribution can be achieved. The land cover information is given as fractions per grid cell where each fraction represents the area covered by the respective land cover within each grid cell (0-1). The dataset is available in two different horizontal resolutions, 0.1° (10km) and 0.018° (2km), whereby the land cover information is resampled using a fractional approach to achieve the desired resolution.

The LANDMATE PFT dataset was carefully developed and designed to meet the present and future requirements of regional climate models and is therefore recommended to be used for regional climate modeling over the European Continent. The LANDMATE PFT dataset (0.1° resolution) serves as basemap for the historical and future land use and land cover dataset LUCAS LUC developed by Hoffmann et al. (2021).

3.3 Publication III

Title	High-resolution land-use land-cover change data for regional climate modelling applications over Europe – Part 1: The plant functional type basemap for 2015
Authors	Vanessa Reinhart, Peter Hoffmann, Diana Rechid, Jürgen Böhner, Benjamin Bechtel
Journal	Earth System Science Data Discussions
DOI	https://doi.org/10.5194/essd-2021-251, in review, 2021
Submitted	July 2021
Published	*in discussion*
Personal contribution	I conceptualized the paper outline and objective with the support of all Co-authors. Together with Peter Hoffmann I developed the cross-walking procedure and the corresponding cross-walking tables. I developed the accuracy assessment design for the LANDMATE PFT map supported by Benjamin Bechtel. I conducted the accuracy assessment and the visualization and interpretation of results. I wrote the original draft of the manuscript, which was revised by all co-authors.

Abstract

The concept of plant functional types (PFTs) is shown to be beneficial in representing the complexity of plant characteristics in land use and climate change studies using regional climate models (RCMs). By representing land use and land cover (LULC) as functional traits, responses and effects of specific plant communities can be directly

coupled to the lowest atmospheric layers. To meet the requirements of RCMs for realistic LULC distribution, we developed a PFT dataset for Europe (LANDMATE PFT Version 1.0, Reinhart et al. (2021c)). The dataset is based on the high-resolution ESA-CCI land cover dataset and is further improved through the the additional use of climate information. Within the LANDMATE PFT dataset, satellite-based LULC information and climate data are combined to achieve the best possible representation of the diverse plant communities and their functions in the respective regional ecosystems while keeping the dataset most flexible for application in RCMs. Each LULC class of ESA-CCI is translated into PFT or PFT fractions including climate information by using the Holdridge Life Zone concept. Through the consideration of regional climate data, the resulting PFT map for Europe is regionally customized. A thorough evaluation of the LANDMATE PFT dataset is done using a comprehensive ground truth database over the European Continent. A suitable evaluation method has been developed and applied to assess the quality of the new PFT dataset. The assessment shows that the dominant LULC groups, cropland and woodland, are well represented within the dataset while uncertainties are found for some less represented LULC groups. The LANDMATE PFT dataset provides a realistic, high-resolution LULC distribution for implementation in RCMs and is used as basis for the LUCAS LUC dataset introduced in the companion paper by Hoffmann et al. (2021) which is available for use as LULC change input for RCM experiment setups focused on investigating LULC change impact.

4. Discussion

The original publications stand on their own but are closely connected. The discussion is structured along the initial research questions in section 2.1 including the reference to the respective publication. Research question 1 and 2 are discussed in section 4.1 and research question 3 and 4 are discussed in section 4.2.

4.1 Selection and validation of LULCC products for the use in RCMs

The selection of ESA-CCI LC as an answer to research question 1, followed extensive research on available LULCC data. The decision to build the LANDMATE PFT map on this ready-to-use product instead of using remotely sensed raw data followed pertinent reasons of practicability and the scope of this thesis. The most transparent approach would have been building the PFT dataset for RCMs on remotely sensed raw data. An obvious choice for a remote sensing dataset would have been the use of one of the oldest and longest satellite programs, the LANDSAT program (Woodcock et al., 2008; Wulder et al., 2019). Since the early 70s, the overlapping missions (LANDSAT 1-8, LANDSAT 9 since September 2021) cover the global land area with multiple fly overs per year capturing image quality of 30 m horizontal resolution. Although the multispectral images are freely available, the raw imagery requires careful processing in order to generate a classified LULCC product (ED Chaves et al., 2020; Leinenkugel et al., 2019). Beside the selection of suitable algorithms, supervised or unsupervised, the overall image availability is a factor that determines the quality of the final product (Prishchepov et al., 2012). Following the processing, validation is needed to determine the success of the processing chain. There are several regional approaches using customized algorithms to classify the LANDSAT imagery over the respective focus region in order to identify changes in landscape patterns (Cheţan et al., 2018). Furthermore, a pan-European map for 2015 based on Landsat 8 imagery processed using machine learning algorithms was published (Pflugmacher et al., 2019). Quality information for the map is given, however, the classification, which contains eleven LULC classes, is not diverse enough to meet the requirements of the RCM community. As the classification approaches show, the focus of the production of LULCC maps from remotely sensed raw data is necessarily laid on the development of high-performing classification algorithms, which need to be developed through elaborate procedures. However, the focus of this thesis is the identification of uncertainties within existing LULCC maps that are in use in the RCM community in order to improve the LULCC representation, which makes the development of classification algorithms to be clearly out of scope. With regard to the time and resource consuming procedure of preparing classified maps from LANDSAT data and the following design and implementation of a quality assessment, the approach was not followed further. Since the ESA-CCI dedicated multiple years into the production and global validation of their classified LC time series, it was decided to be the ideal baseline product for the production of a customized high-resolution PFT map for Europe. Nevertheless, comprehensive validation tailored to the respective purpose is incumbent on the user and is therefore tackled in the following research.

Research question 2 is focused on the ESA-CCI LC product itself and the LULCC suggested by the annual maps, where the hot-spot LULCC locations in Europe were found to be in Eastern Europe. While the ESA-CCI LC map for 2010 is validated on a global scale using the GLOBCOVER 2009 validation database (global accuracy is 70 % to over 74 %, depending on the assessment design (ESA, 2017)), an extensive validation on regional level was not done for Europe. In order to verify small scale LULCC, the comparison to independent LULC data for the respective region of interest is necessary and the quality figures of a coarse scale assessment cannot be transferred to small scale studies, especially when it comes to the validation of changes. Small scale changes that might be missed by a coarse scale assessment could lead to over- or rather underestimation of LULCC. With the proportional area comparison (PAC) we aim to overcome one major challenge in accuracy assessment of LULC maps, namely the modification of map resolution in order to fit the reference to the assessed map product.

A challenge that could not be tackled to a sufficient extent in this approach is the harmonization of classifications, that always bears the risk of changing accuracy measures. Assessed map and reference are mostly provided by different institutions and are created to serve different purposes. Therefore, LULC map classifications or precisely, class definitions may differ. The ESA-CCI LC maps are classified following the protocol of the UN LCCS classification system (Di Gregorio, 2005). According to the user manual the classification is based on automated classifiers that are arranged in a hierarchical structure. One of eight main LC types is defined as a first step followed by specification, where the set of classifiers is determined by the previously defined LC type. The final classification of ESA-CCI LC includes 37 LC classes including various mixed classes which is also an implication of the 300 m resolution. In a highly heterogeneous landscape like it is found in Europe, there is a high probability that a patch of 300x300 m does include more than one LC type. Therefore, the classes within the ESA-CCI LC legend, such as class 70 (Tree cover, needleleaved, evergreen, closed to open (>15 %)) may contain tree cover in a range between 15 % and 100 % where the actual proportion is not further defined.

The CLC database that is treated as reference in the approach by Reinhart et al. (2021a) is classified based on the interpretation of objects within the recorded images, precisely the interpretation of texture and patterns. This technique of classification requires regional expert knowledge, which is given through the decentralized approach of CLC (Büttner et al., 2004). The CLC datasets legend is available in three levels of detail, where level I includes the five main LC types, level II includes 15 subclasses and the most detailed level III includes 44 subclasses. The fine resolution of CLC (~100 m) makes the occurrence of mixed pixels less likely than for ESA-CCI LC. Nevertheless, the level III classification includes mixed classes but the proportions on LC types within a pixel are defined differently. For example, class 3.2.1 (Coniferous forest) is defined with a crown cover of at least 30 % within a pixel, which means the tree cover can range between 30 % and 100 %. The difference between the tree proportions in the two datasets might not be relevant when only one pixel is considered. When a continuous map is compared to another, the differences may increase to an extent that influences the overall LC type proportions.

The use of independent map data or ground truth reference data is necessary in order to verify the quality of a LULC map. However, when independent map products are not based on the same classification system, classification harmonization is inevitable. Associated with a modification of a map legend, including aggregation or grouping of individual classes are changes in the representation of LULC of the respective map. The harmonization of ESA-CCI LC with the CLC classification is only one example of this challenge. While in the present approach an established harmonization table was utilized, there are multiple concepts successfully tested, such as the one-to-many approach by Fritz et al. (2008), where a degree of overlap between classes is calculated based on the class descriptions of each product in the comparison. Another approach is the use of a Latent Dirichlet Allocation (LDA) model, where a newly created harmonized class is allocated to each pixel (Z. Li et al., 2021). For future work, the application of such advanced methods for comparing differently classified products should be considered for the validation of LULCC maps that are used in RCMs.

4.2 A validated PFT map for climate modelling

In order to answer research question 3, the method selected is a connection between previous approaches, where the ESA-CCI LC PFT translation is used as guideline for the new translation and the concept for the inclusion of climate data is adopted from Wilhelm et al. (2014). By the addition of climate data the LULC representation is adapted regionally.

The uncertainties arising from the translation of LULC maps into PFT maps was investigated by Hartley et al. (2017) where the cross-walking procedure is identified as a main source of uncertainty. The translation of a LULC map into a PFT map clearly bears the risk of changing the map accuracy. Therefore, validation of the final product should be key within the production process. The default ESA-CCI PFT translation was developed by Poulter et al. (2015) who relied on expert knowledge to prepare the cross-walking tables. However, detailed documentation on the production process of the CWTs as well as information on quality is not given. Instead, the ESA-CCI PFT map is compared to other available PFT maps in order to point out the differences between the maps in use.

In the approach by Reinhart et al. (2021b), the default CWT provided by ESA-CCI is revisited, modified and supplemented by high-resolution climate data. Nevertheless, uncertainties remain in the LANDMATE PFT map. In order to evaluate the success of the modification of the CWTs, the validation procedure that was developed for the validation of fractional, gridded LULC maps is also applied to the default ESA-CCI PFT map for 2015 and the results are compared to the spatial accuracy figures for LANDMATE PFT (Fig. 4.1).

The maps show the differences in agreement with the ground truth dataset between LANDMATE PFT 2015 and ESA-CCI PFT 2015, dependent of the respective count of evaluable points per grid cell. While for cropland and urban areas no large improvement is noticed, considerable improvement was achieved for the representation of woodland,

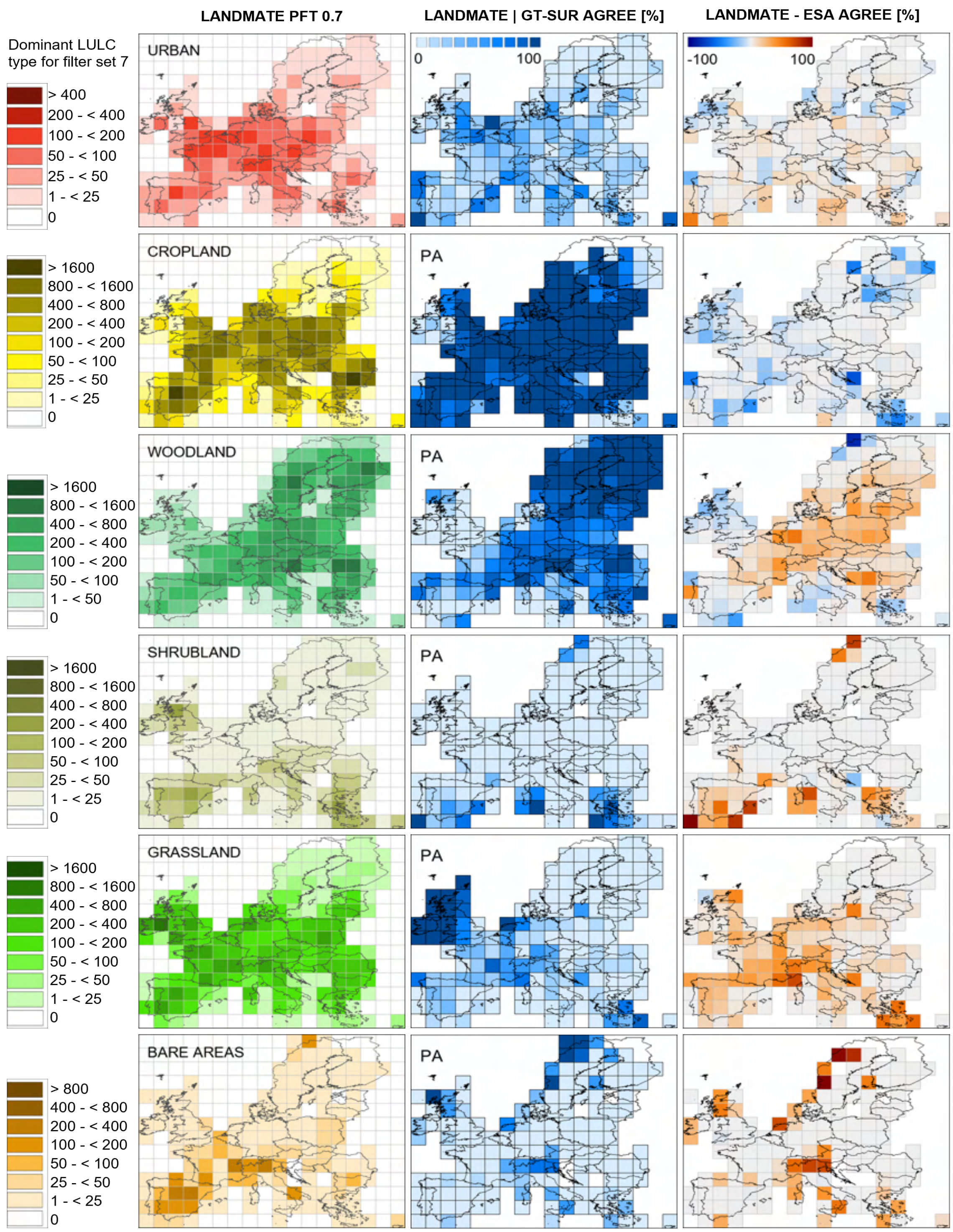


Figure 4.1: The LANDMATE PFT 2015 cell count where the dominant LULC types cover 70 % or more of a grid cell aggregated as count per 2.5° grid cell for improved visualization (left column), the producers accuracy (PA) for LANDMATE PFT 2015 compared to the ground truth database LUCAS ground truth survey (GT-SUR, expressed in percentage, middle column) and the difference between the PA of LANDMATE PFT 2015 and ESA-CCI PFT 2015 per 2.5° grid cell where a positive value means an advantageous LULC type representation by LANDMATE PFT 2015

shrubland and grassland. The bare areas representation has improved as well. Nevertheless, the grid cells with improved representation include a low number of bare area cells, which is why the results cannot not be weighted too heavily and might require further investigation. All further results derived by the validation of ESA-CCI PFT are prepared according to the workflow introduced in Reinhart et al. (2021b) and presented in Appendix B.

The validation of LANDMATE PFT using the LUCAS ground truth survey (GT-SUR, d'Andrimont et al., 2020) gives information on the uncertainty of the map for climate modellers that implement the map within RCM simulations. The approach followed in this thesis is designed to produce this uncertainty information especially for fractional LULC maps that are used in RCMs. The challenges that occur when conducting such an accuracy assessment are already addressed in section 4.1. While these challenges remain, the validation of fractional maps brings additional issues into the production of useful accuracy measures. With the selection of subsets within the LANDMATE PFT map, where the selection criterion is the lower threshold for minimum coverage of a LC type one map unit (grid cell), a range is determined for each accuracy measure, based on the minimum fraction size. The ranges for each LULC type are published as well as spatial distribution of accuracy and uncertainties. The accuracy measures are further useful when the PFT map is transferred into an RCM that is not able to represent subgrid fractions. In such RCMs, the fractional map is resampled with a majority approach. Precisely, only the dominant LULC type is represented in the model. Since the measures are calculated for different minimum coverage of the respective dominant LULC type (from 10-90 %) the user can directly derive areas of LULC uncertainty within the respective RCM.

However, one of the biggest challenges, the classification harmonization addressed in section 4.1, is still present for the quality assessment of LANDMATE PFT. The GT-SUR consists of nine main LULC types, while two LULC types are water and marine areas, which are not directly comparable to the LANDMATE PFT map. On the other hand, LANDMATE PFT contains the special vegetation types tundra and swamps, which are not comparable to the GT-SUR types. It was decided to focus the comparison on the main LULC types to make the most use of the extensive database GT-SUR. Further validation might not be achievable through the use of this database, but could be done using specialized datasets. For example, the validation of forest type distribution could not be included within the validation workflow. Within the LANDMATE PFT map, a distinction is made between six different forest types. This unique feature is possible through the inclusion of additional data, such as climate data and is highly important for RCM studies focused on risk assessment (Albert et al., 2017; Honkaniemi et al., 2020; Terrier et al., 2013) or vegetation change (Ahmed et al., 2021; Lehtonen et al., 2019; Morin et al., 2018) Although the distribution of forest types within the LANDMATE PFT map is supported by additional climate data, validation is crucially needed in order to identify possible uncertainties and to modify the map to approximate realistic LULC over Europe and needs to be addressed in future work.

ESA-CCI LC is available globally, which implies that the CWP used to create LANDMATE PFT 2015 can be used to produce the PFT map also for other regions, in case high-resolution climate data is available. However, this step is not trivial as the regional adaptation of LANDMATE PFT for Europe might not be working out for other continents, considering the highly diverse ecosystems and plant communities depending on climate and global location. The transfer of the CWP to outside of Europe was done using the global climate dataset CRU (Harris et al., 2014) but it was decided that the global data will not be published due to major inconsistencies for the other continents, when compared to other global PFT datasets (Reinhart et al., 2020). The CWP necessarily needs to be adjusted before application, depending on the continent of interest. The modification of the CWP in order to adjust the CWP to other regions than Europe is associated with thorough validation through comparison to regional ground truth data and local expert knowledge, which is possible for future work, but is beyond the scope of this thesis.

The fact that ESA-CCI does only provide maps for the past annual time steps and therefore, does not provide continuity for future LULCC projections, lead to the decision that the ESA-CCI LC time series was not used for the follow-up production of the PFT time series. The associated work by Hoffmann et al. (2021) relied on the well established LUH2 dataset to extend the LANDMATE PFT map for 2015 into the past and future [16]. In order to gain more detail to LULC type distribution and spatial resolution, additional datasets are implemented, instead of considering the high-resolution ESA-CCI LC annual maps. Future work should include a quality assessment for more time steps of the LUCAS LUC time series in order to validate the LULCC suggested by the dataset for the use in RCMs.

[16] LUCAS LUC historical land use and land cover change dataset (Version 1.0) & LUCAS LUC future land use and land cover change dataset (Version 1.0) available at https://cera-www.dkrz.de/WDCC/ui/cerasearch/project?acronym=LANDMATE

5. Conclusion

The ESA-CCI LC time series was found to be suitable as baseline product for the production of a PFT time series for the use in RCMs. The LULCC suggested by the ESA-CCI time series is validated by Proportional Area Comparison (PAC) to the CORINE Land Cover Inventory (CLC). LULCC within ESA-CCI LC was found to be reliable for the most common LULC types while uncertainties were found for the less common LULC types. A workflow was developed for the translation of the ESA-CCI LC map 2015 into PFTs. In order to compensate the uncertainties found in the previous assessment, an existing cross-walking procedure (CWP) was revisited and modified according to regional expert knowledge, scientific insights from literature and knowledge from preliminary studies. Further, additional data was added, including high-resolution climate data. In order to confirm the success of the modified CWP and the inclusion of additional data, a quality assessment workflow was developed, tailored to the challenge of validating gridded LULC maps with a fractional structure.

Applying the novel validation approach designed for fractional LULC data, it was demonstrated that the LANDMATE PFT map for 2015 is able to represent the assessed LULC types woodland and cropland very well while minor uncertainties are found for the other assessed LULC types. Limitations of the conceptual validation framework arise from the classification harmonization and the structural differences between LANDMATE PFT and the ground truth reference. While the structural differences could be overcome by the implementation of filters, the classification harmonization can be improved for future validation work. Nevertheless, the newly developed validation approach can be transferred to other PFT maps and regions. The implementation of LANDMATE PFT 2015 into RCMs is currently ongoing and is still is to be documented thoroughly.

The methods applied to answer the research questions that the present thesis is built on are tailored to fill the research gaps identified in chapter 1. Firstly, the LANDMATE PFT dataset for 2015 is a PFT dataset that fulfils the requirements of the regional climate modelling community. Through the involvement of high-resolution climate data it is possible to tailor the dataset to regional climate conditions. Precisely, the CWP can be adjusted to non-European regions to serve as LULC input for non-European domains in regional climate modelling. Further, the sophisticated validation procedure that was developed in the course of the creation of the dataset makes it possible to generate quality information for fractional LULC datasets, as they are used frequently as input for climate models. When ground truth data is available in the respective region of interest, the validation workflow can be applied to all regional PFT datasets before implementation into an RCM.

The LANDMATE PFT map for 2015 for Europe is originally developed to serve as baseline map for a LULCC time series that is used within the CORDEX Flagship Pilot Study Land Use Across Scales (CORDEX FPS LUCAS). The FPS LUCAS is dedicated to investigate the impact and feedback mechanisms of LULCC on regional climate, where it is crucial to use LULCC input that is thoroughly quality checked. In the course of this thesis the sensitivity to the impact of the quality of LULCC input in RCMs was raised within the FPS LUCAS and will be investigated in the upcoming project phases.

The present approach contributes valuable insights to the whole RCM community regarding the identification of uncertainties in RCM simulations caused by LULC input. The LANDMATE PFT dataset for 2015 is the first high-resolution PFT map for Europe that is thoroughly validated by comparison to an extensive ground truth database. The validation workflow can be used to investigate the quality of PFT maps, that are currently used. The generation of quality information of PFT maps can be majorly improved by inclusion of the validation workflow in the production process, instead of just comparing the respective PFT dataset to other products.

The added value of the LANDMATE PFT map even does extend beyond the RCM community with having direct impact on climate services. With the improvement of LULCC representation in RCMs the provision of climate information to stakeholders can be improved. Under the consideration of the overall purpose of RCMs, to provide high-resolution climate information to society, all components of an RCM need to be constantly challenged for reliability. With the LANDMATE PFT map for 2015, a LULCC component is introduced to the RCM community that provides validated, spatial quality information for the European continent, which represents a great added value in contrast to the other LULC products currently used.

Related publications

Hoffmann, P., **Reinhart, V**, Rechid, D., de Noblet-Ducoudré, N., Davin, E. L., Asmus, C., Bechtel, B., Böhner, J., Katragkou, E., AND Luyssaert, S.: High-resolution land-use land-cover change data for regional climate modelling applications over Europe – Part 2: Historical and future changes, Earth Syst. Sci. Data Discuss. [preprint],https://doi.org/10.5194/essd-2021-252,inreview,2021.

Personal contribution: I was part of the development of the workflow. I developed the cross-walking procedure together with Peter Hoffmann. I downloaded and prepared the ESA POULTER and the MODIS PFT dataset for comparison with LUCAS LUC. Further, I wrote the data description in the manuscript (i.e. Sect. 3.1.1 575 and Sect. 3.1.2). I visualized the LANDMATE PFTs and prepared tables A2 and A3. All authors reviewed the paper draft and contributed to the final paper.

Hoffmann, P., **Reinhart, V.**, Rechid, D. (2021). LUCAS LUC historical land use and land cover change dataset (Version 1.0). World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.26050/WDCC/LUC_hist_landCovChange_v1.0

&

Hoffmann, P., **Reinhart, V.**, Rechid, D. (2021). LUCAS LUC future land use and land cover change dataset (Version 1.0). World Data Center for Climate (WDCC) at DKRZ. https://doi.org/10.26050/WDCC/LUC_future_landCovChange_v1.0

Personal contribution: I and Peter Hoffmann developed the cross-walking pro-cedure and the corresponding cross-walking tables to generate the basemap for LUCAS LUC, LANDMATE PFT 2015. I revised the inclusion of climate data within the workflow. I co-wrote the supplementary data description document together with Peter Hoffmann.

Conference contributions

Reinhart, V., Hoffmann, P., Rechid, D., & Böhner, J. (2020, December). Uncertainties in plant functional type (PFT) products for use in regional climate models-assessing the sensitivity of global PFT time series to different input data and cross-walking procedures. In AGU Fall Meeting Abstracts (Vol. 2020, pp. GC018-07). https://agu.confex.com/agu/fm20/meetingapp.cgi/Paper/721156

Hoffmann, P., Rechid, D., Reinhart, V., Asmus, C., Davin, E., Katragkou, E., de Noblet-Ducoudré, N., Bechtel, B. & Böhner, J. (2020, December). Long-term high-resolution land-use/land-cover change dataset for regional climate modeling. In AGU Fall Meeting Abstracts (Vol. 2020, pp. GC007-0008). https://agu.confex.com/agu/fm20/meetingapp.cgi/Paper/709275

Reinhart, V., Hoffmann, P., Bechtel, B., Rechid, D. & Böhner, J. (2020, June). Accuracy assessment of ESA CCI LC over Eastern Europe and the Baltic States from a climate modelling perspective–identification of spatial inaccuracy patterns and misclassification issues using a fuzzy comparison method. Earth system changes and Baltic Sea coasts, 19(7), 145. https://archive.baltic.earth/hel2020/material/3rd_BalticEarth_Conference_Proceedings.pdf

Reinhart, V., Böhner, J., Rechid, D., Bechtel, B. & Hoffmann, P. (2019, May). How is Europe's Landscape Changing? Analysis of ESA-CCI Land Cover for its potential use in high-resolution climate modelling. ESA Living Planet Symposium 2019. https://drive.google.com/file/d/1OILXHvP5Q0uDvinKwK6bNIj4WqIAX-jz/view?usp=sharing

Bibliography

Adinolfi, M., Raffa, M., Reder, A., & Mercogliano, P. (2021). Evaluation and expected changes of summer precipitation at convection permitting scale with cosmo-clm over alpine space. *Atmosphere*, *12*(1), 54.

Ahmed, H. et al. (2021). *Characterization of spring thaw for different forest types in the southern boreal forest under current and future climate* (Doctoral dissertation). University of Saskatchewan.

Albert, M., Nagel, R.-V., Nuske, R. S., Sutmöller, J., & Spellmann, H. (2017). Tree species selection in the face of drought risk—uncertainty in forest planning. *Forests*, *8*(10), 363.

Arino, O., Ramos Perez, J. J., Kalogirou, V., Bontemps, S., Defourny, P., & Van Bogaert, E. (2012). Global land cover map for 2009 (globcover 2009).

Bartholome, E., & Belward, A. S. (2005). Glc2000: A new approach to global land cover mapping from earth observation data. *International Journal of Remote Sensing*, *26*(9), 1959–1977.

Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., & Defourny, P. (2012a). Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences*, *9*(6), 2145–2157. https://doi.org/10.5194/bg-9-2145-2012

Bontemps, S., Herold, M., Kooistra, L., Groenestijn, A. v., Hartley, A., Arino, O., Moreau, I., & Defourny, P. (2012b). Revisiting land cover observation to address the needs of the climate modeling community. *Biogeosciences*, *9*(6), 2145–2157.

Büttner, G., Feranec, J., Jaffrain, G., Mari, L., Maucha, G., & Soukup, T. (2004). The corine land cover 2000 project. *EARSeL eProceedings*, *3*(3), 331–346.

Cheţan, M. A., Dornik, A., & Urdea, P. (2018). Analysis of recent changes in natural habitat types in the apuseni mountains (romania), using multi-temporal landsat satellite imagery (1986–2015). *Applied geography*, *97*, 161–175.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Joebges, C., Lemoine, G., et al. (2020). Harmonised lucas in-situ land cover and use database for field surveys from 2006 to 2018 in the european union. *Scientific Data*, *7*(1), 1–15.

Davin, E. L., Rechid, D., Breil, M., Cardoso, R. M., Coppola, E., Hoffmann, P., Jach, L. L., Katragkou, E., de Noblet-Ducoudré, N., Radtke, K., et al. (2020). Biogeophysical impacts of forestation in europe: First results from the lucas (land use and climate across scales) regional climate model intercomparison. *Earth System Dynamics*, *11*(1), 183–200.

De Meij, A., & Vinuesa, J. (2014). Impact of srtm and corine land cover data on meteorological parameters using wrf. *Atmospheric Research*, *143*, 351–370.

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L., & Leroy, M. (2006). Globcover: A 300 m global land cover product for 2005 using envisat meris time series. *Proceedings of ISPRS Commission VII Mid-Term Symposium: Remote Sensing: from Pixels to Processes, Enschede (NL)*, 8–11.

de Noblet-Ducoudré, N., Boisier, J.-P., Pitman, A., Bonan, G., Brovkin, V., Cruz, F., Delire, C., Gayler, V., Van den Hurk, B., Lawrence, P., et al. (2012). Determining robust impacts of land-use-induced land cover changes on surface climate over north america and eurasia: Results from the first set of lucid experiments. *Journal of Climate*, *25*(9), 3261–3281.

Di Gregorio, A. (2005). *Land cover classification system: Classification concepts and user manual: Lccs* (Vol. 2). Food & Agriculture Org.

Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M., Mellab, K., & Saint, G. (2014). Proba-v mission for global vegetation monitoring: Standard products and image quality. *International Journal of Remote Sensing*, *35*(7), 2589–2614.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., et al. (2012). The global monitoring for environment and security (gmes) sentinel-3 mission. *Remote Sensing of Environment*, *120*, 37–57.

ED Chaves, M., CA Picoli, M., & D Sanches, I. (2020). Recent applications of landsat 8/oli and sentinel-2/msi for land use and land cover mapping: A systematic review. *Remote Sensing*, *12*(18), 3062.

ESA. (2017). *Land cover cci product user guide version 2* (tech. rep.). European Space Agency. maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf

ESA, E. A. P. (2002). Available online: Http://envisat. esa. int/handbooks/asar. *CNTR. html (accessed on 27 January 2020)*.

Fonte, C., See, L., Laso-Bayas, J., Lesiv, M., & Fritz, S. (2020). Assessing the accuracy of land use land cover (lulc) maps using class proportions in the reference data. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, *5*(3).

Foody, G. M. (2015). Valuing map validation: The need for rigorous land cover map accuracy assessment in economic valuations of ecosystem services. *Ecological Economics*, *111*, 23–28.

Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. (2010). Modis collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote sensing of Environment*, *114*(1), 168–182.

Fritz, S., & See, L. (2008). Identifying and quantifying uncertainty and spatial disagreement in the comparison of global land cover for different applications. *Global Change Biology*, *14*(5), 1057–1075.

Giri, C., Zhu, Z., & Reed, B. (2005). A comparative analysis of the global land cover 2000 and modis land cover data sets. *Remote sensing of environment*, *94*(1), 123–132.

Harris, I., Jones, P. D., Osborn, T. J., & Lister, D. H. (2014). Updated high-resolution grids of monthly climatic observations–the cru ts3. 10 dataset. *International journal of climatology*, *34*(3), 623–642.

Hartley, A., MacBean, N., Georgievski, G., & Bontemps, S. (2017). Uncertainty in plant functional type distributions and its impact on land surface models. *Remote Sensing of Environment*, *203*, 71–89.

Hastings, D. A., & Emery, W. J. (1992). The advanced very high resolution radiometer (avhrr)-a brief reference guide. *Photogrammetric Engineering and Remote Sensing*, *58*(8), 1183–1188.

Herold, M., Mayaux, P., Woodcock, C., Baccini, A., & Schmullius, C. (2008). Some challenges in global land cover mapping: An assessment of agreement and accuracy in existing 1 km datasets. *Remote Sensing of Environment*, *112*(5), 2538–2556.

Hoffmann, P., Rechid, D., de Noblet-Ducoudré, N., Davin, E. L., Asmus, C., Bechtel, B., Böhner, J., Katragkou, E., & Luyssaert, S. (2021). High-resolution land-use land-cover change data for regional climate modelling applications over europe–part 2: Historical and future changes. *Earth System Science Data Discussions*, 1–43.

Honkaniemi, J., Rammer, W., & Seidl, R. (2020). Norway spruce at the trailing edge: The effect of landscape configuration and composition on climate resilience. *Landscape ecology*, 1–16.

Hua, T., Zhao, W., Liu, Y., Wang, S., & Yang, S. (2018). Spatial consistency assessments for global land-cover datasets: A comparison among glc2000, cci lc, mcd12, globcover and glcnmo. *Remote Sensing*, *10*(11), 1846.

Jiménez-Esteve, B., Udina, M., Soler, M. R., Pepin, N., & Miró, J. R. (2018). Land use and topography influence in a complex terrain area: A high resolution mesoscale modelling study over the eastern pyrenees using the wrf model. *Atmospheric Research*, *202*, 49–62.

Karvonen, V., Ribard, C., Sädekoski, N., Tyystjärvi, V., & Muukkonen, P. (2018). Comparing esa land cover data with higher resolution national datasets. *Creating, managing, and analysing geospatial data and databases in geographical themes*, 26–45.

Koubodana, D., Diekkrüger, B., Näschen, K., Adounkpe, J., & Atchonouglo, K. (2019). Impact of the accuracy of land cover data sets on the accuracy of land cover change scenarios in the mono river basin, togo, west africa. *International*

Journal of Advanced Remote Sensing and GIS, *8*(1), 3073–3095. https://cloudpublications.org/journals/index.php/RemoteSensing/article/view/461

Lehtonen, I., Venäläinen, A., Kämäräinen, M., Asikainen, A., Laitila, J., Anttila, P., & Peltola, H. (2019). Projected decrease in wintertime bearing capacity on different forest and soil types in finland under a warming climate. *Hydrology and Earth System Sciences*, *23*(3), 1611–1631.

Leinenkugel, P., Deck, R., Huth, J., Ottinger, M., & Mack, B. (2019). The potential of open geodata for automated large-scale land use and land cover classification. *Remote Sensing*, *11*(19), 2249.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A., & Peng, S. (2018). Gross and net land cover changes in the main plant functional types derived from the annual esa cci land cover maps (1992–2015). *Earth System Science Data*, *10*(1), 219–234.

Li, Z., White, J. C., Wulder, M. A., Hermosilla, T., Davidson, A. M., & Comber, A. J. (2021). Land cover harmonization using latent dirichlet allocation. *International Journal of Geographical Information Science*, *35*(2), 348–374.

Liang, D., Zuo, Y., Huang, L., Zhao, J., Teng, L., & Yang, F. (2015). Evaluation of the consistency of modis land cover product (mcd12q1) based on chinese 30 m globeland30 datasets: A case study in anhui province, china. *ISPRS International Journal of Geo-Information*, *4*(4), 2519–2541.

Liu, X., Yu, L., Si, Y., Zhang, C., Lu, H., Yu, C., & Gong, P. (2018). Identifying patterns and hotspots of global land cover transitions using the esa cci land cover dataset. *Remote Sensing Letters*, *9*(10), 972–981.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., & Merchant, J. W. (2000). Development of a global land cover characteristics database and igbp discover from 1 km avhrr data. *International Journal of Remote Sensing*, *21*(6-7), 1303–1330.

Maisongrande, P., Duchemin, B., & Dedieu, G. (2004). Vegetation/spot: An operational mission for the earth monitoring; presentation of new standard products. *International Journal of Remote Sensing*, *25*(1), 9–14.

Majasalmi, T., & Rautiainen, M. (2021). Representation of tree cover in global land cover products: Finland as a case study area. *Environmental Monitoring and Assessment*, *193*(3), 1–19.

McCallum, I., Obersteiner, M., Nilsson, S., & Shvidenko, A. (2006). A spatial comparison of four satellite derived 1 km global land cover datasets. *International Journal of Applied Earth Observation and Geoinformation*, *8*(4), 246–255.

Morin, X., Fahse, L., Jactel, H., Scherer-Lorenzen, M., Garcıa-Valdés, R., & Bugmann, H. (2018). Long-term response of forest productivity to climate change is mostly driven by change in tree species composition. *Scientific Reports*, *8*(1), 1–12.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42–57.

Pérez-Hoyos, A., García-Haro, F., & San-Miguel-Ayanz, J. (2012). Conventional and fuzzy comparisons of large scale land cover products: Application to corine, glc2000, modis and globcover in europe. *ISPRS Journal of Photogrammetry and Remote Sensing*, *74*, 185–201. https://doi.org/https://doi.org/10.1016/j.isprsjprs.2012.09.006

Perugini, L., Caporaso, L., Marconi, S., Cescatti, A., Quesada, B., de Noblet-Ducoudre, N., House, J. I., & Arneth, A. (2017). Biophysical effects on temperature and precipitation due to land cover change. *Environmental Research Letters*, *12*(5), 053002.

Pflugmacher, D., Rabe, A., Peters, M., & Hostert, P. (2019). Mapping pan-european land cover using landsat spectral-temporal metrics and the european lucas survey. *Remote sensing of environment*, *221*, 583–595.

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S., Boettcher, M., Brockmann, C., Defourny, P., et al. (2015). Plant functional type classification for earth system models: Results from the european space agency's land cover climate change initiative. *Geoscientific Model Development*, *8*, 2315–2328.

Prishchepov, A. V., Radeloff, V. C., Dubinin, M., & Alcantara, C. (2012). The effect of landsat etm/etm+ image acquisition dates on the detection of agricultural land abandonment in eastern europe. *Remote Sensing of Environment*, *126*, 195–209.

Reinhart, V., Fonte, C. C., Hoffmann, P., Bechtel, B., Rechid, D., & Böhner, J. (2021a). Comparison of esa climate change initiative land cover to corine land cover over eastern europe and the baltic states from a regional climate modeling perspective. *International Journal of Applied Earth Observation and Geoinformation*, *94*, 102221.

Reinhart, V., Hoffmann, P., Rechid, D., Böhner, J., & Bechtel, B. (2021b). High-resolution land-use land-cover change data for regional climate modelling applications over europe–part 1: The plant functional type basemap for 2015. *Earth System Science Data Discussions*, 1–54.

Reinhart, V., Hoffmann, P., & Rechid, D. (2021c). Landmate pft land cover dataset for europe 2015 (version 1.0). https://doi.org/10.26050/WDCC/LM_PFT_LandCov_EUR2015_v1.0

Reinhart, V., Hoffmann, P., Rechid, D., & Böhner, J. (2020). Uncertainties in plant functional type (pft) products for use in regional climate models-assessing the sensitivity of global pft time series to different input data and cross-walking procedures. *AGU Fall Meeting Abstracts*, *2020*, GC018–07.

Samasse, K., Hanan, N. P., Tappan, G., & Diallo, Y. (2018). Assessing cropland area in west africa for agricultural yield analysis. *Remote Sensing*, *10*(11), 1785.

Santos-Alamillos, F., Pozo-Vázquez, D., Ruiz-Arias, J., & Tovar-Pescador, J. (2015). Influence of land-use misrepresentation on the accuracy of wrf wind estimates: Evaluation of glcc and corine land-use maps in southern spain. *Atmospheric Research*, *157*, 17–28.

Schicker, I., Arias, D. A., & Seibert, P. (2016). Influences of updated land-use datasets on wrf simulations for two austrian regions. *Meteorology and Atmospheric Physics*, *128*(3), 279–301.

Sertel, E., Robock, A., & Ormeci, C. (2010). Impacts of land cover data quality on regional climate simulations. *International Journal of Climatology*, *30*(13), 1942–1953.

Stehman, S. V., & Foody, G. M. (2019). Key issues in rigorous accuracy assessment of land cover products. *Remote Sensing of Environment*, *231*, 111199.

Strandberg, G., & Kjellström, E. (2019). Climate impacts from afforestation and deforestation in europe. *Earth Interactions*, *23*(1), 1–27.

Sulla-Menashe, D., & Friedl, M. A. (2018). User guide to collection 6 modis land cover (mcd12q1 and mcd12c1) product. *USGS: Reston, VA, USA*, 1–18.

Terrier, A., Girardin, M. P., Périé, C., Legendre, P., & Bergeron, Y. (2013). Potential changes in forest composition could reduce impacts of climate change on boreal wildfires. *Ecological Applications*, *23*(1), 21–35.

Tölle, M. H., & Churiulin, E. (2021). Sensitivity of convection-permitting regional climate simulations to changes in land cover input data: Role of land surface characteristics for temperature and climate extremes. *Frontiers in Earth Science*, 954.

Törmä, M., Markkanen, T., Hatunen, S., Härmä, P., Mattila, O.-P., & Arslan, A. N. (2015). Assessment of land-cover data for land-surface modelling in regional climate studies.

Vilar, L., Garrido, J., Echavarrıa, P., Martinez-Vega, J., & Martın, M. P. (2019). Comparative analysis of corine and climate change initiative land cover maps in europe: Implications for wildfire occurrence estimation at regional and local scales. *International Journal of Applied Earth Observation and Geoinformation*, *78*, 102–117.

Wilhelm, C., Rechid, D., & Jacob, D. (2014). Interactive coupling of regional atmosphere with biosphere in the new generation regional climate system model remo-imove. *Geoscientific Model Development*, *7*(3), 1093–1114.

Woodcock, C. E., Allen, R., Anderson, M., Belward, A., Bindschadler, R., Cohen, W., Gao, F., Goward, S. N., Helder, D., Helmer, E., et al. (2008). Free access to landsat imagery. *SCIENCE VOL 320: 1011*.

Wulder, M. A., Loveland, T. R., Roy, D. P., Crawford, C. J., Masek, J. G., Woodcock, C. E., Allen, R. G., Anderson, M. C., Belward, A. S., Cohen, W. B., et al. (2019). Current status of landsat program, science, and applications. *Remote sensing of environment*, *225*, 127–147.

Yang, Y., Xiao, P., Feng, X., & Li, H. (2017). Accuracy assessment of seven global land cover datasets over china. *ISPRS Journal of Photogrammetry and Remote Sensing*, *125*, 156–173.

A. Appendix: Original publications

A.0.1 Publication I

International Journal of Applied Earth Observations and Geoinformation 94 (2021) 102221

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Comparison of ESA climate change initiative land cover to CORINE land cover over Eastern Europe and the Baltic States from a regional climate modeling perspective

V. Reinhart[a,b,*], C.C. Fonte[c], P. Hoffmann[a,b], B. Bechtel[d], D. Rechid[a], J. Boehner[b]

[a] Helmholtz Zentrum Geesthacht, Climate Service Center Germany (GERICS), Germany
[b] Universität Hamburg, Institute of Geography, Germany
[c] University of Coimbra, Portugal
[d] Ruhr-Universität Bochum, Germany

ARTICLE INFO

Keywords:
Land Use
Land Cover
Quality Assessment
Reference Data
Proportional Area Comparison
Regional Climate Modelling
Eastern Europe

ABSTRACT

High-quality land use and land cover (LULC) information is of crucial importance for the performance of regional climate models (RCMs), in particular at high spatial resolutions down to convection permitting scales below 4 km. Several satellite-based high-resolution products are currently available for implementation into RCMs. One of the most recent products is the European Space Agency Climate Change Initiative Land Cover (ESA CCI LC) dataset. While the ESA CCI LC has been assessed globally, an evaluation against regional, independent LULC datasets is necessary to identify LULC inaccuracies in the respective region of interest and to give regional climate modelers estimates for the uncertainty in the land use forcing. In the present work the ESA CCI LC dataset is compared to the COoRdination and INformation on the Environment (CORINE) Land Cover (CLC). Agreement between the datasets is assessed by proportional area comparison (PAC). The resulting agreement measures are compared to the results of a majority approach (MA) to explore possible differences between the methods. Three timesteps of ESA CCI LC matching the timesteps of CLC are assessed to take a change in agreement over time into account. In addition to the quantification of agreement, spatial patterns of possible issues with ESA CCI LC are identified through utilization of geospatial information systems (GIS). Using the PAC, the agreement of ESA CCI LC with CLC is found to be ~76 % for the research area (RA). Although the agreement decreases slightly using the PAC, no substantial differences in agreement measures were found compared to the results of the MA. Dominant LULC categories agriculture and forest show an agreement of over 80 % with CLC. A few major issues were found for grassland, wetlands, settlements, and water bodies in the RA of which some might influence RCM performance if the dataset is implemented without adjustment. We highly recommend to apply the PAC to other regions in Europe and further globally to investigate if the found issues are also found elsewhere. The use of more independent regional and specified datasets for validation but also for possible improvement of the ESA CCI LC dataset is suggested.

1. Introduction

Production of high-quality and high-resolution land use and land cover (LULC) information has received increased attention in the last decades due to the importance of land cover representation for numerous fields of research. One important application is in regional climate modelling, which is moving towards higher resolution and therefore needs high-quality land cover information with high spatial and temporal resolution and coverage. For regional climate models (RCMs) a realistic LULC representation is crucial to realistically model subsurface and near-surface energy and moisture fluxes as well as to investigate feedback mechanisms and coupling effects between LULC and regional climate (Chu et al., 2011; Verburg et al., 2011; Houghton et al., 2012; Bontemps et al., 2013; Brovkin et al., 2013; Davin et al., 2019; Georgievski and Hagemann, 2019). Several RCM studies focusing on the quantification of uncertainties in near-surface climate parameters caused by LULC showed the benefits of using more precise LULC information in RCMs (Gao et al., 2015; Santos-Alamillos et al., 2015; Sertel et al., 2010).

Continuous LULC information is nowadays mostly produced by a

* Corresponding author at: Helmholtz Zentrum Geesthacht, Climate Service Center Germany (GERICS), Germany.

https://doi.org/10.1016/j.jag.2020.102221
Received 16 April 2020; Received in revised form 10 August 2020; Accepted 17 August 2020
Available online 9 September 2020

combination of manual and automatic interpretation of satellite imagery with inclusion of existing ground truth data and regional or country based LULC information (Loveland and Belward, 1997; Bontemps et al., 2011). For the reconstruction of past LULC and the projection of future LULC, supplementary data such as population development and distribution or changes in climate are used as proxies to assimilate LULC development with a model approach (Pongratz et al., 2008; Hurtt et al., 2011). Historical "ground truth" documents such as cadastral maps are an important input source or a validation instrument when reconstructing historical LULC with a model approach (Ramankutty and Foley, 1999; Petit and Lambin, 2002; Fuchs et al., 2013).

Uncertainties in a final LULC dataset can arise from various well-known sources of error during the development process. Different classification procedures, atmospheric disturbances as well as changing satellite sensors and algorithms can contribute to uncertainties between datasets but also within multiannual datasets (Castilla and Hay, 2007; Verburg et al., 2011; Fuchs et al., 2013). Therefore, throughout the development process of a LULC dataset, a comprehensive assessment is required to gain information about the quality of the product. However, there is no uniformly applied method and the procedure itself is widely discussed (Foody, 2002; Olofsson et al., 2014; Sarmento et al., 2015).

One of the most recent and detailed LULC products is the European Space Agency Climate Change Initiative Land Cover (ESA CCI LC) dataset, a continuous global dataset with 23 annual time steps (1992–2015) at 300 m grid resolution (ESA, 2017). Previous work found the ESA CCI LC epoch time steps (2000, 2005, 2010 and 2015) to be relatively accurate on a global scale (75.1 % for 2015, (Achard et al., 2017). Regional quality assessments gave consistent results when comparing ESA CCI LC to other high-resolution LC products in the investigated regions respectively (Pérez-Hoyos et al., 2017; Yang et al., 2017; Samasse et al., 2018; Koubodana et al., 2019). Hua et al. (2018) indicated a high consistency with other global state of the art land cover datasets for ESA CCI LC over Europe (over 60 % agreement with GLC2000 and GLOBCOVER globally and slightly higher when only Europe was compared). Due to its continuous global and annual availability and the detail in LULC description ESA CCI LC is most promising to be implemented in RCMs for various regional domains. However, a comparison of ESA CCI LC with independent reference data for Eastern Europe on a regional scale is missing.

In this study, a detailed comparison of the global ESA CCI LC product over Eastern Europe and the Baltic States with the COoRdination and INformation on the Environment (CORINE) Land Cover (CLC) dataset is carried out. We consider CLC as reference, since this dataset is more detailed, has a higher resolution and is independent from ESA CCI LC. Further, the quality of CLC is well known (Jaffrain et al., 2017). However, due to the different characteristics of the ESA CCI LC and the reference data, resampling and nomenclature harmonization techniques are necessary to adjust data sets of different resolution and classification for comparison. As a consequence of the application of these techniques at least one of the products is modified majorly during the process which can bias agreement measures (Foody, 2002; Tchuenté et al., 2011; Yang et al., 2017).

When applying a majority resampling approach (MA) to adjust the spatial resolution of two or more products, LULC class areas are changed. However, a recent approach in the region of Coimbra (Portugal) showed the bias in agreement measures due to use of the majority resampling is considerably high (Fonte et al., 2020). Against this background, an alternative approach is required, to reduce the bias due to resampling of LULC products and to improve the information value of LULC map comparisons.

The present work uses an innovative approach to assess the quality of high resolution LULC products in a way that is not restrained by the static grid structure using proportional area comparison (PAC) (Sarmento et al., 2015). By applying PAC, the quantity of correctly classified cells is transformed to a quantity of area that is correctly classified and that is independent from the grid structure. The method can be applied to compare gridded datasets with different resolutions (Fritz et al., 2010, 2011). In addition, the method can provide a spatial pattern of inaccuracies for the individual LULC classes.

In the present work, the method is applied to compare ESA CCI LC to CLC for all time steps that were available for both datasets respectively. In section 2, methods and data used in this study as well as the classification harmonization method are described. The agreement measures for the PAC and the MA for all time steps respectively, followed by maps of spatial disagreement patterns of the LULC categories are presented in section 3. Ways to deal with the identified issues in the individual categories and their possible implications for regional climate modelling are discussed in section 4. Finally, section 5 closes with conclusions and an outlook for further research.

2. Methods and data

2.1. ESA CCI land cover

ESA CCI LC is a continuous global land cover product with 300 m grid resolution. The product has been available online since 2017 and was developed over nine years by the ESA Climate Change Initiative (CCI) program. It provides global annual LC maps for 23 consecutive years from 1992 to 2015 (ESA, 2017). ESA CCI LC is a combination product of global surface reflectance from different satellite missions (Table 1). One of the major purposes of ESA CCI LC was to create a land cover product that meets the requirements of the climate modeling community (Li et al., 2017).

Validation is achieved through use of a dedicated tool provided by ESA and Google Earth images as background data. (Achard et al., 2017). Globally regularly distributed two-stage stratified random sampling including primary and secondary point sample units are validated by a network of experts for the respective region. Satellite based Google Earth data covering the respective time period is utilized as reference and validation is still ongoing. When validated using the GlobCover 2009 validation database, on which the validation tool is based, it is found that the overall accuracy for the ESA CCI LC 2015 map is 75.4 %. (Achard et al., 2017). Global consistency with other existing high resolution satellite based global land cover datasets was found to be relatively high for the European Continent (Hua et al., 2018). In addition, the ESA CCI LC is compared to existing, validated LULC products over the African Continent (Koubodana et al., 2019; Pérez-Hoyos et al., 2017; Samasse et al., 2018) and China (Yang et al., 2017) but up to now, there are no comprehensive assessment activities for ESA CCI LC over Central and Eastern Europe published. Extension of the ESA CCI LC map series until 2018 was provided in October 2019 (ESA, 2019).

Table 1
ESA CCI Product information (ESA, 2017).

Time period	Satellite products
Baseline Production 2003–2012	MERIS FR/RR global SR composites
1992–1999	Baseline 10-year global map AVHRR global SR composites for back-dating baseline
1999–2013	Baseline 10-year global map SPOT-VGT global SR composites for up and back-dating the baseline MERIS FR global SR composites to delineate the identified changes at 300 m spatial resolution PROBA-V global SR composites at 300 m for year 2013 to delineate the identified changes at 300 m spatial resolution
2013–2015	Baseline 10-year global map PROBA-V global SR composites at 1 km for years 2014 and 2015 for updating the baseline PROBA-V time series at 300 m for 2014 and 2015 to delineate the identified changes at 300 m spatial resolution

2.2. CORINE land cover (CLC)

The CLC database was initiated by a European program and includes land cover information for all EU member states in 1985 (Heymann, 1994). The first dataset from 1990 therefore covers 27 countries, while the most recent CLC map from 2018 covers 39 countries (GISAT, 2019).

The CLC development process includes satellite image interpretation (LANDSAT, SPOT, TM and MSS) and regional land cover information such as aerial photography, local knowledge and statistics. Table 2 shows the technical specifications for each available CLC dataset (Copernicus Land Monitoring Service., 2020).

Validation of CLC 2000 was done by reinterpretation based on IMAGE2000 data and by comparison to LUCAS LULC data. Reliability of CLC2000 was found to be 87.0 ± 0.8 % and agreement with LUCAS LULC data to be 74.8 ± 0.6 % respectively (Büttner and Maucha, 2006).

For CLC2006, no individual validation was done but a change reliability study from 2000 to 2006 was carried out (Büttner et al., 2011). The changes were found to be small (1.25 % of total CLC area) and it was concluded, that for CLC 2006 a similar accuracy can be expected. The CLC change product (2000–2006) was validated using stratified random sampling and including a weighted proportion of all occurring change types. Accuracy of changes was found to be 87.8 ± 3.3 % which confirms the assumption that CLC 2006 accuracy is similar to the CLC 2000 accuracy.

For CLC 2012, validation is done by evaluation of more than 25,000 sampling locations which were evaluated by experts (Jaffrain et al., 2017). Overall accuracy was found to be 85 %.

CORINE CLC is considered as one of the most consistent and most carefully prepared land cover product for Europe. Nevertheless, the product is only available for a few countries in Europe and few time steps. Therefore, CLC might be rather unsuitable for the use in high resolution RCMs investigating LULC change induced feedback mechanisms over continental scale domains. Yet, CLC provides a valuable source of high resolution LULC information. This information can be used to compare coarser, global LULC products over these countries to investigate their quality in a comparative assessment.

Its availability for three timesteps of the ESA CCI LC time series (i.e. 2000, 2006 and 2012) makes CLC a most valuable product for validating ESA CCI LC. CLC is classified into 3 levels with 44 land cover and land use classes on level 3 (Heymann, 1994).

2.3. Dataset harmonization

In land cover comparison, agreement measures depend on the semantic resolution of the chosen land cover typology where a lower number of classes is resulting in higher agreement/ accuracy (Bechtel et al., 2019). In order to avoid this issue, we decided to use an established harmonization method. Harmonization of classifications is done following Vilar et al. (2019) who provided a robust categorization method for both, ESA CCI LC and CLC to eight LULC categories in total (Table 3).

Overall, modification of utilized datasets was kept to a minimum leaving the resolution of both datasets unchanged.

ESA CCI LC and CLC are available in different projections. Therefore, the projection of the ESA CCI LC maps was transformed to fit the CLC ETRS89 Lambert Azimuthal Equal Area (LAEA). The projection is suitable for all approaches where true area representation is required and is further suited to the research area. Since the map product classification consists of discrete categories on a nominal scale, a nearest neighbor resampling strategy was applied. Both datasets were clipped to the extent of the research area in Eastern Europe. In order to account for geoprecision of the used data, the offset between datasets was looked into by comparing certain landmarks like coastlines and rectangular features. The offset was found not to exceed 20 m, therefore the geoprecision was found to be sufficient for the present analysis.

Table 3
Classification harmonization of ESA CCI LC and CLC[a].

	Category	ESA CCI LC	CLC
1	Agriculture	10, 11, 12, 20, 30, 40	2
2	Forest	50, 60, 61, 62, 70, 71, 72, 80, 81, 82, 90, 100, 160, 170	3.1
3	Grassland	110, 130	3.2.1
4	Wetland	180	4
5	Settlement	190	1
6	Shrubland	120, 121, 122	3.2.2, 3.2.3, 3.2.4
7	Sparse vegetation, bare areas, permanent snow and ice	140, 150, 152, 153, 200, 201, 202, 220	3.3
8	Water bodies	210	5

[a] Nomenclature can be found at http://maps.elie.ucl.ac.be/CCI/viewer/download/CCI-LC_Maps_Legend.pdf (ESA CCI LC) and https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/html (CLC).

2.4. Methodology

In the present study two LC data comparison approaches are carried out, a PAC and a comparison method using only the majority CLC class per ESA CCI LC pixel. For both approaches, overall accuracy (OA), user's accuracy (UA) and producer's accuracy (PA) are evaluated for all assessed timesteps and countries. In the following assessment the measures OA, UA and PA are referred to as overall agreement (OA = OA), precision (PR = UA), recall (RE = PA), since they are used as comparative measures and not as accuracy measures. Spatial comparison of ESA CCI LC and CLC is carried out using the PAC maps. All calculations were performed using SAGA GIS (Conrad et al., 2015).

In the PAC, the proportion of each CLC category per ESA CCI LC pixel is counted and added. Table 4 shows the calculated area of three CLC classes for the proportional area method (Ref_{prop}) and for the majority method (Ref_{maj}) for four example ESA CCI LC pixels (Fig. 1).

The area proportions are added per assessed class or category. Agreement is measured by utilizing an area based confusion matrix (Story and Congalton, 1986; Stehman, 1997) where the rows of the matrix correspond to the assessed dataset (ESA CCI LC) and the columns correspond to the reference dataset. In the following comparison assessment, the confusion matrix is referred to as contingency table, since the confusion matrix term is rather to be used in the context of an accuracy assessment. The cell values are calculated as follows in Eq. (1):

$$c_{ij} = \sum_{s=1}^{r} pp_{ij}(s) \quad (1)$$

Table 2
CORINE (CLC) Product information.

	Satellite data	Time consistency	Geometric accuracy satellite data	Geometric accuracy	Thematic accuracy
CLC 1990	Landsat-5 MSS/TM single date	1986–1998	≤50 m	100 m	≥85 % (probably not achieved)
CLC 2000	Landsat-7 ETM single date	2000 +/- 1 year	≤25 m	>100 m	≥85 %
CLC 2006	SPOT-4/5 and IRS LISS III dual date	2006 +/- 1 year	≤25 m	>100 m	≥85 %
CLC 2012	IRS LISS III and RapidEye dual date	2011–2012	≤25 m	>100 m	≥85 %
CLC 2018	Sentinel 2A/ 2B	2017 mid-spring to mid-autumn	≤10 m	>100 m	≥85 %

Table 4
Proportional area (Ref_{prop}) of three LULC categories (see Table 3 for category descriptions) for four example ESA CCI LC pixels (Fig. 1) in comparison to a majority comparison method (Ref_{maj}).

Pixel ID *(k)*	Map Class *i* (ESA CCI LC)	Ref_{prop} Reference Class *j* (CLC)	Ref_{prop} Proportion in pixel pp_{ij} *(k)*	Ref_{maj} Reference Class *j* (CLC)	Ref_{maj} Proportion in pixel pp_{ij} *(k)*
1	2	1	pp_{11} (1) = 0.20	1	pp_{11} (1) = 0
		2	pp_{12} (1) = 0.45	2	pp_{12} (1) = 1
		5	pp_{15} (1) = 0.35	5	pp_{15} (1) = 0
2	1	1	pp_{11} (2) = 0.60	1	pp_{11} (2) = 1
		2	pp_{12} (2) = 0.25	2	pp_{12} (2) = 0
		5	pp_{15} (2) = 0.15	5	pp_{15} (2) = 0
3	5	1	pp_{11} (3) = 0.30	1	pp_{11} (3) = 0
		2	pp_{12} (3) = 0.10	2	pp_{12} (3) = 0
		5	pp_{15} (3) = 0.60	5	pp_{15} (3) = 1
4	2	1	pp_{11} (4) = 0.25	1	pp_{11} (4) = 0
		2	pp_{12} (4) = 0.6	2	pp_{12} (4) = 1
		5	pp_{15} (4) = 0.15	5	pp_{15} (4) = 0

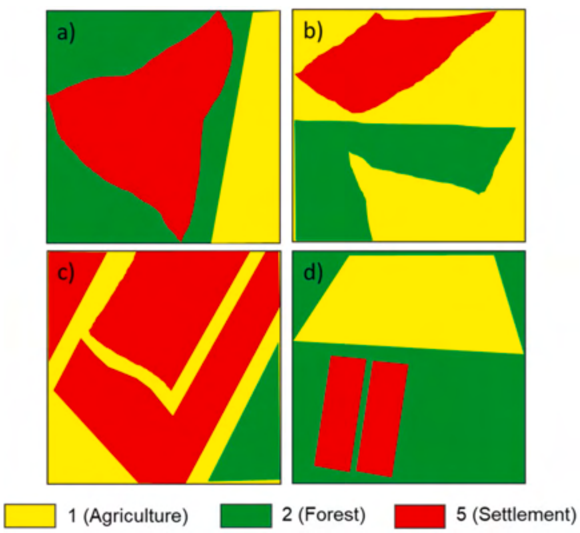


Fig. 1. PAC with three LULC classes for four example pixels (a-d). Proportional area covered by CLC categories agriculture, forest and settlement can be found in Table 4. Pixel IDs refer to example pixels as follows: a) Pixel ID = 1; b) Pixel ID = 2; c) Pixel ID = 3; d) Pixel ID = 4.

where c_{ij} is the value of cell in row i and column j, r is the number of spatial units in the reference dataset and $pp_{ij}(s)$ is the proportion of class j in the spatial unit s assigned to class i in the assessment.

The contingency table for the four example cells is shown in Table 5. In comparison, Table 6 shows the contingency table for the majority method. The main differences between the two matrices are the column sums appearing in decimals that correspond to the added proportions of area per class.

Indices of comparison are then derived from the resulting contingency tables. OA, PR and RE are calculated using Eqs. (2),(3) and (4).

$$OA_i = \frac{\sum_{k=1}^{n} c_{kk}}{n} \tag{2}$$

$$PR_i = \frac{c_{ii}}{\sum_{k=1}^{n} c_{ik}} \tag{3}$$

Table 5
Example contingency table using the PAC. Values derived from Table 4 (Ref_{prop}).

	1 (Agriculture)	2 (Forest)	5 (Settlements)	Sum
1 (Agriculture)	0.6	0.25	0.15	1
2 (Forest)	0.45	1.05	0.5	2
3 (Settlements)	0.3	0.1	0.6	1
Sum	1.35	1.4	1.25	4

Table 6
Example contingency table using the majority method. Values derived from Table 4 (Ref_{maj}).

	1 (Agriculture)	2 (Forest)	5 (Settlements)	Sum
1 (Agriculture)	1	0	0	1
2 (Forest)	0	2	0	2
3 (Settlements)	0	0	1	1
Sum	1	2	1	4

$$RE_j = \frac{c_{jj}}{\sum_{k=1}^{n} c_{kj}} \tag{4}$$

where c_{ij} is the value of the cell in row i in column j of the contingency table and n is the number of classes or categories in the map.

In addition, visual analysis of proportional overlay maps (according to Fig. 1) can indicate spatial patterns of agreement between ESA CCI LC and CLC and therefore reveal inconsistencies for certain categories.

2.5. Research area

The total size of the Research Area (RA) in Eastern Europe is ~867.000 km^2. Countries included in the RA investigated in this paper are Estonia, Latvia, Lithuania, Poland, Hungary, Romania and Slovakia (Fig. 2) for the available CLC time steps 2000, 2006 and 2012. Cells that are not covered by one of the used datasets are left out of the analysis.

The RA in Eastern Europe and the Baltic States was chosen following preliminary investigations that show extensive land use changes suggested by ESA CCI LC (Figs. 2 & 3). The bars in Fig. 3 shows the area gains and losses of the eight LULC categories according to the dataset. The colors show the respective categories that the area was gained from or lost to. Most dynamics are found in the categories agriculture and forest. Most agricultural area loss is due to forest gain and most of agricultural area gain is due to forest area loss. The dynamics might indicate a spatial shift of LULC categories, where net area is not lost but moved. Most of settlement area gain is due to agricultural area loss which is a common dynamic all over the European continent, where urban areas are known to expand.

3. Results

3.1. Contingency tables and agreement measures

Table 7 shows the proportion of each harmonized LULC category for each assessed time step according to ESA CCI LC and CLC. Dominant categories are agriculture and forest with total share of more than 80 % in both datasets. Sparse vegetation is almost non-existent in the RA in both datasets after classification harmonization while shrubland is completely vanished in ESA CCI LC. For agriculture, the difference between the two datasets is ~2 % and for forest ~4% respectively. For grassland and shrubland the proportional areas differ widely.

The proportional areas of categories do not change substantially between the assessed time steps for most of the categories in the individual data sets. The largest relative changes are found for settlement areas, which double the percentage share over the assessed time period in the ESA CCI LC dataset. However, CLC settlement area proportions are ~5 % while the area in ESA CCI LC is much lower with 1.5–2.5 %.

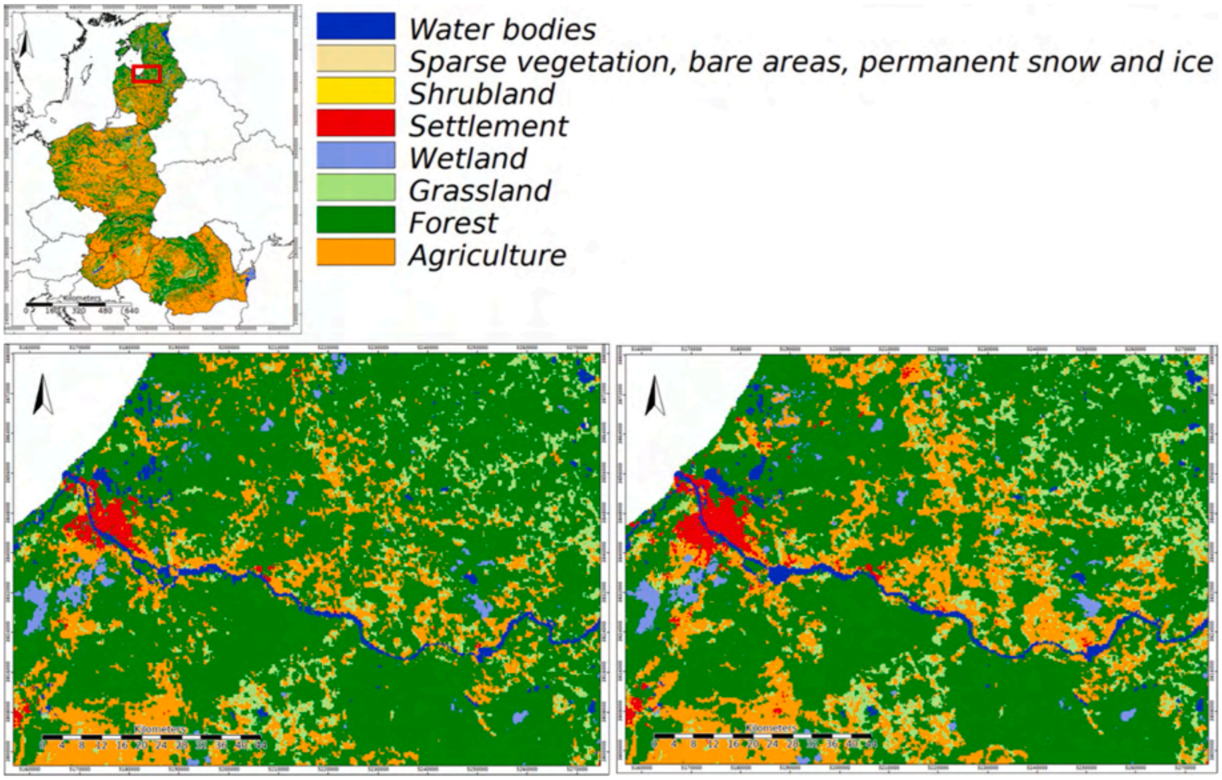


Fig. 2. Research area (upper left) with ESA CCI LC representation of the city Riga (Latvia) in 1992 (lower left) and 2015 (lower right) as an example for extensive LULC changes in the whole RA.

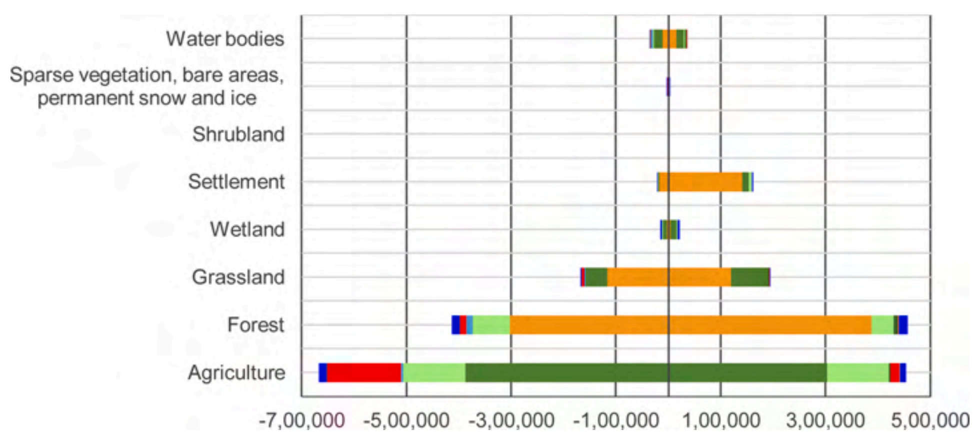


Fig. 3. Category to category changes from 1992-2015 in the RA according to ESA-CCI LC (in 300 × 300 m cell units).

A contingency table gives quantitative information on agreement between the datasets and on issues between LULC categories. Table 8 shows the contingency table for the PAC between ESA CCI LC and CLC for the year 2012. Table 9 shows the same matrix but for the MA. Since there are no significant changes in agreement measures RE and PR among the assessed time steps, only one pair of matrices is exemplarily shown.

Agreement (PA) of ESA CCI LC agriculture and forest with CLC is over 80 %. ~58.000 pixels of CLC agriculture are classified as settlement by ESA CCI LC which is not a major share for agriculture areas but makes a considerable difference for settlements. ESA CCI LC forest areas show the highest agreement with CLC. Nevertheless, the classification of CLC forest areas as agriculture and grassland areas might be not negligible for the performance of RCMs and needs to be discussed. The low accuracies of categories 6 and 7 (shrubland and sparse vegetation) might occur because they are not (or almost not) present in the RA in one or both datasets used after classification harmonization. With ~36 %, agreement for settlements is very low. Most of the CLC settlement areas are classified as agriculture by ESA CCI LC. Grassland and wetlands accuracies are also not on a high level. For both categories, ~50 % of the respective area is classified as forest or agriculture. Possible reasons for this very high disagreement in three categories between the two datasets are investigated in the visual map analysis. Table 9 shows a very similar picture in the contingency table for the majority method. The differences between the methods are not apparent using only the confusion matrices. Nevertheless, the modification step of using only the dominant LULC class of the reference dataset in the MA can give a biased picture of the agreement between two datasets, depending on resolution and

Table 7
Proportional area of every LULC category for ESA CCI LC and CLC [%]. Assessed time steps are 2000, 2006 & 2012.

	Area per category in the RA [%]						
		2000		2006		2012	
	Category	ESA CCI LC	CLC	ESA CCI LC	CLC	ESA CCI LC	CLC
1	Agriculture	54.99	57.74	53.81	56.48	53.63	55.54
2	Forest	34.82	30.87	35.02	31.24	35.05	31.38
3	Grassland	6.08	0.79	6.11	1.05	6.15	1.04
4	Wetland	0.90	1.13	0.90	1.09	0.93	1.07
5	Settlement	1.54	4.57	2.47	4.86	2.58	5.19
6	Shrubland	0	3.02	0	3.38	0	3.86
7	Sparse vegetation, bare areas, permanent snow and ice	0.04	0.10	0.04	0.07	0.04	0.07
8	Water bodies	1.60	1.74	1.61	1.79	1.60	1.81

complexity of classification.

A summary of agreement measures OA, RE and PR is given in Tables 10–12. Like for the confusion matrices, only minor differences are found between the two methods for the agreement measures. OA (Table 10) gives a quantification of the overall agreement of the two datasets. It is steady over all assessed timesteps and for both methods. A more distinguished picture is given by PA and PR (Tables 11 and 12).

Table 11 shows the PR, the proportional share of reference pixels of CLC correctly classified by ESA CCI LC for each category. PR for agricultural and forest is highest. Also, the PR for settlements is high in 2000 but is decreasing considerably per timestep for both approaches. PR for water bodies appears to increase slightly but remains below 75 %. A relatively low proportion of wetlands and an extremely low proportion of grassland is classified correctly by ESA CCI LC.

The RE in Table 12 shows the proportion of area classified by ESA CCI LC that agrees with the respective category in CLC. While PR gives information on the reliability of the assessed dataset, RE gives

Table 10
Overall agreement for all assessed categories and time steps.

Overall agreement (%)					
2000		2006		2012	
Ref_{prop}	Ref_{maj}	Ref_{prop}	Ref_{maj}	Ref_{prop}	Ref_{maj}
76.38	76.59	76.64	76.68	76.21	76.26

Table 11
Precision for all assessed categories and time steps.

	Precision (%)					
	2000		2006		2012	
	Ref_{prop}	Ref_{maj}	Ref_{prop}	Ref_{maj}	Ref_{prop}	Ref_{maj}
1 Agriculture	85.30	85.48	85.51	85.47	84.72	84.69
2 Forest	75.12	75.91	75.90	76.64	75.99	76.68
3 Grassland	5.83	5.93	7.75	7.94	7.64	7.74
4 Wetland	61.66	62.88	65.76	65.89	64.04	64.37
5 Settlement	79.24	78.61	71.46	71.00	72.60	72.06
6 Shrubland	0.00	–	0.00	–	0.00	–
7 Sparse vegetation	36.08	37.35	33.90	33.92	35.01	36.55
8 Water bodies	72.47	72.48	73.25	73.38	73.84	73.88

Table 12
Recall for all assessed categories and time steps.

	Recall (%)					
	2000		2006		2012	
	Ref_{prop}	Ref_{maj}	Ref_{prop}	Ref_{maj}	Ref_{prop}	Ref_{maj}
1 Agriculture	81.24	80.86	81.48	81.04	81.81	79.46
2 Forest	84.73	85.51	85.09	85.77	84.88	84.75
3 Grassland	44.81	46.08	44.87	46.05	45.01	45.78
4 Wetland	48.93	50.69	54.25	55.46	55.36	56.68
5 Settlement	26.79	27.68	36.35	37.13	36.05	36.87
6 Shrubland	–	0	–	0	–	0
7 Sparse vegetation	15.33	18.13	20.55	23.09	20.55	22.20
8 Water bodies	66.29	68.30	65.34	67.48	65.34	66.76

Table 8
contingency table for the year 2012 - Ref_{prop}. Proportional area and SUM are given in thousands (e.g. 8,6 = 8600 spatial units). Recall (RE) and precision (PR) are given in percentage.

	1	2	3	4	5	6	7	8	SUM	PR [%]
1 Agriculture	4373.9	379.0	20.3	14.6	264.0	79.5	1.8	29.9	5163.0	84.72
2 Forest	437.9	2564.0	32.6	21.6	34.0	262.3	1.7	19.8	3374.0	75.99
3 Grassland	450.8	51.0	45.2	5.5	15.7	18.7	1.7	3.3	592.0	7.64
4 Wetland	8.5	8.1	1.6	57.4	0.8	8.6	0.2	4.4	89.6	64.04
5 Settlement	58.7	5.2	0.0	0.3	180.4	1.0	0.1	2.7	248.4	72.60
6 Shrubland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.00
7 Sparse vegetation	0.2	0.2	0.2	0.1	1.5	0.5	1.5	0.3	4.4	35.01
8 Water bodies	16.3	13.2	0.5	4.1	4.1	1.8	0.3	113.9	154.2	73.84
SUM	5346.4	3020.7	100.5	103.6	500.4	372.3	7.5	174.3		
RE [%]	81.81	84.88	45.01	55.36	36.05	0.00	20.55	65.34		

Table 9
contingency table for the year 2012 - Ref_{maj}. Area of categories and SUM are given in thousands (e.g. 8,6 = 8600 spatial units). Recall (RE) and precision (PR) are given in percentage.

	1	2	3	4	5	6	7	8	SUM	PR [%]
1 Agriculture	4173.1	358.7	20.6	13.7	255.7	77.6	1.8	26.7	4927.8	84.68
2 Forest	415.6	2551.4	30.7	20.8	32.5	256.8	1.7	19.2	3328.9	76.64
3 Grassland	452.6	47.5	45.5	5.5	15.1	17.9	1.5	3.1	588.6	7.74
4 Wetland	8.6	7.3	1.7	58.2	0.7	8.6	0.3	5.0	90.5	64.37
5 Settlement	60.4	5.6	0.1	0.4	180.5	0.9	0.1	2.6	250.5	72.08
6 Shrubland	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	50.00
7 Sparse vegetation	0.3	0.2	0.2	0.1	1.4	0.4	1.6	0.3	4.4	36.56
8 Water bodies	16.4	13.6	0.5	4.1	3.8	1.6	0.4	114.3	154.7	73.90
SUM	5127.0	2984.2	99.4	102.7	489.9	363.7	7.3	171.2		
RE [%]	81.39	85.50	45.84	56.70	36.86	0.00	22.20	66.77		

information on the probability of a pixel in the reference map being classified correctly by the assessed dataset. Previously identified issues with grassland, wetlands, settlements and water bodies appear for each assessed timestep. A considerable increase of RE for wetlands and settlements can be noted between 2000 and 2006 but the RE is still below a reasonable level.

It needs to be considered, that slight changes in proportional area influence the agreement measures especially for categories that are rather rarely present in the RA (categories 3, 4,5 and 8). Therefore, in addition to the agreement measures, we provide the visual map analysis, which can give insight on the spatial occurrence of disagreement, also for categories that are rarely occurring in the respective RA.

3.2. Visual map analysis

A spatial analysis of the data, in addition to the raw quantification shown in the confusion matrices, reveals in which areas the disagreements between ESA CCI LC and CLC occur. Since agricultural areas and forest areas are the most well represented categories in the RA with a RE of 81.8 % and 84.88 % respectively, special focus of the maps is on grassland, wetlands, settlements and water bodies. The categories shrubland and sparse vegetation, bare areas, permanent snow and ice are not relevant for the RA and the present analysis and will therefore be neglected in the visual map analysis. The effects of disagreement occur throughout the whole RA. Since the mostly small LULC features are not displayable in a sufficient way for the whole RA, Figs. 4–7 show map sections which highlight typical effects of the disagreeing categories grassland, wetlands, settlements and water bodies for the most recent assessed timestep 2012.

3.2.1. Grassland

RE of grassland for ESA CCI LC reaches around 45 % in the RA. For 2012, ~20 % of the CLC grassland areas are classified as agricultural areas and ~32 % are classified as forest while only ~45 % are classified correctly. According to CLC most of the grassland areas in the RA are in Hungary and around the Carpathian Arc. Fig. 4 shows that the grassland areas classified as forest and agriculture by ESA CCI LC are located around correctly classified areas. Due to the location of the disagreeing areas mostly in mountainous regions, the orography of the area might have major influence on classification results of grassland in ESA CCI LC. Another reason for the low agreement is the classification harmonization. The ESA CCI LC classification includes many mixed classes where for example grassland and low tree density occur as one class. In the harmonization, many of these classes are assigned to the forest class, although the classes might not incorporate actual forest surface properties. The same applies to agriculture in ESA CCI LC, where mixed agriculture-shrubland or agriculture-forest classes are present.

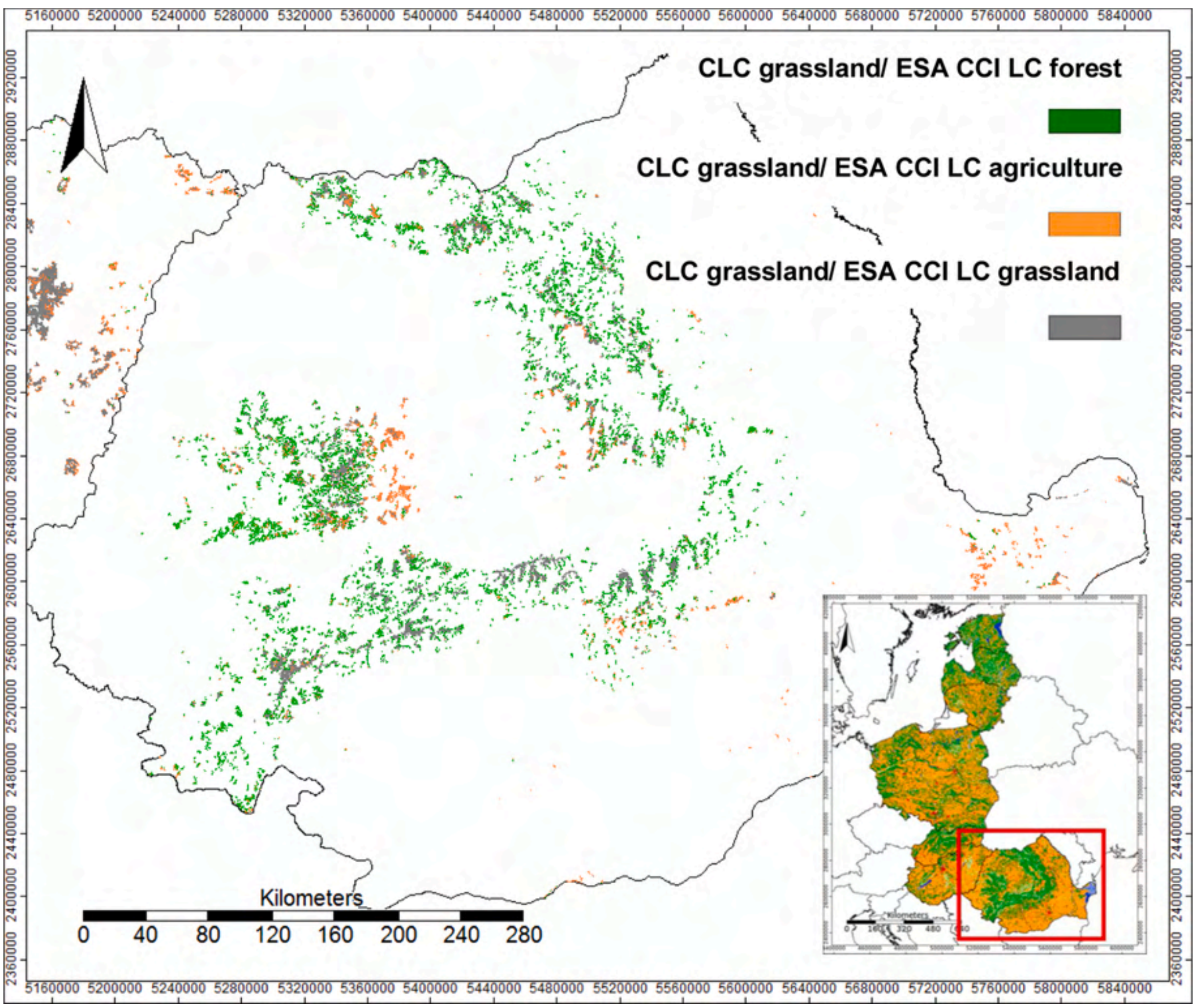


Fig. 4. Romania and parts of Hungary including the grassland proportional area overlay. Grey areas show the agreement of ESA CCI LC with CLC in the grassland category, colors green and orange show the disagreement and the respective other ESA CCI LC category (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

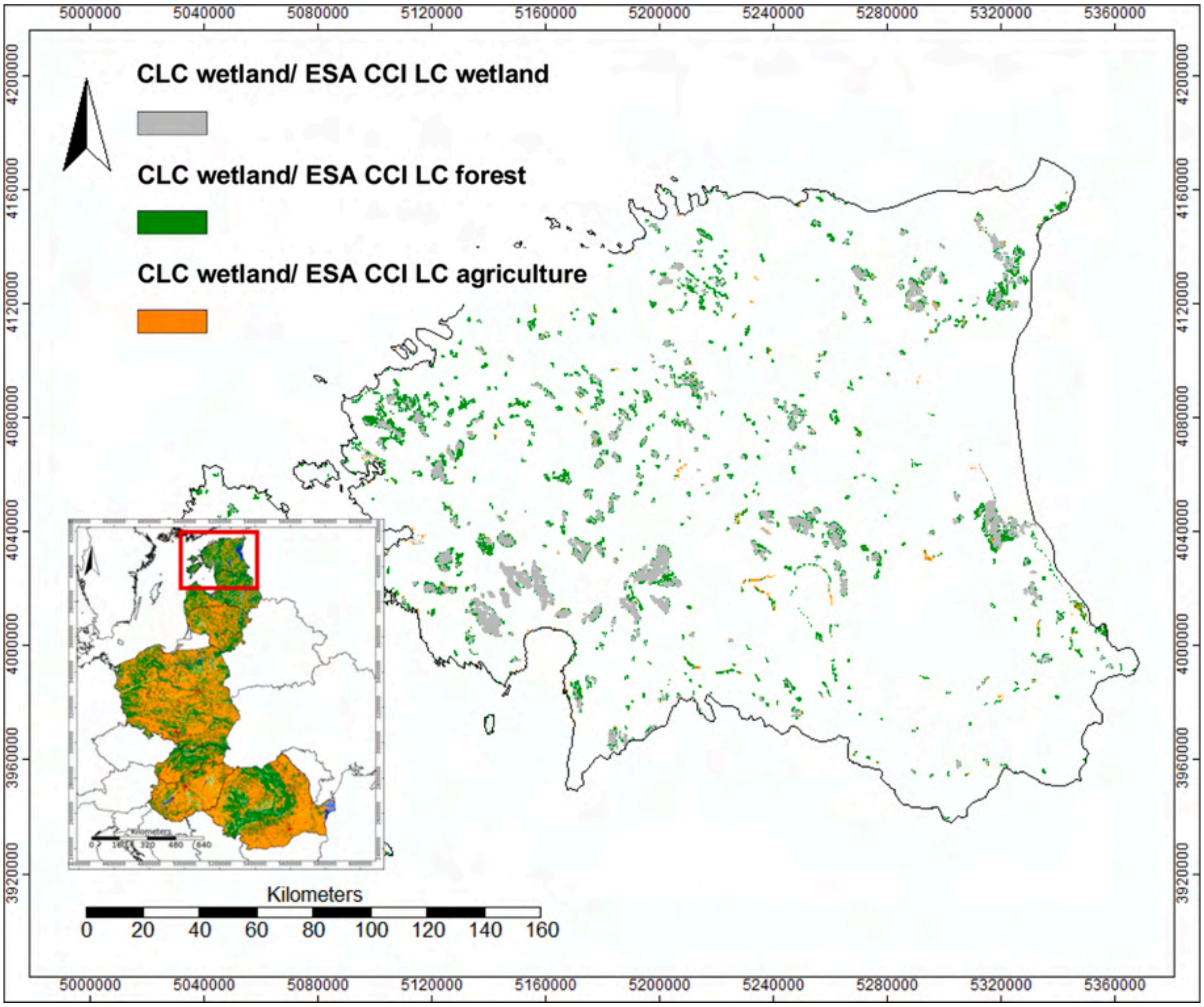


Fig. 5. Proportional area overly map for wetlands in Continental Estonia. Grey areas show the agreement of both data sets.

3.2.2. Wetlands

Wetlands in the RA are found in the Danube river delta in Romania and in the Baltic States Estonia, Latvia and Lithuania where they appear as fens and bogs (Fig. 4). Most of the falsely classified areas are assigned to the categories forest and agriculture by ESA CCI LC (reference year 2012: ~14 % to agriculture and ~21 % to forest, respectively. ~57 % of wetland areas were correctly classified). The map shows the proportional overlay map of ESA CCI LC and CLC for wetlands in Estonia for 2012. Here, the wetlands themselves are reasonably well recognized but the surrounding transition areas are often confused with forest or agricultural areas. Although the overall proportion of wetlands in the RA is relatively small (~1 %) an underestimation of wetlands might have an influence on RCM performance.

3.2.3. Settlements

In the reference year 2012, ~52 % of the total settlement area was classified as agriculture by ESA CCI LC. Fig. 6 shows the Romanian cities Bucharest, Ploiești and the surrounding area as an example. ESA CCI LC works quite well for metropolitan areas but cannot capture smaller, rural settlements which are rather classified as agriculture. Further, linearly distributed features and infrastructure like streets are not identified as settlements or urban areas. There is no clear pattern indicated regarding an influencing spatial factor. The 300 m pixel size of ESA CCI LC has most definitely an influence on the classification results. However, there are also settlement structures seen in the map that are not recognized as such but that are clearly bigger than the minimum mapping unit (MMU) of ESA CCI LC (9 ha). ESA CCI LC includes only one urban LULC class while CLC includes a set of more distinguished artificial areas, which might also be partly responsible for the low agreement of the datasets in the RA.

3.2.4. Water bodies

Around 66 % of water bodies in the RA are correctly classified by ESA CCI LC. As for all other categories, most of the disagreeing areas (~28 % in total) are assigned to agriculture and forest. Fig. 6 shows that most of the disagreeing areas are in or around narrow streams that could not be identified as such by ESA CCI LC, presumably due to surface reflection or surrounding vegetation. Nevertheless, the biggest issue here might be the resolution. The ESA CCI LC resolution of 300 m is in fact not able to capture micro scale landscape features, especially when the features are line shaped.

4. Discussion

The present work investigated the ESA CCI LC dataset regarding agreement with CLC and spatial disagreement patterns of LULC in Eastern Europe and the Baltic States for 2000, 2006 and 2012. The aim of the study was to test if ESA CCI LC agrees with a regional, high-resolution LULC product to investigate the datasets suitability for implementation in regional climate models. Classification harmonization was achieved through transformation of both dataset classifications into eight LULC categories. A PAC was tested against a majority

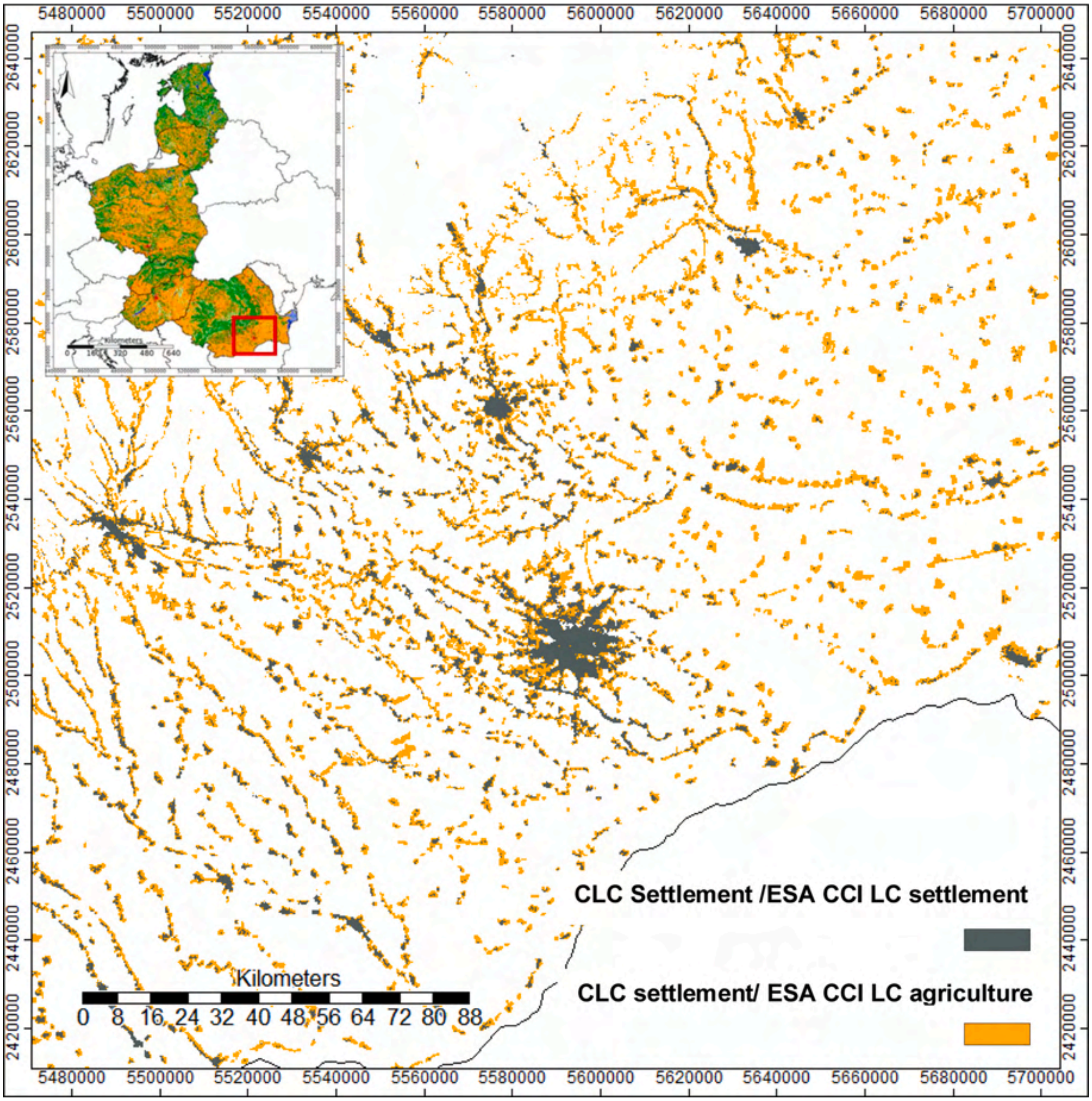


Fig. 6. Proportional overlay map for settlements. ~50 % of the CLC settlement areas are classified as agriculture by ESA CCI LC. Biggest urban agglomerations are the Romanian cities Bucharest (middle of map) and Ploiești (North of Bucharest).

comparison. Main benefit of the PAC for a LC comparative assessment of two or more LC products is that it makes datasets with different resolution and structure comparable without performing preliminary spatial modification. Following the quantitative assessment, spatial disagreement issues for the all categories were investigated through visual analyses. Several issues could be identified for all categories except for shrublands and sparse vegetation, which are not relevant for the RA.

Previous work found the ESA CCI LC epoch time steps (2000, 2005, 2010 and 2015) to be relatively accurate on a global scale (75.4 % for 2015 according to Achard et al., 2017). The present analysis shows consistent OA results for the assessed region (~76 %) for both approaches and all timesteps respectively. However, the comparability between the agreement measures is limited because in all comparative assessment approaches, different data harmonization techniques are used. Nevertheless, the results of the global studies can give a good measure of reliability of the present results. A further restriction is the absence of the categories shrublands and sparse vegetation in the RA. To achieve full comparability with other approaches regarding overall agreement, all existing categories need to be present in the investigated region. Since the aim of the present analysis was to investigate agreement of ESA CCI LC with a regional LULC product in a certain region and from a climate modelling perspective, this criterion can be neglected.

Based on the results for the RA, there are no significant differences in agreement measures between the majority and the PAC for the investigated time steps and categories. The comparison was carried out on the basis of recent findings that showed a considerable bias in agreement measures for the MA due to the loss of small landscape features (Fonte et al., 2020). The findings could not be confirmed in the present study. A reason could be the MMU of CLC which is in fact larger than the MMU of ESA CCI LC (9 ha and 25 ha respectively). Due to the manufacturing process of CLC, small landscape features still appear which makes the product have greater detail than ESA CCI LC. It is expected that when a high resolution product in vector format is used as reference that the differences in agreement measures between the two methods become more apparent. The PAC is therefore still recommended because it can be applied to combine gridded or vector data sets regardless their

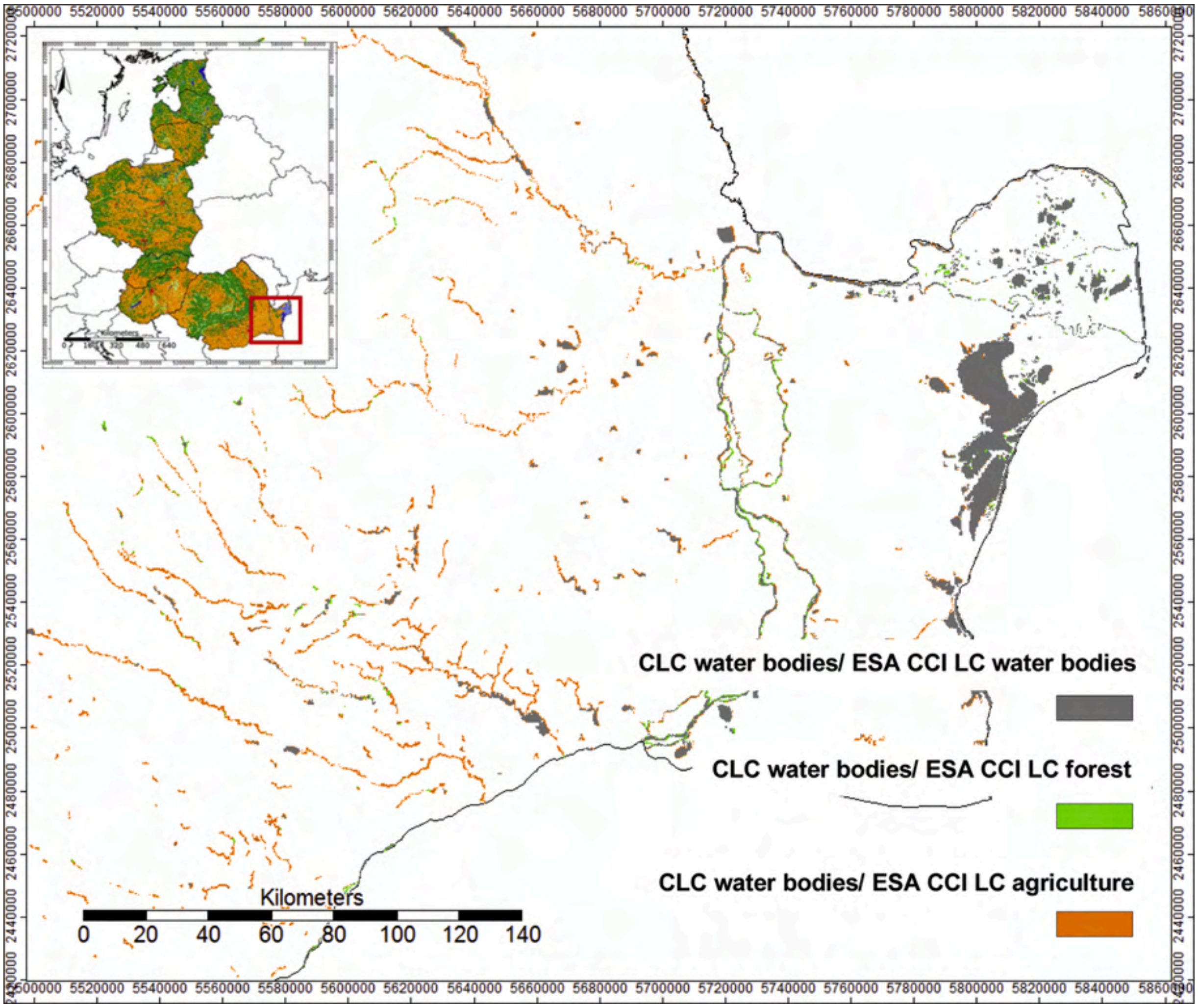


Fig. 7. Proportional overlay map for water bodies. ~28 % of the CLC water bodies are classified as agriculture or forest by ESA CCI LC. The map shows the southeastern part of Romania including the water body proportional agreement results.

structure and resolution. The possibility to compare LULC data products with even less modifications makes the approach advantageous against common LULC dataset comparison techniques. It is therefore highly recommended to be used in LULC dataset comparisons including high-resolution gridded or vector format land cover data.

No substantial changes in overall agreement of ESA CCI LC and CLC over time were found. ESA CCI LC shows a consistent agreement of ~76 % with CLC for every assessed time step. Since the accuracy of CLC was not consistently assessed for all time steps, the reference data set can be biasing the present analysis. Although the CLC database is considered to be one of the most accurate and consistent land cover products for Europe, errors can occur during the production process. Since CLC is only available for certain countries in Europe, independent regional land cover data sets and other available ground truth data need to be used when transferring the PAC to other regions of the globe accordingly. The fact that the data structure of the reference data set does not matter for the PAC solves the issue of finding suitable reference data partly. All types of reference data, which might come in different formats and structures should be used simultaneously for comparison without spatial transformation biasing the agreement measures.

Agriculture and forest are the predominant categories in the RA. Compared to CLC, Recall of ESA CCI LC is over 80 % for both which can be considered as relatively high. However, consistency between other global state of the art LC products seems problematic for both categories (Fritz et al., 2011) Especially for land-climate interaction scenarios that depend highly on LU and LC, overestimation of forest and agriculture has, due to specific LULC related surface properties (e.g. high surface roughness and albedo), most certainly an impact on RCM performance. Therefore, it can be beneficial to use a modification or a combination of the ESA CCI LC data set with other reliable, independent LULC data (Chu et al., 2011). Crowdsourcing approaches like the Geo-Wiki (https://www.geo-wiki.org/) can further be a valuable source of high resolution input data to test reliability and to refine LULC maps in the desired region of interest.

For grassland, mostly present in Hungary and Romania, inaccuracies tend to be found in areas around mountains and in valleys along the Carpathian Arc. Over 50 % of the grassland is differently classified by ESA CCI LC, ~32 % of it as forest. These inaccuracies might occur due to shadowing in the valleys caused by steep and narrow slopes. The constraints of satellite image classifications in mountainous regions are widely known in the remote sensing community and also permanent improvement was achieved during the last decades (Giles, 2001; Mostafa, 2017; Shahtahmassebi et al., 2013). It needs to be investigated if the issue for grassland occurs only in the investigated region or also in other mountainous regions to see whether this is a global issue of ESA CCI LC. Further, the mixed classes of ESA CCI LC are a biasing factor. It needs to be checked, if these disagreements also occur with different reference datasets or with a different dataset harmonization method.

Issues with wetlands are mostly found in the Baltic Countries where a considerable area of wetlands are classified as forest by ESA CCI LC.

These findings are consistent with (Törmä et al., 2015) who discovered misclassification of bogs, mires and marshes as forest in Finland. Against the background of widely differing surface properties of forest and wetland respectively, it should be investigated how this issue can be handled for implementation in RCMs.

Almost half of the settlement areas in the RA (~47 %) are classified as agricultural areas by ESA CCI LC. Mostly missing are rural settlements and infrastructure like roads without an agglomeration center and of linear shape. This might be because ESA CCI LC only incorporates one explicit urban LULC class. Since RCMs are moving to finer resolution, explicit representation of settlement structures is becoming critical. In particular, when RCMs are used on high spatial resolutions down to convection permitting scales below 4 km and are increasingly employing an explicit urban parameterization, relevance of high quality urban input data is extremely increased (Trusilova et al., 2013; Daniel et al., 2019; Langendijk et al., 2019). Modifications of the ESA CCI LC maps, which take the linear shaped infrastructure and rural settlements into account, might be necessary. For instance, LULC datasets that are specialized on urban representation like the urban atlas for Europe (Montero et al., 2014) or equivalent regional and global data products for other regions could be combined with the ESA CCI LC dataset.

The water bodies show a similar picture like the settlements regarding spatial disagreement issues. Coherent features like lakes or larger basins are captured very well but when it comes to rivers and streams, ESA CCI LC classifies water bodies as agriculture. Considering the surface properties of water bodies as well as the influence of rivers and streams on the surrounding landscape features, the missing features in ESA CCI LC will be relevant when implementing the data set into an RCM. Since the disagreement seems to be limited to the streams and rivers that are represented very well by CLC, it might be beneficial to improve the ESA CCI LC dataset through the integration of CLC or other suitable reference data that represents the river network in a more sufficient way. ESA CCI LC addressed the issue themselves with publishing a global water bodies map on 150 m horizontal resolution (Lamarche et al., 2017). That map could be integrated into ESA CCI LC before aggregation into coarser resolution, to preserve the small water body proportions for further use of the data.

5. Conclusion

The present work investigated the agreement of the ESA CCI LC dataset with CLC over Eastern Europe and the Baltic states to explore ESA CCI LCs suitability for implementation into RCMs over Europe. Three timesteps of the annual ESA CCI LC dataset were compared to CORINE LC, applying a PAC and a majority method, respectively. Classification harmonization of the assessed and the reference dataset was achieved through transformation into eight LULC categories.

Taking all results regarding overall agreement of ~76 % and temporal consistency into account the ESA CCI LC is considered to be suitable for implementation into RCMs by taking the following issues found in the present study under consideration. Disagreement with CORINE LC, which is considered a reliable reference for Europe, is for ESA CCI LC ~55 % for grassland, ~43 % for wetland, ~64 % for settlements and ~34 % for water bodies in the investigated RA.

Regional quality of the dataset must be confirmed for each region of interest separately with comparison to independent reference data. Although ESA CCI LC was found to be overall suitable for implementation in RCMs, spatial disagreement patterns were found that might influence RCM performance on certain scales which must also be investigated in each region of interest.

To get a deeper understanding of spatial disagreement not only for the RA but for the whole European Continent a consistent reference database for Europe should be developed. In addition to the continuous LULC product CLC, regional, independent datasets or also datasets specified on one LULC aspect should be included in the analysis which then can be compared using the PAC, regardless spatial structure and resolution.

In order to investigate the effects of detected LULC issues on the performance of RCMs and on regional climate in a region of interest, different LULC distributions and maps could be implemented into an RCM and tested. The testing would include different intensities of LULC over- and underestimation as well as varying spatial resolutions to quantify the impact of inaccuracies of LULC on different scales and to specify how to treat inaccuracies in LULCC products in regional climate modelling.

CRediT authorship contribution statement

V. Reinhart: Writing - original draft, Conceptualization, Validation, Formal analysis, Investigation. **C.C. Fonte:** Methodology, Writing - review & editing. **P. Hoffmann:** Writing - review & editing, Investigation. **B. Bechtel:** Methodology, Writing - review & editing. **D. Rechid:** Writing - review & editing, Supervision, Project administration. **J. Boehner:** Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

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

Acknowledgements

This work was conducted and financed within the framework of the Helmholtz Institute for Climate Service Science (HICSS), a cooperation between Climate Service Center Germany (GERICS) and Universität Hamburg as part of the project LANDMATE. Geospatial analyses were done with overwhelming support of the SAGA GIS User Group (http://www.saga-gis.org/en/index.html) at the Universität Hamburg (Institute of Geography, Section of Physical Geography). We thank Dr. Laurens Bower (GERICS) for the support in the internal review process. Finally, we thank the anonymous reviewers for their highly valuable feedback to our work.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jag.2020.102221.

References

Achard, F., Bontemps, S., Lamarche, C., Da Maet, T., Mayaux, P., Van Bogaert, E., Defourny, P., 2017. Quality Assessment of the CCI Land Cover Maps. https://www.esa-landcover-cci.org/?q=webfm_send/159.

Bechtel, B., Demuzere, M., Stewart, I.D., 2019. A weighted accuracy measure for land cover mapping: comment on Johnson et al. Local climate zone (LCZ) map accuracy assessments should account for land cover physical characteristics that affect the local thermal environment. Remote Sens. (Basel) 12 (11), 1769. https://doi.org/10.3390/rs12111769.

Bontemps, S., Defourny, P., Van Bogaert, E., Arino, O., Kalogirou, V., Perez, J.R., 2011. GLOBCOVER 2009-Products Description and Validation Report. URL: Http://Ionia1.Esrin. Esa.Int/Docs/GLOBCOVER2009_Validation_Report_2,2.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., Zülkhe, M., Kalogirou, V., Arino, O., 2013. Consistent global land cover maps for climate modeling communities: current achievements of the ESA's land cover CCI. ESA Living Planet Symposium 2013 (September), 9–13.

Brovkin, V., Boysen, L., Arora, V.K., Boisier, J.P., Cadule, P., Chini, L., Claussen, M., Friedlingstein, P., Gayler, V., Van den hurk, B.J.J.M., Hurtt, G.C., Jones, C.D., Kato, E., De noblet-ducoudre, N., Pacifico, F., Pongratz, J., Weiss, M., 2013. Effect of anthropogenic land-use and land-cover changes on climate and land carbon storage in CMIP5 projections for the twenty-first century. J. Clim. 26 (18), 6859–6881. https://doi.org/10.1175/JCLI-D-12-00623.1.

Büttner, G., Maucha, G., 2006. The thematic accuracy of Corine land cover 2000. Assessment using LUCAS (land use/cover area frame statistical survey). EEA Tech. Rep. 7/2006 (7), 85. www.eea.europa.eu.

Büttner, G., Maucha, G., Kosztra, B., 2011.). European Validation of Land Cover Changes in CLC2006 Project. EARSeL Symposium, Prague.
Castilla, G., Hay, G.J., 2007. Uncertainties in land use data To cite this version: HAL Id: Hal-00305113 Uncertainties in land use data. Hydrol. Earth Syst. Sci. Discuss. 11 (6), 1857–1868.
Chu, J., Syktus, J., McAlpine, C., Thatcher, M., Scarth, P., Jeffrey, S., Katzfey, J., Zhang, H., McGregor, J., Adams-Hosking, C., 2011. Validation of land surface products for modelling the climate impacts of large-scale revegetation in Queensland. In: MODSIM 2011 - 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living With Uncertainty. December 2011, pp. 2676–2682.
Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., Böhner, J., 2015. System for automated geoscientific analyses (SAGA) v. 2.1. 4. Geosci. Model. Dev. Discuss. 8 (2).
Copernicus Land Monitoring Service, 2020. Evolution of Corine Land Cover. https://land.copernicus.eu/pan-european/corine-land-cover.
Daniel, M., Lemonsu, A., Déqué, M., Somot, S., Alias, A., Masson, V., 2019. Benefits of explicit urban parameterization in regional climate modeling to study climate and city interactions. Clim. Dyn. 52 (5–6), 2745–2764. https://doi.org/10.1007/s00382-018-4289-x.
Davin, E.L., Rechid, D., Breil, M., Cardoso, R.M., Coppola, E., Hoffmann, P., Jach, L.L., Katragkou, E., de Noblet-Ducoudré, N., Radtke, K., Raffa, M., Soares, P.M.M., Sofiadis, G., Strada, S., Strandberg, G., Tölle, M.H., Warrach-Sagi, K., Wulfmeyer, V., 2019. Biogeophysical impacts of forestation in Europe: First results from the LUCAS Regional Climate Model intercomparison. Earth Syst. Dyn. Discuss. (February), 1–31. https://doi.org/10.5194/esd-2019-4.
ESA, 2017. Land Cover CCI Product User Guide Version 2.0. http://maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf.
ESA, 2019. ESA/CCI Viewer. http://maps.elie.ucl.ac.be/CCI/viewer/download.php.
Fonte, C.C., See, L., Laso-Bayas, J.C., Lesiv, M., Fritz, S., 2020. Assessing the accuracy of land use land cover (LULC) maps using class proportions in the reference data. Isprs Ann. Photogramm. Remote. Sens. Spat. Inf. Sci. V-3–2020, 669–674. https://doi.org/10.5194/isprs-annals-V-3-2020-669-2020.
Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80 (1), 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4.
Fritz, S., See, L., Rembold, F., 2010. Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. Int. J. Remote Sens. 31 (9), 2237–2256. https://doi.org/10.1080/01431160902946598.
Fritz, S., See, L., McCallum, I., Schill, C., Obersteiner, M., Velde, Mvander, Boettcher, H., Havlík, P., Achard, F., 2011. Highlighting continued uncertainty in global land cover maps for the user community. Environ. Res. Lett. 6 (4), 044005. https://doi.org/10.1088/1748-9326/6/4/044005.
Fuchs, R., Herold, M., Verburg, P.H., Clevers, J.G.P.W., 2013. A high-resolution and harmonized model approach for reconstructing and analysing historic land changes in Europe. Biogeosciences 10 (3), 1543–1559. https://doi.org/10.5194/bg-10-1543-2013.
Gao, Y., Weiher, S., Markkanen, T., Pietikäinen, J.-P., Gregow, H., Henttonen, H.M., Jacob, D., Laaksonen, A., 2015. Implementation of the CORINE Land Use Classification in the Regional Climate Model REMO.
Georgievski, G., Hagemann, S., 2019. Characterizing uncertainties in the ESA-CCI land cover map of the epoch 2010 and their impacts on MPI-ESM climate simulations. Theor. Appl. Climatol. 137 (1–2), 1587–1603. https://doi.org/10.1007/s00704-018-2675-2.
Giles, P.T., 2001. Remote sensing and cast shadows in mountainous terrain. Photogramm. Eng. Remote Sensing 67 (7), 833–840.
GISAT, 2019. CLC Seamless Data Coverage. https://land.copernicus.eu/user-corner/technical-library/clc-country-coverage-v18.1.
Heymann, Y., 1994. CORINE Land Cover: Technical Guide. Office for Official Publ. of the Europ. Communities.
Houghton, R.A., House, J.I., Pongratz, J., Van Der Werf, G.R., Defries, R.S., Hansen, M.C., Le Quéré, C., Ramankutty, N., 2012. Carbon emissions from land use and land-cover change. Biogeosciences 9 (12), 5125–5142. https://doi.org/10.5194/bg-9-5125-2012.
Hua, T., Zhao, W., Liu, Y., Wang, S., Yang, S., 2018. Spatial consistency assessments for global land-cover datasets: a comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO. Remote Sens. 10 (11), 1846.
Hurtt, G.C., Chini, L.P., Frolking, S., Betts, R.A., Feddema, J., Fischer, G., Fisk, J.P., Hibbard, K., Houghton, R.A., Janetos, A., Jones, C.D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., et al., 2011. Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. Clim. Change 109 (1), 117–161. https://doi.org/10.1007/s10584-011-0153-2.
Jaffrain, G., Sannier, C., Pennec, A., Dufourmont, H., 2017. CORINE land cover 2012 final validation report. Copernicus Land Monitoring 214.
Koubodana, D.H., Diekkrüger, B., Näschen, K., Adounkpe, J., Atchonouglo, K., 2019. Impact of the accuracy of land cover data sets on the accuracy of land cover change scenarios in the Mono River Basin, Togo, West Africa. Int. J. Adv. Remote. Sens. Gis 8 (1), 3073–3095. https://doi.org/10.23953/cloud.ijarsg.422.
Lamarche, C., Santoro, M., Bontemps, S., d'Andrimont, R., Radoux, J., Giustarini, L., Brockmann, C., Wevers, J., Defourny, P., Arino, O., 2017. Compilation and validation of SAR and optical data products for a complete and global map of inland/ocean water tailored to the climate modeling community. Remote Sens. (Basel) 9 (1), 36.
Langendijk, G.S., Rechid, D., Jacob, D., 2019. Urban areas and urban–rural contrasts under climate change: what does the EURO-CORDEX ensemble tell us?—investigating near surface humidity in Berlin and its surroundings. Atmosphere 10 (12), 730.
Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R.A., Peng, S., 2017. Gross and net land cover changes based on plant functional types derived from the annual ESA CCI land cover maps. Earth Syst. Sci. Data Discuss. 1–23. https://doi.org/10.5194/essd-2017-74.
Loveland, T.R., Belward, A.S., 1997. The IGBP-DIS global 1km land cover data set, DISCover: First results. Int. J. Remote Sens. 18 (15), 3289–3295.
Montero, E., Van Wolvelaer, J., Garzón, A., 2014. The European urban atlas. Land Use and Land Cover Mapping in Europe. Springer, pp. 115–124.
Mostafa, Y., 2017. A review on various shadow detection and compensation techniques in remote sensing images. Can. J. Remote. Sens. 43 (6), 545–562.
Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. Remote Sens. Environ. 148, 42–57. https://doi.org/10.1016/j.rse.2014.02.015.
Pérez-Hoyos, A., Rembold, F., Kerdiles, H., Gallego, J., 2017. Comparison of global land cover datasets for cropland monitoring. Remote Sens. (Basel) 9 (11). https://doi.org/10.3390/rs9111118.
Petit, C.C., Lambin, E.F., 2002. Long-term land-cover changes in the Belgian Ardennes (1775–1929): model-based reconstruction vs. Historical maps. Glob. Chang. Biol. 8 (7), 616–630. https://doi.org/10.1046/j.1365-2486.2002.00500.x.
Pongratz, J., Reick, C., Raddatz, T., Claussen, M., 2008. A reconstruction of global agricultural areas and land cover for the last millennium. Global Biogeochem. Cycles 22 (3). https://doi.org/10.1029/2007GB003153.
Ramankutty, N., Foley, J.A., 1999. Estimating historical changes in land cover:North American croplands from 1850 to 1992. Glob. Ecol. Biogeogr. 8 (5), 381–396. https://doi.org/10.1046/j.1365-2699.1999.00141.x.
Samasse, K., Hanan, N.P., Tappan, G., Diallo, Y., 2018. Assessing cropland area in West Africa for agricultural yield analysis. Remote Sens. (Basel) 10 (11). https://doi.org/10.3390/rs10111785.
Santos-Alamillos, F., Pozo-Vazquez, D., Ruiz-Arias, J., Tovar-Pescador, J., 2015. Influence of land-use misrepresentation on the accuracy of WRF wind estimates: evaluation of GLCC and CORINE land-use maps in southern Spain. Atmos. Res. 157 https://doi.org/10.1016/j.atmosres.2015.01.006.
Sarmento, P., Fonte, C.C., Dinis, J., Stehman, S.V., Caetano, M., 2015. Assessing the impacts of human uncertainty in the accuracy assessment of land-cover maps using linguistic scales and fuzzy intervals. Int. J. Remote Sens. 36 (10), 2524–2547. https://doi.org/10.1080/01431161.2015.1043034.
Sertel, E., Robock, A., Ormeci, C., 2010. Impacts of land cover data quality on regional climate simulations. Int. J. Climatol. 30 (13), 1942–1953. https://doi.org/10.1002/joc.2036.
Shahtahmassebi, A., Yang, N., Wang, K., Moore, N., Shen, Z., 2013. Review of shadow detection and de-shadowing methods in remote sensing. Chin. Geogr. Sci. 23 (4), 403–420.
Stehman, S.V., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sens. Environ. 62 (1), 77–89. https://doi.org/10.1016/S0034-4257(97)00083-7.
Story, M., Congalton, R.G., 1986. Remote sensing brief accuracy assessment: a user's perspective. Photogramm. Eng. Remote Sensing 52 (3), 397–399.
Tchuenté, A.T.K., Roujean, J.-L., De Jong, S.M., 2011. Comparison and relative quality assessment of the GLC2000, GLOBCOVER, MODIS and ECOCLIMAP land cover data sets at the African continental scale. Int. J. Appl. Earth Obs. Geoinf. 13 (2), 207–219.
Törmä, Markus, Markkanen, Tiina, Hatunen, Suvi, Härmä, Pekka, Mattila, Olli-Pekka, Arslan, Ali Nadir, 2015. Assessment of land-cover data for land-surface modelling in regional climate studies. Boreal Environment Research 20, 243–260.
Trusilova, K., Früh, B., Brienen, S., Walter, A., Masson, V., Pigeon, G., Becker, P., 2013. Implementation of an urban parameterization scheme into the regional climate model COSMO-CLM. J. Appl. Meteorol. Climatol. 52 (10), 2296–2311. https://doi.org/10.1175/JAMC-D-12-0209.1.
Verburg, P.H., Neumann, K., Nol, L., 2011. Challenges in using land use and land cover data for global change studies. Glob. Chang. Biol. 17 (2), 974–989. https://doi.org/10.1111/j.1365-2486.2010.02307.x.
Vilar, L., Garrido, J., Echavarría, P., Martínez-Vega, J., Martín, M.P., 2019. Comparative analysis of CORINE and climate change initiative land cover maps in Europe: implications for wildfire occurrence estimation at regional and local scales. Int. J. Appl. Earth Obs. Geoinf. 78 (November 2018), 102–117. https://doi.org/10.1016/j.jag.2019.01.019.
Yang, Y., Xiao, P., Feng, X., Li, H., 2017. Accuracy assessment of seven global land cover datasets over China. Isprs J. Photogramm. Remote. Sens. 125, 156–173. https://doi.org/10.1016/j.isprsjprs.2017.01.016.

A.0.2 Publication III

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Earth System Science Data
Open Access Discussions

High-resolution land-use land-cover change data for regional climate simulations over Europe - Part I: The plant functional type basemap for 2015

Vanessa Reinhart[1,2], Peter Hoffmann[1,2], Diana Rechid[1], Jürgen Böhner[3], and Benjamin Bechtel[4]

[1]Helmholtz-Zentrum Hereon, Climate Service Center Germany (GERICS), Fischertwiete 1, 20095 Hamburg, Germany
[2]Universität Hamburg, Institut of Geography, Section Physical Geography, Bundesstraße 55, 20146 Hamburg, Germany
[3]Universität Hamburg, Institute of Geography, Cluster of Excellence “Climate, Climatic Change, and Society” (CLICCS), 805A Geomatikum, Bundesstraße 55, D-20146, Hamburg, Germany
[4]Ruhr-Universität Bochum, Department of Geography, Universitätsstraße 150/ Gebäude IA, 44801 Bochum, Germany

Correspondence: Vanessa Reinhart (vanessa.reinhart@hereon.de)

Abstract. The concept of plant functional types (PFTs) is shown to be beneficial in representing the complexity of plant characteristics in land use and climate change studies using regional climate models (RCMs). By representing land use and land cover (LULC) as functional traits, responses and effects of specific plant communities can be directly coupled to the lowest atmospheric layers. To meet the requirements of RCMs for realistic LULC distribution, we developed a PFT dataset for Europe (LANDMATE PFT Version 1.0; Reinhart et al., 2021b). The dataset is based on the high-resolution ESA-CCI land cover dataset and is further improved through the the additional use of climate information. Within the LANDMATE PFT dataset, satellite-based LULC information and climate data are combined to achieve the best possible representation of the diverse plant communities and their functions in the respective regional ecosystems while keeping the dataset most flexible for application in RCMs. Each LULC class of ESA-CCI is translated into PFT or PFT fractions including climate information by using the Holdridge Life Zone concept. Through the consideration of regional climate data, the resulting PFT map for Europe is regionally customized. A thorough evaluation of the LANDMATE PFT dataset is done using a comprehensive ground truth database over the European Continent. A suitable evaluation method has been developed and applied to assess the quality of the new PFT dataset. The assessment shows that the dominant LULC groups, cropland and woodland, are well represented within the dataset while uncertainties are found for some less represented LULC groups. The LANDMATE PFT dataset provides a realistic, high-resolution LULC distribution for implementation in RCMs and is used as basis for the LUCAS LUC dataset introduced in the companion paper by Hoffmann et al. (submitted) which is available for use as LULC change input for RCM experiment setups focused on investigating LULC change impact.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

1 Introduction

Land use and land cover (LULC), including the vegetation type and function, was declared an Essential Climate Variables (ECVs) by the Global Climate Observing System (GCOS) (Bojinski et al., 2014). Changes in ECVs are crucial factors of climate change and therefore need to be monitored and further represented in climate models to be able to assimilate and understand atmospheric processes and feedback effects on different scales. For LULC, anthropogenic modifications are the most important drivers of change. De- and reforestaion and expansion of urban and cropland areas affect biogeophysical (e.g., albedo, roughness, evapotranspiration, runoff) and biogeochemical (e.g., carbon emissions and sinks) surface properties and processes (Mahmood et al., 2014; Lawrence and Vandecar, 2015; Alkama and Cescatti, 2016; Perugini et al., 2017; Davin et al., 2020). Besides LULC changes, land management practices are being assessed regarding influence of related land surface modifications on regional climate, and also the potential of land management practices regarding climate change adaptation and mitigation efforts (Lobell et al., 2006; Kueppers et al., 2007; Burke and Emerick, 2016).

In order to represent impacts and feedbacks of LULC modifications as realistic as possible, regional climate models (RCMs) require an accurate representation of LULC and its changes. In this context, the concept of plant functional types (PFTs) is increasingly used for the representation of LULC in RCMs. A comprehensive review of the subsequent development of PFTs representing vegetation dynamics in climate models was done by Wullschleger et al. (2014). The need for applicable global PFT maps for vegetation models that are used with atmospheric models was already well emphasized by Box (1996). Moreover, the requirement that a climate model should include a vegetation model representing the biosphere was discussed by Lavorel et al. (2007). One criterion that is highly emphasized is the inter-regional applicability of a preferably simple PFT classification, which has the ability to capture key characteristics of the biosphere from biome to continental scale, regardless of climate zone and individual vegetation composition. A variety of PFT definitions and cross-walking procedures (CWPs), used for translating LULC products into global or regional PFT maps, are currently available. The European Space Agency Climate Change Initiative (ESA-CCI) and the United States Geological Service (USGS) provide the only two ready to use continuous global products to the community (Poulter et al., 2015; Sulla-Menashe and Friedl, 2018). However, the individual PFT definitions and CWPs as well as the mostly satellite based input data differ greatly in complexity and temporal and horizontal resolution (Bonan et al., 2002; Winter et al., 2009; Lu and Kueppers, 2012). Moreover, inter-regional consistency cannot be achieved by products that origin from regionally constrained input data or regionally adapted CWPs. Therefore, the additional use of climate information in the CWP from LULC to PFT is a highly useful step, to create a dynamically customizable product, that can be adapted to various climate and vegetation characteristics (Poulter et al., 2011).

With the present work, we introduce a PFT map for the European Continent that specifically addresses the requirements of the RCM community (Bontemps et al., 2013). The land cover maps of the ESA-CCI are translated into 16 PFTs creating an updated version of the interactive MOsaic-based Vegetation (iMOVE) PFTs that were originally developed for the RCM REMO (Wilhelm et al., 2014). Climate information is implemented into the CWP employing the Holdridge ecosystem classification concept based on the Holdridge Life Zones (HLZs; Holdridge et al., 1967), which provide a global classification of climatic zones in relation to potential vegetation cover. The HLZ concept is commonly used as a tool for ecosystem mapping from

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

various overlapping research communities (Lugo et al., 1999; Yue et al., 2001; Khatun et al., 2013; Szelepcsényi et al., 2014; Tatli and Dalfes, 2021). This paper gives a detailed documentation on the preparation of the PFT map - hereinafter referred to as "LANDMATE PFT" - within the Helmholtz Institute for Climate Service Science (HICSS) project "Modelling human LAND surface Modifications and its feedbacks on local and regional cliMATE" (LANDMATE). The LANDMATE PFT map is prepared in close collaboration with the EURO-CORDEX Flagship Pilot Study Land Use and Climate Across Scales (FPS LUCAS; Rechid et al., 2017). Within the FPS LUCAS, RCM experiments are coordinated among an RCM ensemble to investigate the impact of LULC change for past climate and future climate scenarios. Through creation of LANDMATE PFT and the time series LUCAS LUC (Hoffmann et al., submitted), the need for improved LULC and LULC change representation among the FPS LUCAS RCM ensemble is met. For the preparation of LANDMATE PFT, we developed a CWP for the translation of LULC classes of ESA-CCI into 16 PFTs according to the needs of regional climate modellers from all over Europe (Bontemps et al., 2013). A key issue to address in the map development process is the accuracy of LULC representation in the final product (Hartley et al., 2017). In order to assess the quality of the product, we compared the LANDMATE PFT map to a comprehensive ground truth database for large parts of the European Continent. The quality information derived from the assessment supports the RCM community in addressing and interpreting uncertainties caused by LULC representation in RCMs. The general workflow and subsequently all utilized datasets are summarized in section 2 while the major steps of the CWP are listed in section 3. Section 4 introduces in detail the accuracy assessment procedure followed directly by the results in section 4.3. All CWTs and figures corresponding to the CWP and the accuracy assessment can be found in Appendix A and B.

2 Methods and data

The LANDMATE PFT map (Reinhart et al., 2021b) is a combination of multiple datasets and concepts created using well-established methods and in addition, by considering the expertise of regional climate modellers from all over Europe within the FPS LUCAS.

2.1 General workflow

The workflow to generate the LANDMATE PFT map is summarized in fig. 1, which also includes the steps to generated the LUCAS LUC dataset further described in the companion paper by Hoffmann et al. (submitted). First, the ESA-CCI land cover map (Sect. 2.2.1), which has a native resolution of ~300 m, is aggregated to the 0.1° target resolution using SAGA GIS (Conrad et al., 2015). The target resolution results from the FPS LUCAS ensemble resolution (i.e., EURO-CORDEX domain EUR-11) that is used for LULC change impact studies in FPS LUCAS Phase II. The LULC type information from the original product is preserved in fractions per 0.1° grid cell which is advantageous to common majority resampling methods. The sum of PFT fractions in the whole dataset remains the same in all target resolutions, only the distribution of fractions per grid cell changes depending on the target resolution.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

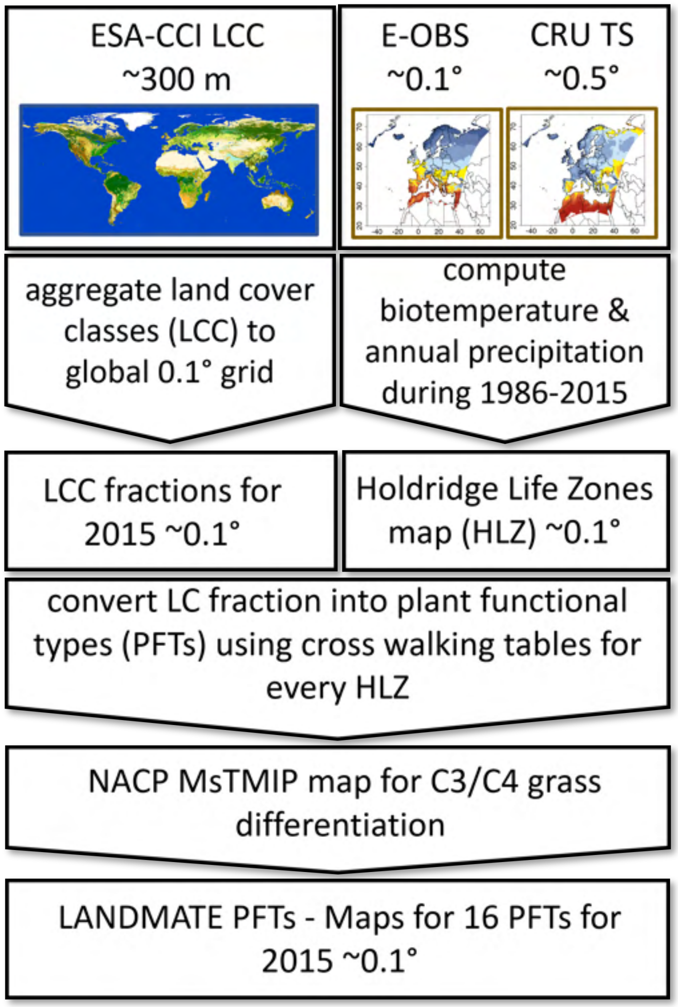


Figure 1. The general workflow to generate LANDMATE PFT 2015 Version 1.0. This workflow is part of the workflow to generate the LUCAS LUC time series as introduced in the companion paper by Hoffmann et al. (submitted)

The E-OBS gridded climate data (Sect. 2.2.2) is utilized for the preparation of the HLZ map over Europe (Sect. 2.2.4). From E-OBS, the ensemble mean 2-meter-temperature and annual precipitation from 1950-2020 are used to create the HLZ map of 0.1° horizontal resolution which is further implemented in the CWTs to prepare the final LANDMATE PFT maps. For regions that are not covered by E-OBS, the respective data of the CRU dataset (Sect. 2.2.3) is used.

For each of the 37 ESA-CCI land cover classes, an individual CWT is created (Sect. 3) that includes a unique translation for each used HLZ. The translation process is based on Wilhelm et al. (2014) where the translation of the Global Land Cover (GLC) 2006 to the 16 REMO-iMOVE PFTs is described. Since the nomenclature of GLC 2006 and ESA-CCI LC are similar and based on the same classification system some of the CWTs were initially adopted from (Wilhelm et al., 2014). For the more diverse ESA-CCI LC classes new CWTs need to be created. The new CWTs follow the translation of Poulter et al.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

(2015) (ESA-CCI PFTs) but were carefully revised and modified during the process. This revision of the CWTs is supported by reference data and visual satellite image interpretation. The quality of the LANDMATE PFT dataset is finally assessed by comparison to a comprehensive ground truth database (Sect. 4).

2.2 Datasets & concepts

2.2.1 ESA-CCI LC

The European Space Agency Climate Change Initiative (ESA-CCI) provides continuous global land cover maps (ESA-CCI LC) on ~300 m horizontal grid resolution. The ESA-CCI LC maps are available for download in annual time steps for the years 1992-2018 (ESA, 2017). The classification of the LC maps follows the United Nations Land Cover Classification System (UN-LCCS) protocol (Di Gregorio, 2005) and consists of 22 level 1 classes and 14 additional level 2 classes, which include regional specifications. More information on ESA-CCI LC data processing can be found at maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf. An overview of the satellite missions involved in the production of ESA-CCI LC is given in table 1. Besides systematic global validation efforts (ESA, 2017; Hua et al., 2018), a few regional approaches investigated the quality of ESA-CCI LC over Europe (Vilar et al., 2019; Reinhart et al., 2021a).

Table 1. Satellite missions involved in the production of ESA-CCI LC according to ESA (2017)

Time period	Satellite product
Baseline Production 2003-2012	MERIS FR/RR [1] global SR [2] composites
1992-1999	Baseline 10-year global map; AVHRR [3] global SR composites for back-dating baseline
1999-2013	Baseline 10-year global map; SPOT-VGT [4] global SR composites for up and back-dating the baseline; PROBA-V [5] global SR composites at 300 m
2013-2015	Baseline 10-year global map; PROBA-V global SR composites at 1 km for years 2014 and 2015 for up-dating the baseline; PROBA-V time series at 300 m
Since 2016	Sentinel-3 OLCI and SLSTR [6] 7-day composites

[1] MEdium Resolution Imaging Spectrometer Full Resolution/Reduced Resolution (ESA, 2002)
[2] Surface Reflectance
[3] Advanced Very-High-Resolution Radiometer (Hastings and Emery, 1992)
[4] SPOT Vegetation satellite program (Maisongrande et al., 2004)
[5] Project for On-Board Autonomy - Vegetation (Dierckx et al., 2014)
[6] Ocean and Land Colour Instrument (OLCI) and Sea and Land Surface Temperature Radiometer (SLSTR) (Donlon et al., 2012)

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

2.2.2 E-OBS Climate data

The E-OBS dataset (Cornes et al., 2018) is a daily gridded observational dataset, derived from station observations from European countries covering the period from 1950 to 2020. The point observations are interpolated using a spline method with random perturbations in order to produce an ensemble of realizations. For the creation of the HLZs that are used for the conversion of ESA-CCI LC classes to PFTs (Section 2.2.5), the ensemble mean of the 2-meter-temperature (TG) and precipitation (RR) on a regular 0.1° grid from E-OBS version 19.0e is used. It covers most of Europe, some parts of the Middle East and a narrow strip of Northern Africa.

2.2.3 CRU

The Climate Research Unit (CRU) TS 4.03 dataset is a global gridded high-resolution climate dataset based on station observations produced and maintained by the CRU of the University of East Anglia (Harris et al., 2014). The dataset provides global monthly means of climate parameters at 0.5° resolution from 1901 to 2019. In order to achieve the target resolution of 0.1° for the global LANDMATE PFT maps, the CRU climate data is downscaled using bilinear interpolation. Following Hoffmann et al. (2016), distance-weighted interpolation was applied to the atmospheric observation dataset CRU to extrapolate the climate data to the coastlines of the ESA-CCI LC maps in order to compensate for the different land-sea-masks of the products. The CRU climate dataset was used within this application for regions where E-OBS is not available.

2.2.4 Holdridge Life Zones

The Holdridge Life Zone (HLZ) concept was initially developed in 1967 (Holdridge et al., 1967) to define all divisions of the global biosphere, depending on the relation of biotemperature (average of monthly temperature above 0°C; since plant activities are idle below freezing, all values below 0°C are adjusted to 0°C), mean annual precipitation and ratio of potential evapotranspiration to mean annual precipitation. By combining threshold values of biotemperature and annual rainfall, the 38 HLZs are created (Table 2). In the present analysis, the tropical and subtropical as well as the polar and subpolar HLZs are mereged. Through the merging of the aforementioned HLZs, 30 individual HLZs in total are available for the creation of the European HLZ map (Fig. 2). The dynamic character of the specific quantitative ranges of the long-term means of the

Table 2. The Holdridge Life Zones following (Holdridge et al., 1967).

Bio-temperature [°C]	Precipitation [mm]					
	<125	125 to <250	250 to <500	500 to <1000	1000 to <2000	>2000
<3	Subpolar dry tundra	Subpolar moist tundra	Subpolar wet tundra	Subpolar rain tundra	-	-
3 to <6	Boreal desert	Boreal dry shrub	Boreal moist forest	Boreal wet forest	Boreal rain forest	-
6 to <12	Cool temperate desert	Cool temperate desert shrub	Cool temperate steppe	Cool temperate moist forest	Cool temperate wet forest	Cool temperate rain forest
12 to <18	Warm temperate desert	Warm temperate desert scrub	Warm temperate thorn steppe/woodland	Warm temperate dry forest	Warm temperate moist forest	Warm temperate wet/rain forest
18 to <24	Subtropical desert	Subtropical desert shrub	Subtropical thorny steppe/woodland	Subtropical dry forest	Subtropical moist forest	Subtropical wet/rain forest
>24	Tropical desert	Tropical desert shrub	Tropical thorny woodland	Tropical very dry forest	Tropical dry forest	Tropical moist/wet/rain forest

utilized climate parameters make the HLZ classification more flexible than other available global ecosystem classifications and therefore makes the HLZs most suitable for the application presented in this article. In addition the requirement for input data is relatively low.

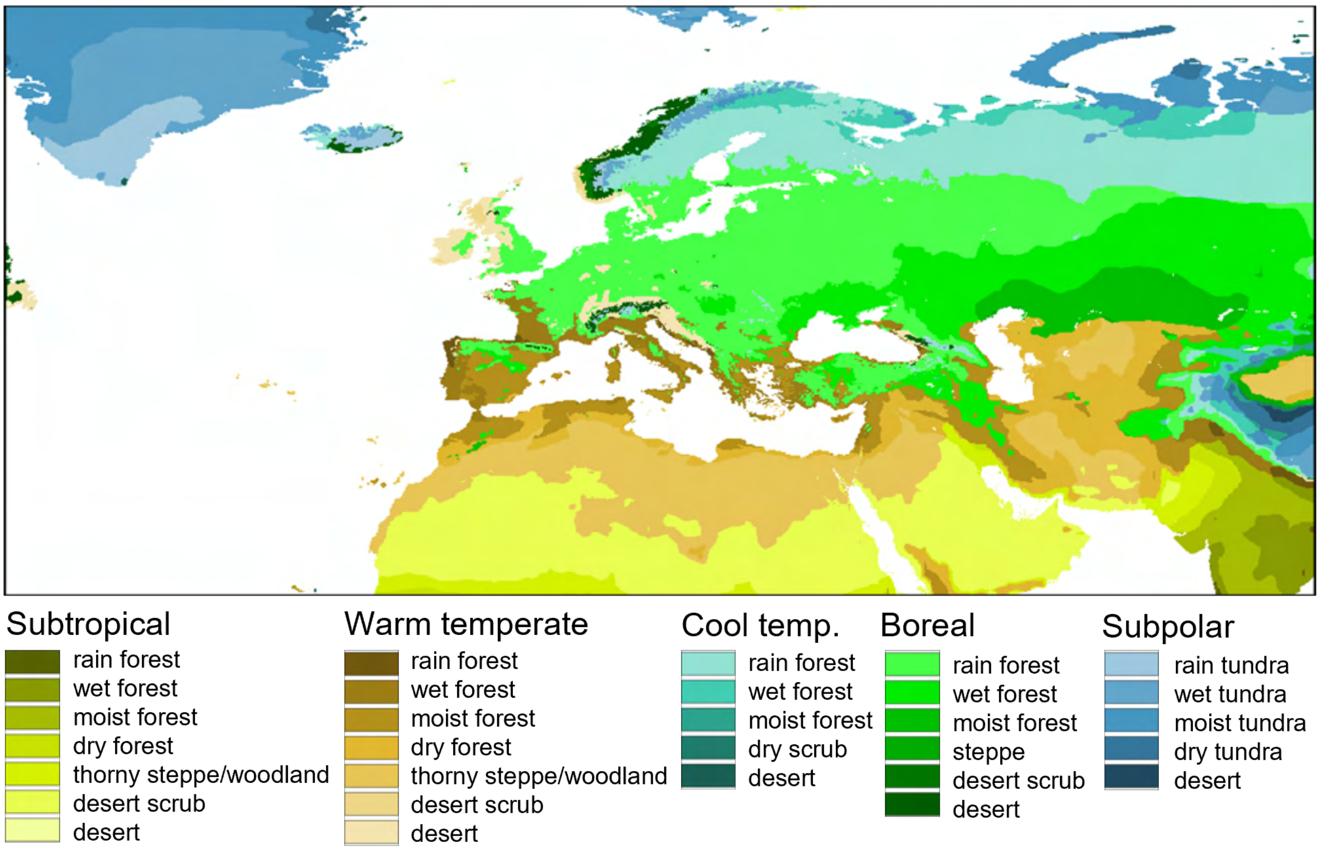


Figure 2. Holdridge Life Zones map for the extent of LANDMATE PFT

In the past, the HLZ concept was not only found useful for global applications but successfully implemented especially for regional mapping approaches due to its ability to capture regional climate features with the support of bioclimatic variables (Daly et al., 2003; Tatli and Dalfes, 2016). Further, the HLZ concept was used for LULC change predictions, such as land use impact assessments, related to current and future climate change scenarios (Chen et al., 2003; Skov and Svenning, 2004; Yue et al., 2006; Saad et al., 2013; Szelepcsényi et al., 2018). With the implementation of climate data through the HLZ concept, the resulting PFT maps become more detailed and can be customized to individual regions without losing global consistency.

2.2.5 Plant Functional Types

Table 3 shows the LANDMATE PFTs that are based on the PFTs introduced by Wilhelm et al. (2014). The implementation of an irrigated cropland PFT (PFT 14) that is currently being developed within the HICSS project LANDMATE will be implemented

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

in a later version of the dataset. In the initial version that is presented in this article, all cropland proportions are assigned to the cropland PFT (PFT 13).

Table 3. LUCAS plant functional types based on Wilhelm et al. (2014) with modified crop types.

PFTs	Names
1	Tropical broadleaf evergreen trees
2	Tropical deciduous trees
3	Temperate broadleaf evergreen trees
4	Temperate deciduous trees
5	Evergreen coniferous trees
6	Deciduous coniferous trees
7	Coniferous shrubs
8	Deciduous shrubs
9	C3 grass
10	C4 grass
11	Tundra
12	Swamp
13	Non-irrigated crops
14	*Irrigated crops* [7]
15	Urban
16	Bare

2.2.6 Potential C4 grass fraction NACP MsTMIP

The initial land cover map from the ESA-CCI LC does not provide a distinction between C3 and C4 grassland. Therefore, an additional product is used after applying the CWP. The map from the North American Carbon Program Multi-scale Synthesis and Terrestrial Model Intercomparison Project (NACP MsTMIP; Wei et al., 2014) is constructed based on the synergetic land cover product (SYNMAP) by Jung et al. (2006). SYNMAP is a combination of multiple high-resolution LULC products using a fuzzy agreement approach. The NACP MsTMIP map uses the grassland fractions from the SYNMAP product and the C4 grass distribution is estimated supported by growing season temperature based on present climate conditions (Wei et al., 2014). The map is provided in 0.5° horizontal grid resolution for the period from 1801 to 2010. For the preparation of LANDMATE PFT the NACP MsTMIP map of 2010 is used.

[7] the irrigated crop PFT is currently empty (see section 3.4)

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

2.3 LUCAS - land use and land cover survey

The harmonized LUCAS *in situ* land cover and use database for field surveys from 2006 to 2018 (d'Andrimont et al., 2020) is the most consistent ground truth database for the European Continent. The survey was carried out at three-yearly intervals between 2006 and 2018. The systematic sampling design of the survey consists of a theoretical, regular grid over the European Continent with ~2 km grid size. The reference point locations are the corner points of the theoretical grid. Not all locations within the survey were easily accessible. Therefore, the survey is supported by in situ photo interpretation, in-office photo interpretation and satellite data in the latest time steps 2015 and 2018 (table 4). However, the main proportion of the reference points was recorded through location visits at all time steps, which makes this land survey the most reliable and consistent ground truth database for Europe.

Table 4. Number and recording method of reference points in the LUCAS land cover and use database per timestep.

Year	Reference points	in situ	in situ PI[8]	in-office PI[9]	GT[10] [%]
2006	168401	155238	13163		92.18
2009	234623	175029	59594		74.6
2012	270272	243603	26669		90.13
2015	340143	242823	25254	71970	71.39
2018	337854	215120	22894	99803	63.67

The extent of the LUCAS survey was increased over time. The 2006 survey covered 11 countries while the 2018 map covers large parts of the European Continent with 28 countries. Throughout the survey, the ground truth data has been continuously checked for quality and plausibility. For the accuracy assessment of the LANDMATE PFT map the ground truth points of the year 2015 are employed (Sect. 4). In order to avoid confusion between the FPS LUCAS and the LUCAS ground truth dataset, the latter will be further referred to as **Ground Truth Survey** or **GT-SUR**.

3 Cross-walking procedure - ESA-CCI LC classes to PFTs

The CWP from ESA-CCI LC classes to PFTs presented in this article is based generally on (1) the translation introduced by Poulter et al. (2015) and (2) the translation by Wilhelm et al. (2014). Both translations are not just combined with each other but modified using additional data. The following sections introduce the PFTs of LANDMATE PFT aggregated into groups and give an overview of the decisions on modifications that are made during the production process based on literature and additional data. The final LANDMATE PFT map is shown in fig. 3.

[8] Photo interpretation close to the reference location
[9] Photo interpretation with supporting data, such as satellite images
[10] Ground truth

3.1 Trees and shrubs, tropical and temperate | PFT 1-8

The LANDMATE PFTs are more diversified regarding tree-PFTs than the generic ESA-CCI PFTs. The expansion of tree-PFTs to six in total was done at the expense of two shrub-PFTs. The increase of tree-PFT diversity is done in order to address the strong biogeophysical impacts of forested areas on regional and local climate, such as decreased albedo and increased roughness length (Bright et al., 2015). The effects of forested areas on near-surface climate are distinctively different to the effects of shrub or grass covered areas, and are also highly depending on tree species composition and latitudinal range (Bonan, 2008; Richardson et al., 2013). Another reason for the six tree-PFTs is the intended use of the PFT maps in RCMs. In the Land Surface Models (LSMs) of current generation RCMs, where a distinction is rather made between different tree or tree community types than between different shrub types. Therefore and with regard to the implementation process that needs to be done for each RCM individually, an increase in the number of tree-PFTs and a decrease in the number of shrub-PFTs is considered to be convenient. Accordingly, the tree and shrub proportions were distributed following both, the needleleaf and broadleaf definitions of the ESA-CCI LC classes as well as the HLZ map, where the HLZ map was decisive for an assignment of forest proportions to the temperate or tropical tree-PFT, respectively. Following a comparison with different forest datasets over Europe (not shown), the tree proportions in the translation of the mixed land cover classes (e.g. lass 61 - Tree cover, broadleaved, deciduous, closed (>40%)) are increased to be in line with the indicated overall forest amount over Europe.

3.2 Grassland | PFT 9 & 10

The generic ESA-CCI PFTs include a natural grassland- and a managed grassland-PFT to include grassland and cropland respectively. The LANDMATE PFTs include two grassland-PFTs, distinguishing between C3 and C4 grass. The contrasting photosynthetic pathways and therefore contrasting synthetic response to CO_2 and temperature determine specific ecosystem functions for both PFTs respectively. The main differences are found in global terrestrial productivity and water cycling (Lattanzi, 2010; Pau et al., 2013). The translation from the LULC classes that contain grassland proportions into C3 or C4 grass-PFTs respectively is supported by a map of potential C4 vegetation by Wei et al. (2014) where the potential global distribution of C4 is estimated using bioclimatic parameters (Sect. 2.2.6).

3.3 Tundra and swamps | PFT 11 & 12

The specific vegetation PFTs tundra and swamps are treated individually in LANDMATE PFT. Tundra is mostly used for the polar and subpolar HLZs, where the climatic conditions require a clear distinction of the land surface properties to the boreal and temperate regions regarding exchange and feedback processes with the atmosphere (Thompson et al., 2004). Chapin Iii et al. (2000) further suggest a differentiation of vegetation composition within these northern vegetation communities, which can also be realized using the introduced translation. The swamp-PFT is mostly used for translating the ESA-CCI LC mosaic tree/shrub/herbaceous classes and also partly for the flooded tree cover classes in most of the HLZs. Swamps occur mainly in the boreal and polar regions.

3.4 Cropland | PFT 13 & 14

Currently, two cropland-PFTs are defined in the LANDMATE PFT map. The cropland-PFT (PFT 13, see table 3) includes all managed, agricultural land surface proportions. The uncertainties of the translation of the ESA-CCI cropland classes and mixed cropland classes into the cropland-PFTs was investigated by Li et al. (2018) where the comparison of LULC change in the ESA-CCI PFT maps against other LULC products showed inconsistencies between global trends and geographical patterns between the products. However, Li et al. (2018) provide a modified CWT that was adjusted in regard to an improved knowledge base on how to translate LULC classes into PFTs for climate models. Particular focus is laid on mosaic classes and the sparsely vegetated classes of which appear numerous in ESA-CCI LC. Therefore, the translation from Li et al. (2018) for cropland is adopted into the present CWP.

The irrigated cropland-PFT (PFT 14, see table 3) is currently empty in the LANDMATE PFT map Version 1.0. This decision is made following intense research on available irrigation information. The ESA-CCI LC map that is used as initial input contains an "irrigated cropland" class but this information was not used in the process. The investigation on irrigated areas included the comparison of ESA-CCI LC to other products that are available, such as the irrigation map from the FAO (Siebert et al., 2005).Although the ESA-CCI LC quality assessment shows a very good agreement of the ESA-CCI LC irrigated cropland with the validation database (ESA, 2017), the comparison showed considerable differences between the products. The success of detection of irrigated areas is highly dependent on the correct detection of the crop types to infer the water needs of the respective crops, on atmospheric and environmental conditions and on the availability of multi-temporal, high resolution imagery (Bégué et al., 2018; Karthikeyan et al., 2020). Further, most remote sensing applications depend highly on ground truth data and local knowledge. Applications using different satellite imagery to detect agricultural management practices, such as irrigation, are only successfully tested and applied in local spatial units (Rufin et al., 2019; Ottosen et al., 2019). Therefore, the irrigated cropland PFT remains unoccupied for now. Nevertheless, PFT 14 is defined within LANDMATE PFT Version 1.0 for the purpose of adding irrigated LULC fractions in the future. For the long term LUCAS LUC dataset (Hoffmann et al., submitted) which is extended backward and forward based on the LANDMATE PFT map for Europe 2015, irrigated cropland areas are already implemented following the irrigated area definition of the Land Use Harmonization (LUH2) dataset (Hurtt et al., 2011).

3.5 Non-vegetated | PFT 15 & 16

The non vegetated-PFTs in the LANDMATE PFT dataset are urban and bare. The urban grid cells from ESA-CCI LC are directly translated into urban fractions for all HLZs in the CWP. The same applies for all bare ground proportions that are translated fully into the bare-PFT. In addition, the ESA-CCI LC mixed classes are split up and the bare ground proportions within the mixed classes are added to the bare-PFT. The explicit treatment of urban areas and especially differentiation from bare ground provides the possibility to resolve urban surface characteristics in RCMs. The treatment of urban areas as a slab surface or as an equal to rock surface as done in several RCM approaches cannot account for the complex geobiophysical

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

processes associated with an urban agglomeration (Daniel et al., 2019; Belda et al., 2018). Due to the distinction of the two surface types, the LANDMATE PFT map can be used for impact studies with an urban focus.

3.6 Water, permanent snow & ice

The LANDMATE PFTs do not include individual PFT definitions for water and snow/ice respectively. Regarding the water representation, most currently used RCMs are utilizing a land-sea-mask to account for oceans and inland water areas. Therefore, an explicit definition of water as individual PFT has not been implemented. Consequently, water grid cells are set to no data. In the present translation, the snow/ice grid cells from ESA-CCI land cover are translated into bare-PFT following Wilhelm et al. (2014).

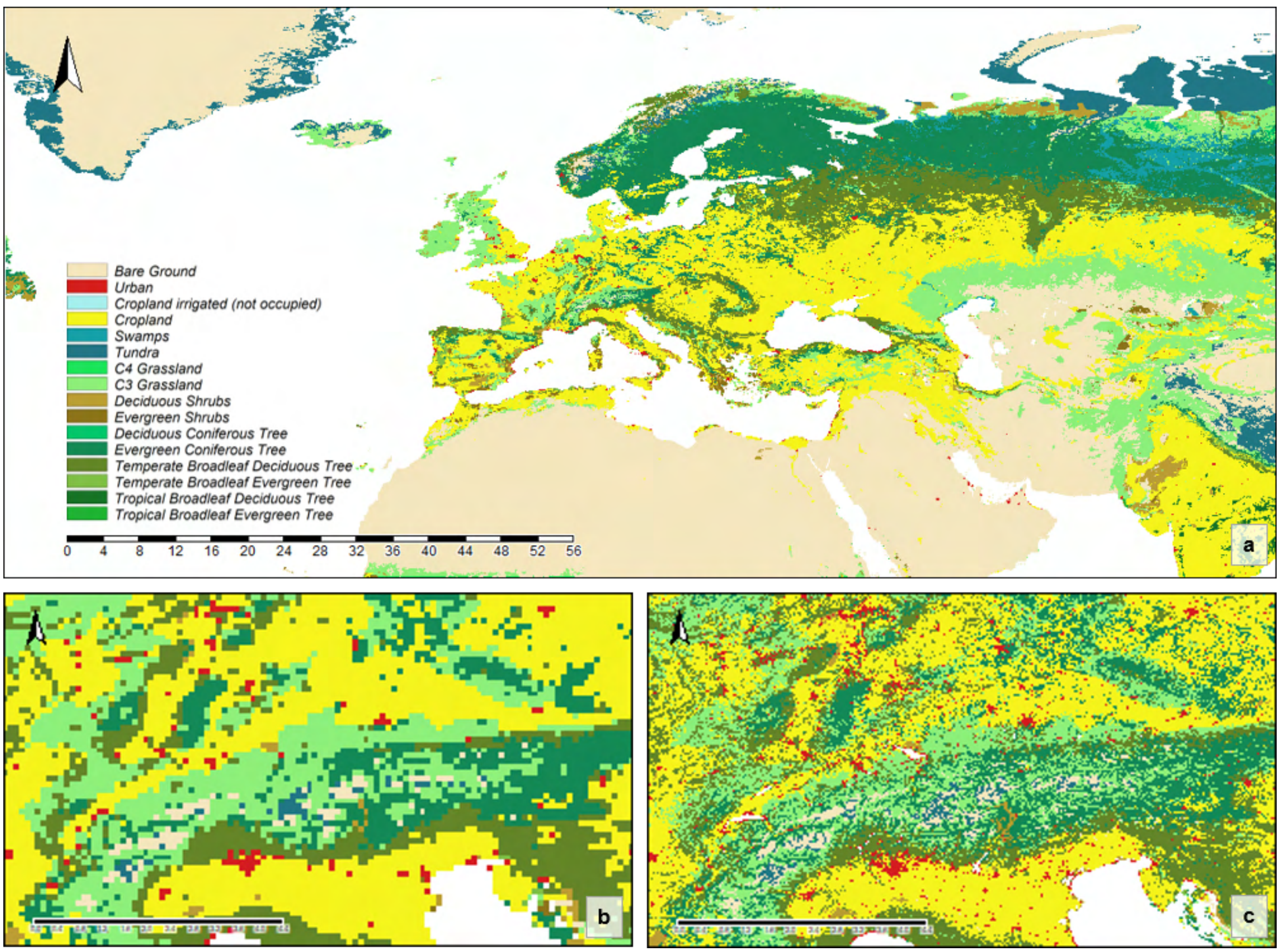


Figure 3. LANDMATE PFT map for Europe for 2015 (a). Below a map section of the Alpine region shows an example of the resolution difference between LANDMATE PFT 0.1 (b) and LANDMATE PFT 0.018 (c). LANDMATE PFT 0.018 is used in the present accuracy assessment. For improved visualization all maps show the majority PFT per grid cell.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

4 Quality assessment of the LANDMATE PFT map

The LANDMATE PFT map is based on the ESA-CCI LC map which was quality checked and compared to similar LULC products on a global (ESA, 2017; Yang et al., 2017; Hua et al., 2018; Li et al., 2018) and regional level (Reinhart et al., 2021a; Vilar et al., 2019). However, the translation from LULC classes to PFTs necessarily results in change of the map. The final product, the LANDMATE PFT map, is intended to be used in RCMs, which means the quality of the final product must be assessed in addition to the available quality assessments of the initial ESA-CCI LC map. In order to overcome the resolution difference, which is non negligible between LANDMATE PFT and the reference data GT-SUR, the LANDMATE PFT map is prepared on 0.018° horizontal resolution, which corresponds closely to the 2 km theoretical grid of GT-SUR.

The design of such a quality assessment of a large scale map product is not trivial, especially since the map product itself and the reference data are often different in structure and nomenclature, given that ground truth reference data is mostly collected as point data and independently from the assessed map product Foody (2002); Wulder et al. (2006); Olofsson et al. (2014). In order to produce reliable quality information for LANDMATE PFT, the present assessment follows closely the well established good-practice recommendations. Nevertheless, adjustments are done to account for the fractional structure of LANDMATE PFT. Section 4.2 provides additional information on the requirements of a "good practice" accuracy assessment, the key components and the selected sampling design and metrics.

4.1 Research area

The coverage of GT-SUR in the year 2015 includes 28 countries which are highlighted in dark grey in fig. 4.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

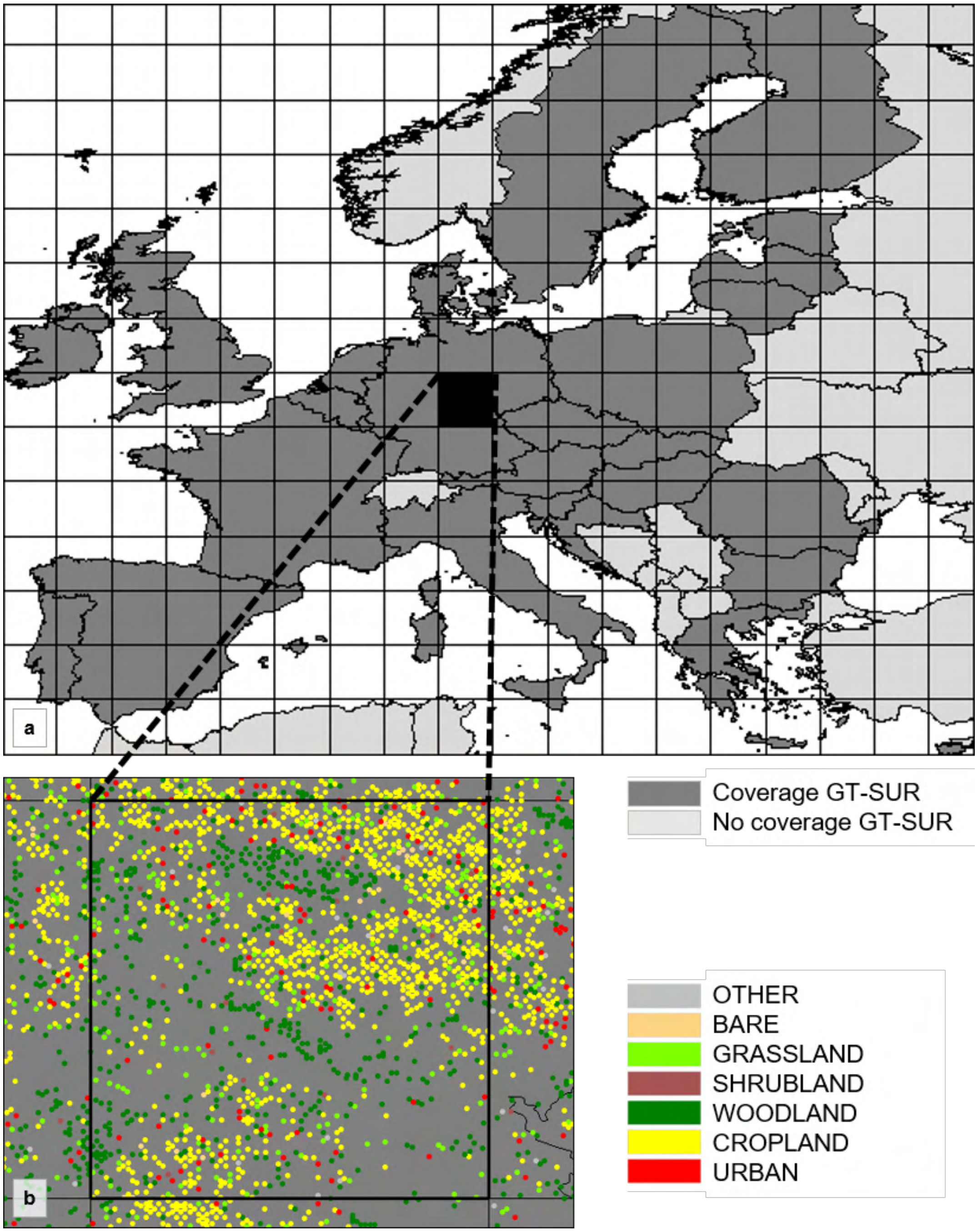


Figure 4. Coverage of the Land Use and Coverage Area frame Survey (LUCAS) for reference year 2015 (top). The lower figure shows the points and LULC group representation within the grid cell highlighted in black color in the top map as an example for the whole research area.

The total number of GT-SUR points for 2015 is 340,143. Out of these points, 338,619 points (~99.55%) are covered with valid LANDMATE PFT grid cells of the assessed LULC groups and can be used in the analysis. Countries located within the contiguous area but missing in the assessment are Switzerland, Norway, the Russian Oblast Kaliningrad, Bosnia and Herzegovina, Montenegro, Albania, Serbia, Kosovo, North Macedonia, and Belarus. Figure 4 also shows the 2.5° grid that was used for the analysis of the accuracy assessment results (Sect. 4.3). Due to the fine scale and the high number of points over the whole research area, the visualization of the spatial analyses on continental scale is challenging. Therefore, the research area is split up through an overlay of a 2.5° grid (as shown in fig. 4). The overall and class-wise accuracy results for all points within each 2.5° grid cell are aggregated in order to identify large scale spatial quality differences for the analyzed LULC groups. Additionally, the total number of points for each LULC group per grid cell are displayed in section 4.3.

4.2 Accuracy assessment - background & design

The key components of the accuracy assessment of a large-scale land cover product are **objective**, **sampling design**, **response design** and the final **analyses and estimation** (Wulder et al., 2006). All of the key components have great impact on the quality of the assessment and further, on the final metrics, especially in the present assessment, where reference and assessed dataset differ widely in structure. LANDMATE PFT is a gridded dataset with fractional LULC classes but no information on the subgrid location within the grid cell. Other than that, the points of GT-SUR have fixed locations expressed through exact coordinates, but no (exact) information on the spatial extent of this class. Another challenge is the fractional structure of LANDMATE PFT itself, where one unit (grid cell) possibly contains multiple fractions. Therefore, the design of the accuracy assessment needs to be customized to the **objective**, which is to determine the overall quality of the LANDMATE PFT map for Europe 2015 as well as the quality of individual LULC type representation within the map in order to derive recommendations for the use of LANDMATE PFTs in RCMs.

When it comes to the **sampling design**, sampling size, spatial distribution of the respective sample and the representation of each LULC group or class within the sample are crucial to produce reliable quality information about a LULC product (Stehman, 2009). However, the collection of ground truth data is a rather expensive procedure regarding time and money, which needs to be considered during the process. The sample size is therefore a compromise size and cost. In the present assessment, an existing ground truth database containing over 340,000 records is used as reference which eliminates the possible issue of a too small sample size. It is also known that all assessed LULC groups are represented in a sufficiently high number (Table 6). Nevertheless, the present assessment is a special case situation with every unit of LANDMATE PFT containing more than one LULC group potentially. Therefore, the subsets are selected through application of a filter to capture the map accuracy in a way that accounts for the fractional structure within the grid cells in the LANDMATE PFT map (see section 4.2.1).

The **response design** deals with the spatial support regions (SSR) and the labelling protocol or classification harmonization. The SSR is a buffer region around a sampling unit that is selected to account for small-scale landscape heterogeneity that is likely not captured by larger scale map products. In the present case, the sampling design is selected in a way that the grid cells of LANDMATE PFT serve as SSR for each GT-SUR point. A fraction is not located precisely at one location within the respective grid cell but evenly distributed over the whole grid cell. Assuming, the uniformly distributed fraction can occur

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

in small patches or in one large patch within the grid cell, the whole grid cell is defined as SSR for the respective LULC group. The labelling protocol needs to be determined to deal with the different legends of the reference and the assessed map. The harmonization of legends is selected in regard to the objective of the respective assessment, as in this case, to provide information about the quality of representation of the most dominant LULC types in LANDMATE PFT. The labelling protocol used in the present assessment is summarized in table 5.

The **analyses and estimation** used are error matrices, that give an overview of the overall and LULC group-wise accuracy of the LANDMATE PFT map. For both resolutions of LANDMATE PFT, the error matrices and the resulting accuracy measures overall accuracy (OA), producer's accuracy (PA) and user's accuracy (UA) are calculated, where PA and OA are calculated group-wise. The error matrix is a cross-tabulation between map and reference of the size q x q, where q stands for the number of land cover classes or groups. The map classes are placed in the rows and the reference classes in the columns so that the diagonal of the matrix gives the sum of the correctly classified map units. The off-diagonal cell values represent the disagreement between the map and the reference. The overall accuracy is calculated according to equation 1:

$$OA = \frac{\sum_{i=1}^{q} n_{ii}}{n} * 100 \tag{1}$$

The sum of the agreeing diagonal elements n_{ii} of all LULC groups is divided by the number of all observations n. The PA represents the accuracy from the view of the map producer. The PA stands for the probability, that a LULC feature in the reference is classified as the respective feature by the map. The PA is calculated using equation 2 where the number of correctly classified units per LULC group n_{ii} is divided by the total number of LULC group occurrences of the reference n_{+i}:

$$PA_i = \frac{n_{ii}}{n_{+i}} * 100 \tag{2}$$

While the PA gives the proportion of features in the reference that are actually represented as those in the produced map, the UA is the accuracy from the perspective of the map user. It is the probability of a feature classified as such in the map is actually present in the reference. The UA is calculated using equation 3, where the number of correctly classified pixels n_{ii} per LULC group is divided by the row sum n_{i+} $\sum_{i=1}^{p} n_{ji}$:

$$UA_i = \frac{n_{ii}}{n_{i+}} * 100 \tag{3}$$

4.2.1 Dataset harmonization & filter

The quality assessment is done assigning the PFT type with the maximum fraction per grid cell to the GT-SUR points located within respective grid cell. The classifications of both datasets need to be harmonized in order to make the comparison as detailed as possible but also to be able to produce reliable and robust results for the RCM community. For the analysis, the classifications of LANDMATE PFT and the GT-SUR are harmonized as shown in table 5.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table 5. Classification harmonization between LANDMATE PFT map and GT-SUR

GT-SUR LC group	GT-SUR group name	LANDMATE PFT number	LANDMATE PFT name	Harmonization group number	Harmonization name
A	Artificial Land	15	Urban	1	URBAN
B	Cropland	13 14	Non-irrigated Crops Irrigatred crops	2	CROPLAND
C	Woodland	1 2 3 4 5 6	Tropical broadleaf evergreen trees Tropical deciduous trees Temperate broadleaf evergreen trees Temperate deciduous trees Evergreen coniferous trees Evergreen deciduous trees	3	WOODLAND
D	Shrubland	7 8	Coniferous shrubs Deciduous shrubs	4	SHRUBLAND
E	Grassland	9 10	C3 Grass C4 Grass	5	GRASSLAND
F	Bare land	16	Bare	6	BARE AREAS
G H Other	Water Wetlands Marine areas	11 12	Tundra Swamps	7	OTHER

The LULC groups URBAN, CROPLAND, WOODLAND, SHRUBLAND, GRASSLAND, and BARE ARES are harmonized without applying modifications to the classifications. The LANDMATE PFTs can easily be grouped or directly adopted while the GT-SUR level one classification (letters A-H) is completely adopted into the harmonized groups. The LANDMATE PFT map is a product developed for the use in RCMs. In general, RCMs implement a land-sea-mask to determine aquatic areas for both, inland and marine water. Therefore, the categories WATER and MARINE areas are neglected in the analyses. The LANDMATE PFTs "Tundra" and "Swamp" can not be assigned with a sufficient agreement to the GT-SUR class definitions. Therefore, the GT-SUR groups water, wetlands and marine areas as well as the LANDMATE PFTs Tundra and Swamps are merged into the group "OTHER" for the assessment. Although the group cannot be evaluated regarding the quality of the

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

LANDMATE PFT map, the group needs to be involved in the assessment to keep the numbers in the assessment correct and reliable for all other groups.

Both datasets are provided in a regular Gaussian grid (WGS84 EPSG:4326) so that no reprojection of the datasets needs to be done for the comparison. The descriptive statistics for each LULC group for the reference GT-SUR and the LANDMATE PFT dataset are summarized in table 6.

Table 6. General information on data in the comparison

LULC group[11]	GT-SUR[12]	LANDMATE PFT 0.018°[13]	Dominant LANDMATE PFT 0.018°[14]
URBAN	14,393	65,000	7,577
CROPLAND	83,295	248,301	136,970
WOODLAND	124,374	277,290	124,437
SHRUBLAND	27,298	302,035	19,790
GRASSLAND	66,541	333,948	44,244
BARE AREAS	10,395	31,756	4,148
OTHER	12,340	28,823	1,470
Sums	338,636		338,636

The LANDMATE PFT dataset includes multiple LULC fractions per grid cell. Accordingly, the area proportion of the dominant LULC group varies widely and thus the likelihood that the GT-SUR point sample falls within this area. The filter applied is categorizing the grid cells regarding the proportion of the dominant LULC group, the higher the threshold, the stricter the filter and the more likely a specific sample falls into the subgrid fraction of the dominant class. The filter set numbers 1-10 are representing the cells containing a minimum of 10 - 100% of the dominant LULC group according to table 7.

[11] LULC group analyzed in the quality assessment
[12] GT-SUR points assigned per LULC group
[13] number of grid cells in LANDMATE PFT that have a share of the respective LULC group >0%
[14] Sum of LANMDATE PFT grid cells where the respective LULC group is represented dominantly

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table 7. Filter sets with varying dominant LULC group share per grid cell from >10% to 100%

Filter set number	fraction size of dominant LULC group	LANDMATE PFT cells within filter set
1	>10 %	338619
2	>20 %	338619
3	>30 %	336703
4	>40 %	311238
5	>50 %	259073
6	>60 %	203343
7	>70 %	137412
8	>80 %	74765
9	>90 %	26993
10	100 %	1449

4.3 Results

In order to show the impact of the applied spatial filter, the spatial distribution of agreement and disagreement of LANDMATE PFT with the reference GT-SUR is investigated. The point counts and percentage agreement are aggregated and averaged, respectively, per 2.5° grid cell. After giving an overview over the overall accuracy measures the individual LULC group results are discussed in the following subsections. Note that due to the low overall point count as shown in table 6, the LULC groups SHRUBLAND and BARE AREAS are discussed together in section 4.3.5. Figure 5 shows the spatial distribution of the filter sets over Europe while fig. 6 shows the overall accuracy for the filter sets.

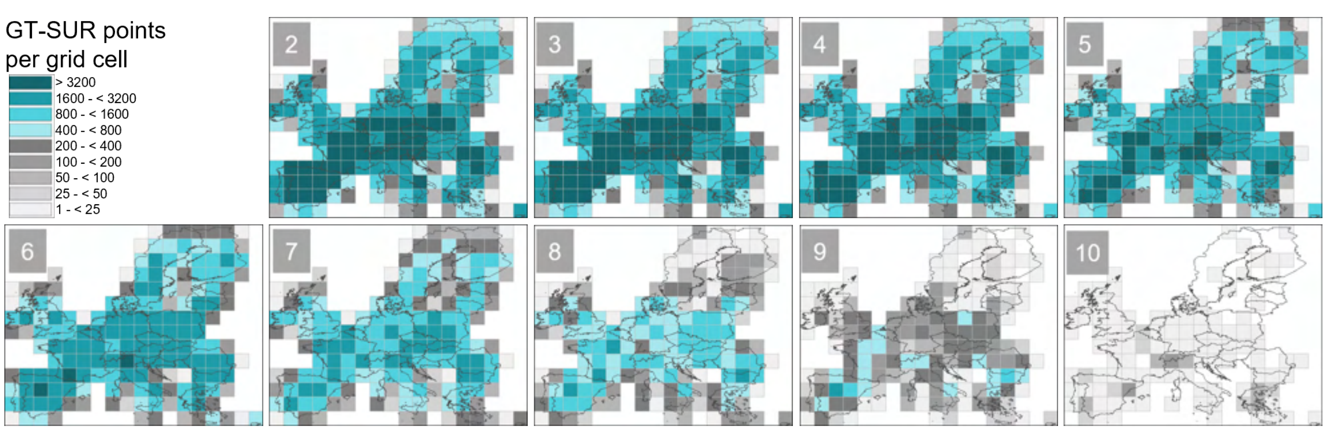


Figure 5. The distribution of the varying dominant LULC group filter sets over the research area in Europe. Since the >10% and the >20% filter set share the same number of points the >10% filter set is not shown.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

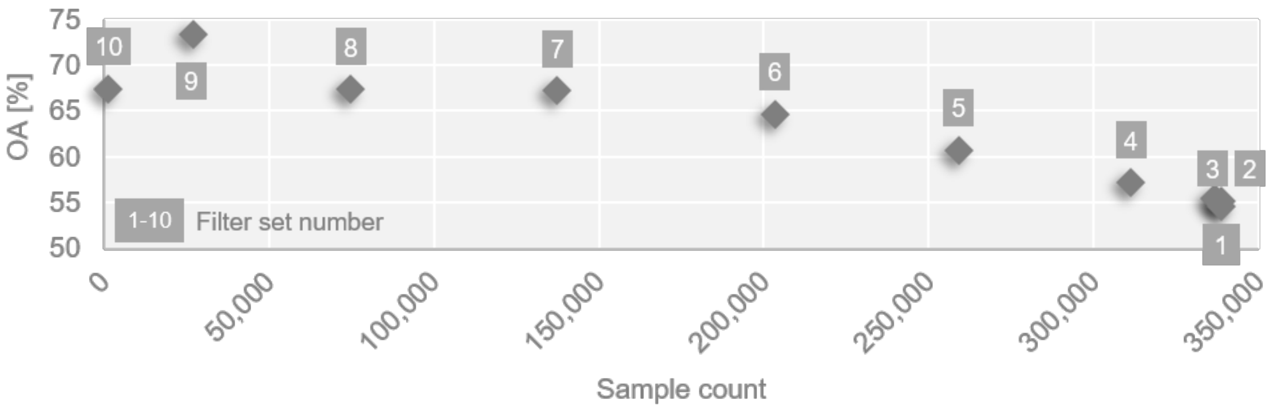
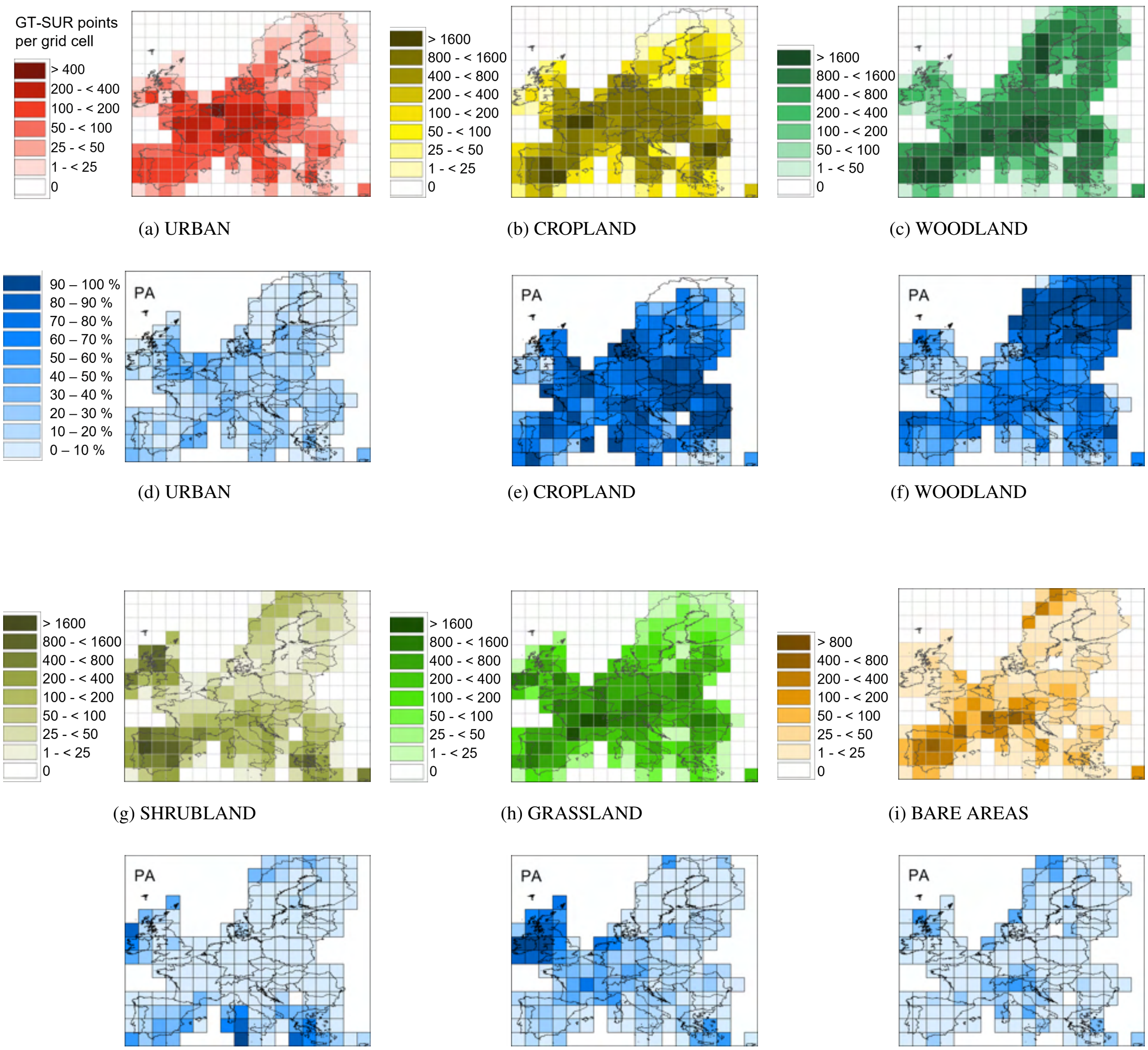


Figure 6. Overall accuracy for the full domain of the 10 filter sets as introduced in table 7 as function of filter set size. Filter set numbers are shown in grey boxes.

In order to be able to capture the LULC group diversity and distribution characteristics of LANDMATE PFT, the filter set must be distributed well over the respective area and contain a sufficiently large proportion of the total cells. Figure 5 shows that the filter sets are distributed reasonably well up to filter set 7. Filter set 8 shows a quite patchy pattern and a strongly decreasing sample number in Northern Europe. Within filter set 9, the patchy pattern of low sample count per 2.5° grid cell spreads over the whole continent. While filter set 9 could still be used for evaluation of LANDMATE PFT for limited regions in Europe, filter set 10 is clearly not evaluable due to the overall small sample count (< 1500). This pattern is also found for filter sets 9 & 10 of the individual LULC groups. The filter set sizes as well as the applied filter itself have direct impact on the OA shown in fig. 6. The decreasing OA towards the higher sample count is an effect of the LANDMATE PFT grid cell heterogeneity representation in each filter set. Filter set 1-3 include all LANDMATE PFT grid cells where the dominant LULC group occupies a minimum of 10% to 30% respectively. Therefore, the probability that the GT-SUR point sample within the respective grid cell represents a location that is occupied by one of the non-dominant LULC groups is relatively high. The applied filter accounts for the impact of the structure difference of the two datasets. The higher probability of agreement is reflected in the increasing OA for the samples that include only grid cells with an occupation of 50% or more of the dominant LULC group. Sample 9 & 10 represent the LANDMATE PFT dataset not adequately regarding distribution and diversity while sample 8 shows a poor coverage in northern Europe. In order to include the largest proportion of the total sample in the analysis, the point count per LULC group as well as the PA per 2.5° grid cell for filter set 2 is analyzed in the results section (fig. 7). In order to give an overview of the spatial accuracy for the evaluable filter range, The respective figures for filter set 5 and 7 are shown in Appendix B (tables B1 & B2).

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021



(a) URBAN (b) CROPLAND (c) WOODLAND

(d) URBAN (e) CROPLAND (f) WOODLAND

(g) SHRUBLAND (h) GRASSLAND (i) BARE AREAS

(j) SHRUBLAND (k) GRASSLAND (l) BARE AREAS

Figure 7. Total count of GT-SUR points per 2.5° grid cell (a-c; g-i) and producer's accuracy for the individual LULC groups (d-f;j-l) for filter set 2 (dominant LULC group occupies > 20% per LANDMATE PFT grid cell)

4.3.1 URBAN

The urban representation in LANDMATE PFT for filter set 2 is shown in fig. 7a and 7d. The PA for the filter sets 1-10 is shown in fig. 8 where an overall low PA for all filter sets is found. With increasing proportion of the dominant LULC group URBAN

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

the PA increases slightly but is still lower than 40% for samples that include enough points to be considered representative for the research area. The overall low PA is reflected in the URBAN maps in fig. 7 as well as in fig. B1 and B2.

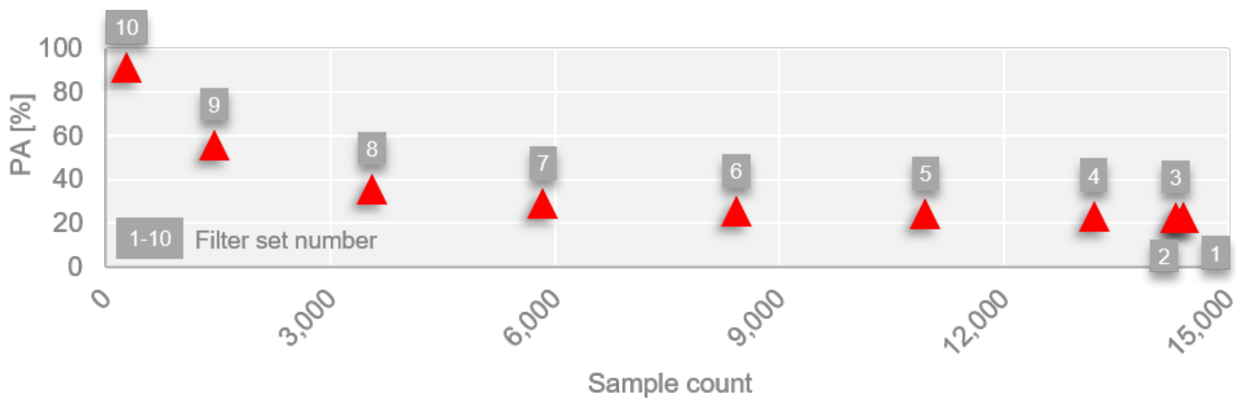


Figure 8. Producer's accuracy of the 10 filter sets (Filter set numbers in grey boxes) for the LULC group URBAN as a function of sample count per filter set.

A visual check of the map agreement between LANDAMTE PFT and GT-SUR revealed the issue that leads to the overall low PA. Figure 9 shows four large URBAN agglomerations in different areas of Europe where the red points represent GT-SUR urban points while the white points represent GT-SUR point representing non-urban LULC groups. The grey-scaled squares represent the LANDMATE PFT URBAN fractions from zero (no coverage, white) to one (full coverage, black) within one grid cell.

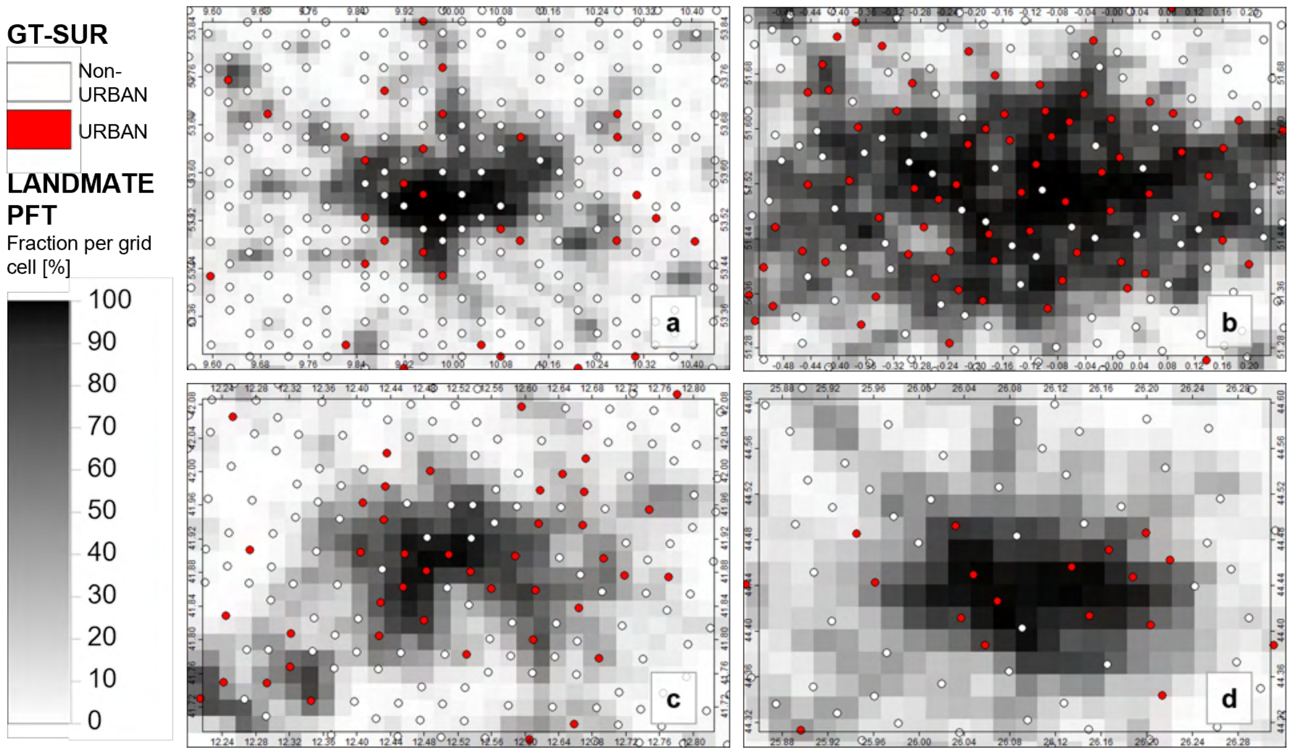


Figure 9. Examples of URBAN representation in LANDMATE PFT (greyscale grid) and GT-SUR (points). Cities shown are Hamburg (a), London (b), Rome (c) and Bukarest (d).

The LANDMATE PFT grid cells with a large urban fraction indicate the respective city core while the GT-SUR points that are located within the city core are mostly not classified as URBAN. However, the GT-SUR points do not fail to represent the structure of urban areas because they are characterized through a heterogeneous pattern of sealed surfaces, recreational areas (e.g. parks) and different building types and density, not through a homogeneous sealed area. The LANDMATE PFT map represents this heterogeneous structure through the varying fractions of non-urban PFTs within the grid cell. However, in order to make the impact of a larger city visible in an RCM simulation, it is beneficial for LANDMATE PFT to represent a larger city with a dense core structure. In order to verify the representation of the large URBAN agglomerations in Europe, a comparison with the World Settlement Footprint for 2015 (WSF, Marconcini et al., 2020) dataset was done (not shown). The comparison showed that not only larger agglomerations but also smaller patches of settlements are represented well in LANDMATE PFT. Therefore, despite the low agreement with GT-SUR in the present assessment, the URBAN PFT of LANDMATE PFT 2015 is of sufficiently good quality and suitable to represent urban land cover in high resolution (~2 km) RCM simulations. Due to the abovementioned comparability issues the UA of the LULC group URBAN will not be further discussed.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

4.3.2 CROPLAND

The CROPLAND representation in LANDMATE PFT shows, together with WOODLAND the highest PA for the research area. As shown in fig. 10 the PA for all filter sets is > 80% which is to be considered as a very good agreement with the reference.

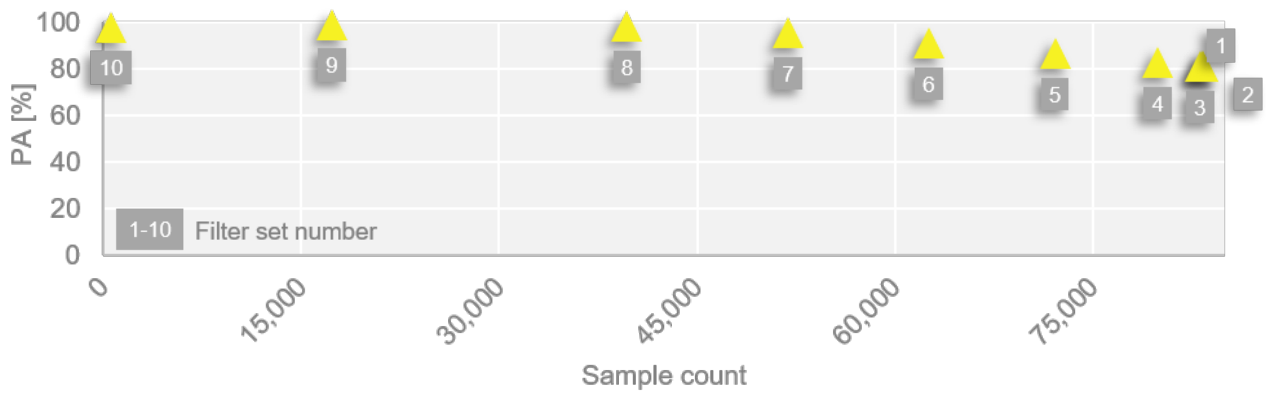


Figure 10. Producer's accuracy of the 10 filter sets (Filter set numbers in grey boxes) for the LULC group CROPLAND as a function of sample count per filter set.

Figure 7b shows the distribution of CROPLAND points in GT-SUR over the research area. CROPLAND points are the second most frequent LULC group in GT-SUR and are mainly distributed over middle and southern Europe. Although the northern European grid cells show a lower count of CROPLAND points, figure 7e shows that the PA is still very high in these areas. The PA increases with increasing filter set homogeneity (Fig. B1 and B2). Regarding the UA for CROPLAND, LANDMATE PFT shows a strong overestimation, where ~51% of the LANDMATE PFT CROPLAND cells in filter set 2 are actually another LULC group in the reference. More than half of the LANDMATE PFT CROPLAND areas are mostly WOODLAND, GRASSLAND, and a mix of the other LULC groups in the reference. The UA for CROPLAND increases rapidly towards the more homogeneous filter sets (~61% for filter set 7). However, the confusion with WOODLAND and GRASSLAND is non-negligible and will be discussed in section 6.

4.3.3 WOODLAND

For the representation of WOODLAND, the PA shows the second highest values with > 70% for all filter sets with a reasonably high point count (filter sets 1-7, fig. 11). Similar to CROPLAND, the sampling filter does not have a large impact on PA. The highest PA is reached over the northern European regions (Fig. 7f). Deficits are visible over the southern British Isles, some parts of France and the coastline along Belgium and the Netherlands. Further, the Mediterranean Coast shows a low PA within grid cells that have an overall small point count (Fig. 7c).

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

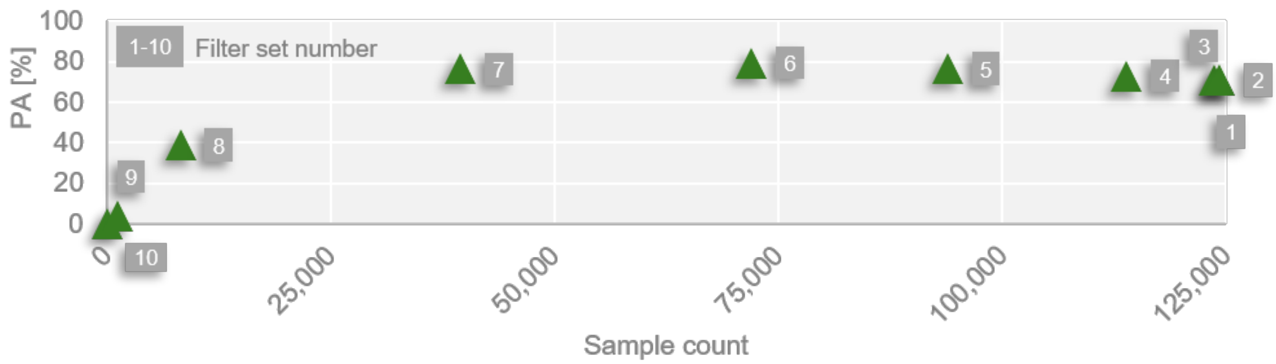


Figure 11. Producer's accuracy of the 10 filter sets (Filter set numbers in grey boxes) for the LULC group WOODLAND as a function of sample count per filter set.

The differences between northern and southern regions tends to increase towards the more homogeneous filter sets as shown in figures B1f and B2f. Agreement over the northern regions increases while agreement over the Iberian Peninsula decreases together with a rapid decrease of the filter set count within the corresponding grid cells. The UA for WOODLAND is noticeably higher than for all other LULC groups (> 70% for filter set 2 and increasing towards the more homogeneous filter sets) which emphasises the very good quality of WOODLAND representation in LANDMATE PFT. (~10% for filter set 2). Further, ~4% of the total LANDMATE PFT cells representing WOODLAND are actually CROPLAND or OTHER.

4.3.4 GRASSLAND

The GT-SUR sampling points show the highest GRASSLAND coverage in central Europe with the highest occurrence in Ireland and the southern part of France (Fig. 7h). The PA for LANDMATE PFT GRASSLAND according to fig. 7k is not noticeably higher in these areas but overall highest in the Southwest of the British Isles. For all filter sets, the PA ranges between 32 and 34% which is considerably low (Fig. 12.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

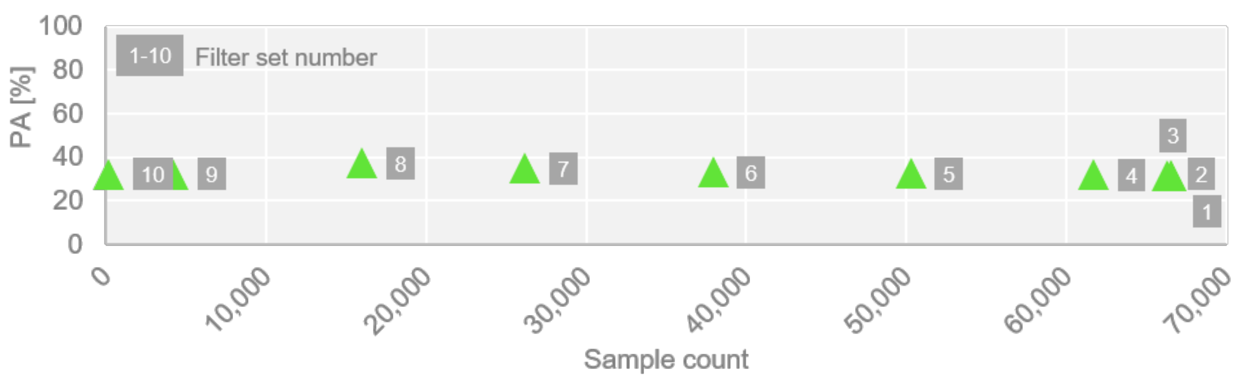


Figure 12. Producer's accuracy of the 10 filter sets (Filter set numbers in grey boxes) for the LULC group GRASSLAND as a function of sample count per filter set.

One reason for this low accuracy of LANDMATE PFT regarding GRASSLAND can be found looking at the results of sections 4.3.2 and 4.3.3. The UAs of CROPLAND and WOODLAND reveal that ~20% of the LANDMATE PFT CROPLAND cells and ~10% of the LANDMATE PFT WOODLAND cells are actually representing GRASSLAND in the reference, which adds up to over 60% of the total GT-SUR GRASSLAND points. Another reason is found in the dataset structure of LANDMATE PFT. A considerable amount of GRASSLAND is not part of the assessment because GRASSLAND does not make the dominant but the second dominant PFT in many grid cells (~45% of all LANDMATE PFT grid cells). Therefore,the seemingly weak GRASSLAND representation in LANDMATE PFT rather shows a weakness of the present assessment that is caused by the different dataset structures.

4.3.5 SHRUBLAND & BARE AREAS

The PA for SHRUBLAND and BARE AREAS is the lowest of all assessed LULC groups with < 20% for all filter sets of both LULC groups respectively (Fig. 13 and 14). The low point count of both LULC groups might be one reason for the low PA. However, looking at the distribution of the SHRUBLAND and BARE AREA points in fig. 7i, LANDMATE PFT is not able to capture the LULC groups even in grid cells with a relatively high point count. The GT-SUR shows ~27,000 SHRUBLAND points while LANDMATE PFT shows only ~19,000. Therefore, one reason for the poor SHRUBLAND representation lies within the base map (ESA-CCI LC) used for the creation of LANDMATE PFT, where the known small count of SHRUBLAND proportions was inherited by LANDMATE PFT. It must be noted, that a large proportion of SHRUBLAND in ESA-CCI LC is part of the mixed LC classes, such as Shrubland/Cropland or Shrubland/Forest. The known deficit was partly compensated by the translation into the PFTs, where SHRUBLAND proportions were added to the total as proportions of the mixed ESA-CCI LC classes. Further SHRUBLAND makes the second dominant PFT in ~20% of the total LANDMATE PFT grid cells in the assessment. Just like for GRASSLAND, these SHRUBLAND proportions can not be addressed within the present assessment.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

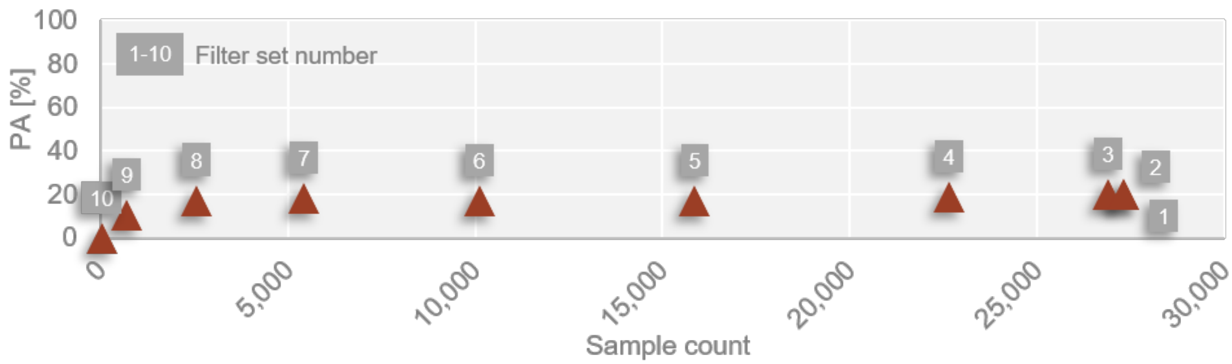


Figure 13. Producer's accuracy of the 10 filter sets (Filter set numbers in grey boxes) for the LULC group SHRUBLAND as a function of sample count per filter set.

The overall BARE AREAS sample count in LANDMATE PFT in filter set 2 is < 50% of the actual BARE AREA points in GT-SUR. Almost half of the GT-SUR BARE AREAS points are identified as CROPLAND while ~30% are identified as WOODLAND or GRASSLAND. Only ~17% (< 2,000 points for filter set 2) of the GT-SUR BARE AREAS are actually identified by LANDMATE PFT with the largest agreement in the Alps, Northern Great Britain, and Northern Scandinavia (Fig. 7l. However, due to the comparably low sample count the spatial assessment is not robust. Just like for SHRUBLAND, the homogeneity of LANDMATE PFT cells does not have a large impact on the PA. UA is higher than PA with ~43% and increasing towards the more homogeneous filter sets. However, considering the rapidly decreasing sample count for the more homogeneous filter sets, the accuracy measures are becoming even less representative for the BARE AREA representation in LANDMATE PFT. Nevertheless, the confusion with the other LULC groups is further discussed in section 6.

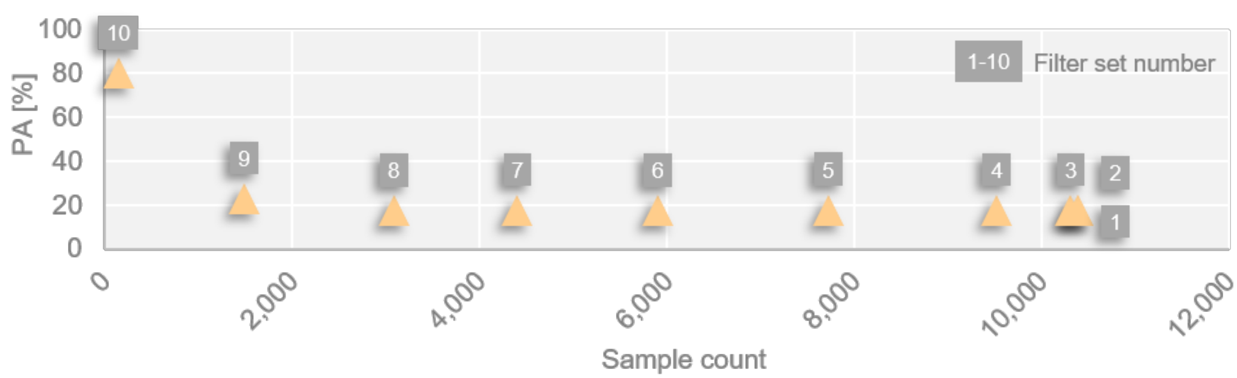


Figure 14. Producer's accuracy of the 10 filter sets (Filter set numbers in grey boxes) for the LULC group BARE AREAS as a function of sample count per filter set.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

5 Data availability

The LANDMATE PFT dataset for Europe 2015 is published with the Long Term Archiving Service (LTA) for large research datasets, which are relevant for climate or earth system research, of the German Climate Computing Service (DKRZ). As World Data Center for Climate (WDCC), the DKRZ LTA is accredited as regular member of the World Data System. The LANDMATE PFT dataset for Europe 2015 is available within the LANDMATE project data at https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=LM_PFT_LandCov_EUR2015_v1.0 (Reinhart et al., 2021b). Within the LANDMATE project, a short documentation summarizes the technical information corresponding to LANDMATE PFT.

6 Discussion & conclusion

The present work introduces the preparation of the LANDMATE PFT map for the European Continent based on several LULC datasets and climate data.

The LANDMATE PFT Version 1.0 is prepared in order to provide realistic, high-resolution LULC representation for RCMs. The dataset includes LULC information from different, validated sources as well as regional climate information through involvement of the HLZs. For each ESA-CCI land cover class, an individual CWT is developed to translate the original LULC classes into PFTs. The various mixed LULC classes included in the base map ESA-CCI LC are extremely difficult to resolve within RCMs. Through the developed CWP, the mixed LULC classes can be disaggregated into PFT fractions, which improves the realistic representation of these classes in RCMs. The involvement of the climate data further allows a customized translation of LULC classes for individual regions. The 16 LANDMATE PFTs are selected to provide simple transferability into various RCM families in order to be able to conduct coordinated RCM experiments where the implementation of a common, high quality LULC map provides minimum uncertainty for a multi-model ensemble.

The accuracy assessment of LANDMATE PFT is conducted in the form of a comparison with the ground truth dataset GT-SUR. In order to account for the different structure of the reference GT-SUR and the assessed LANDMATE PFT map and further the fractional structure of the LANDMATE PFT grid cells, a filter is applied. All filtered LANDMATE PFT subsets are analyzed in terms of agreement with the reference (i.e., GT-SUR). In order to investigate regional differences in accuracy measures, a spatial analysis supported by gridded maps of the research area is done. The quality of the LANDMATE PFT map is assessed using the overall accuracy (OA) and the producer's and user's accuracy (PA and UA) for the individual LULC groups. Overall, the assessment serves as recommendation and uncertainty information for regional climate modellers that use LANDMATE PFT, or the time series LUCAS LUC (Hoffmann et al., submitted), which is based on LANDMATE PFT, in RCMs.

Within the accuracy assessment, the OA does not change considerably between the evaluable filter sets of the respective LULC groups which shows that the dataset structure has no noticeable impact on that accuracy measure. The highest PA is found for CROPLAND and WOODLAND which are the dominant LULC groups in the research area. The lowest PA is found for SHRUBLAND and BARE AREAS, which are also the LULC groups with the lowest overall sample count. The UA is found to be highest for WOODLAND, followed by CROPLAND, GRASSLAND and BARE AREAS. Both accuracy measures, PA

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

and UA are highly influenced by the proportion of the dominant LULC group in the individual grid cell. The difference between the filter sets for UA of the LULC groups is 10 to 20% per group while the difference for PA is noticeable but considerably lower, which means that the applied filter has a higher influence on the former.

The URBAN representation in LANDMATE PFT represents a special case in the present assessment due to the heterogeneous structure of urban areas. Both datasets, GT-SUR and LANDMATE PFT are able to represent the LULC group URBAN very well for their respective purpose. Nevertheless, the PA for URBAN reflects the limitations of the present assessment method. The fine scale point data of GT-SUR represents the patchwork structure of recreational areas, building blocks, and other urban elements at the location of the respective points while LANDMATE PFT represents the urban area as an agglomeration of grid cells with URBAN as the dominant LULC group. The additional comparison with a high resolution dataset (WSF2015) showed that not only large but also small agglomerations of urban areas are represented well in LANDMATE PFT. Therefore and despite of the accuracy assessment results for the LULC group URBAN, the LANDMATE PFT dataset can be recommended to be used in RCMs that resolve urban features over the European Continent.

A limitation of LANDMATE PFT is the overestimation of CROPLAND to the expense of WOODLAND and GRASSLAND and the overestimation of WOODLAND to the expense of mostly GRASSLAND. This overestimation has a minor impact on the overall WOODLAND and CROPLAND representation but a major impact on the representation of GRASSLAND in LANDMATE PFT. The representation of GRASSLAND is comparably low due to the aforementioned reasons. Further, the LULC groups with the lowest point counts SHRUBLAND and BARE AREAS are not well represented, which happens due to the low overall sample size but also due to the overall too low representation in LANDMATE PFT, which is partly inherited from the base map ESA-CCI LC. The representation of these LULC groups needs to be considered when using LANDMATE PFT in RCM simulations using the supporting maps in fig. 7,B1 and B2.

The representation of LULC groups in LANDMATE PFT is assessed through the comparison with ground truth data. The structural differences of the datasets, where gridded data is compared to point data, is a major weakness of this assessment. Although the fractional structure does not have a major influence on the OA, the LULC group-wise PA and even more the UA is affected.

The present assessment takes into account the dominant LULC group per grid cell of LANDMATE PFT. Depending on the proportion of this LULC group, the second or third-most represented LULC group can occupy a considerable area of the respective grid cell. Therefore, a follow up assessment, where these LULC group proportions are also considered and compared to the ground truth is needed in order to investigate, if the PA of the less dominant LULC groups GRASSLAND, SHRUBLAND, and BARE AREAS is increased. The use of additional LULC data, like it was done for URBAN in this assessment, would be an additional useful step to validate the quality of GRASSLAND, SHRUBLAND and BARE AREAS representation in LANDMATE PFT.

The results show that the LANDMATE PFT map is able to represent LULC over large parts of Europe in a sufficient quality. Especially the dominant LULC groups are represented overall well which is highly beneficial for RCM experiments that require realistic, high-resolution LULC representation. Nevertheless, there are uncertainties found for the less represented LULC groups. When using LANDMATE PFT in an RCM it is crucial to consider these uncertainties when interpreting simu-

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

lation results. Especially the spatial distribution of uncertainties in LANDMATE PFT needs to be considered when comparing simulation results to observations because the input parameters in the employed land-surface schemes are influenced by the individual LULC, which subsequently considerably impacts on lower-atmosphere processes, such as the intensity of heat and moisture exchange. Thus, by carefully considering the issue of uncertainty introduced by the LULC input, misconclusions about RCM model performance and about small-scale interconnections can be avoided (Ge et al., 2007; Sertel et al., 2010; Santos-Alamillos et al., 2015; Reinhart et al., 2021a).

Beside the quality of the LULC product, the implementation process of each individual RCM is crucial for the realistic representation of LULC in regional climate model experiments. When translating a LULC product into the model specific LULC classes and structure, modifications are done that can change the map characteristics. When the LANDMATE PFT product is used in an RCM that only uses the dominant LULC fraction per grid cell, the overall LULC proportions can change. The same applies when LANDMATE PFT is used in a model with limited fractions per grid cell or a different classification system. The present assessment gives a guideline on the quality of LANDMATE PFT (Version 1.0) when used unaltered. Through the involvement of the ground truth data, regional deficits of LANDMATE PFT are presented that can be compensated during the implementation process into the individual RCM or RCM family.

The findings of the present assessment support the identification of uncertainties within the LANDMATE PFT map for Europe. Nevertheless, user feedback is crucial for the future overall improvement of LANDMATE PFT. The RCM community within the WCRP FPS LUCAS is already participating in the feedback process where implementation of LANDMATE PFT and the LUCAS LUC time series into different RCMs is comprehensively documented. The future work on LANDMATE PFT also includes the extension of the dataset to other CORDEX regions. Although, the dataset is based on various globally available datasets and therefore, can be created globally, the introduced quality assessment method must be performed for each region individually, desirably using region-specific expert knowledge. Further, the assessment should be expanded in order to include the second or third-most represented LULC group per grid cell to possibly achieve more accurate quality information about LANDMATE PFT.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Appendix A

Table A1. Cross-walking table for ESA-CCI LC class 10 - Cropland, rainfed and LC class 11 -Cropland, herbaceous cover. For LC class 10 and 11, no HLZ were assigned

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30									10				90			

Table A2. Cross-walking table for ESA-CCI LC class 12 - Cropland, tree or shrub cover. For LC class 12, no HLZ were assigned

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30								70					30			

Table A3. Cross-walking table for ESA-CCI LC class 20 - Cropland, irrigated or post flooding. For LC class 20, no HLZ were assigned

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30													100			

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A4. Cross-walking table for ESA-CCI LC class 30 - Mosaic cropland (>50%) / natural vegetation (tree, shrub, herbaceous cover)(<50%).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											20	20	60			
7-9									40				60			
10					10				30				60			
11,12					30				10				60			
13,14									40				60			
15				5	5			20	10				60			
16				7.5	7.5			10	15				60			
17,18				20				10	10				60			
19									40				60			
20							20		20				60			
21,22				10	10		10		10				60			
23,24				10	10		20						60			
25									40				60			
26							20		20				60			
27		20					10		10				60			
28		10					15		15				60			
29	15						10		15				60			
30	20							10	10				60			

Table A5. Cross-walking table for ESA-CCI LC class 40 - Mosaic natural vegetation (tree, shrub, herbaceous cover)(>50%) / cropland(<50%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1,2											35	30	35			
3-5											30	35	35			
6											25	40	35			
7									60				40			
8					10				50				40			
9,10					15				45				40			
11					20				40				40			
12					30			20	10				40			
13				10	10			10	30				40			
14,15				20	20			10	10				40			
16				25	20				15				40			
17				25	25				10				40			
18				30	30								40			
19									60				40			
20							35		25				40			
21			20		15		15		10				40			
22			25		10		15		10				40			
23,24			20		20		20						40			
25									60				40			
26							30		30				40			
27		10					50						40			
28		40					20						40			
29	40						20						40			
30	50						10						40			

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A6. Cross-walking table for ESA-CCI LC class 50 - Tree cover, broadleaved, evergreen, closed to open (>15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6			12.5				12.5				75					
7-18			90	10												
19-24			100													
25-30	100															

Table A7. Cross-walking table for ESA-CCI LC class 60 - Tree cover, broadleaved, deciduous, closed to open (>15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6								100								
7-24				70				15	15							
25-30		70						15	15							

Table A8. Cross-walking table for ESA-CCI LC class 61 - Tree cover, broadleaved, deciduous, closed (>40%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6								85	15							
7-24				70				15	15							
25-30		70						15	15							

Table A9. Cross-walking table for ESA-CCI LC class 62 - Tree cover, broadleaved, deciduous, open (15-40%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6								65	35							
7-24				30				25	45							
25-30		30						25	45							

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A10. Cross-walking table for ESA-CCI LC class 70 - Tree cover, needleleaved, evergreen, closed to open (>15%) and LC class 71 - Tree cover, needleleaved, evergreen, closed (>40%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6					35	35	15		15							
7-18					70		10	5	15							
19-24			35		35		10	5	15							
25-30					70		10	5	15							

Table A11. Cross-walking table for ESA-CCI LC class 72 - Open (15-40%) needleleaved deciduous or evergreen forest (>5m)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6					15	15	25		45							
7-18					30		20	5	45							
19-24			15		15		20	5	45							
25-30					30		20	5	45							

Table A12. Cross-walking table for ESA-CCI LC class 80 - Tree cover, needleleaved, deciduous, closed to open (>15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30						50	5	15	30							

Table A13. Cross-walking table for ESA-CCI LC class 81 - Treecover, needleleaved, deciduous, closed (>40%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30						70		15	15							

Table A14. Cross-walking table for ESA-CCI LC class 82 - Tree cover, needleleaved, deciduous, open (15-40%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30						30	5	20	45							

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A15. Cross-walking table for ESA-CCI LC class 90 - Tree cover, mixed leaf type (broadleaved and needleleaved)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-12				20	70				10							
13-24				70	20				10							
25-30	45	45							10							

Table A16. Cross-walking table for ESA-CCI LC class 100 - Mosaic tree and shrub (>50%) / herbaceous cover(<50%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1					30		30				30	10				
2,3					30		25				25	20				
4-6					30		20				20	30				
7-9				20	20		20		40							
10				25	25		20		30							
11				30	30		20		20							
12				30	30		25		15							
13				15	15			35	35							
14				20	20			30	30							
15				25	25			25	25							
16-18				25	25			30	20							
19,20			30				30		40							
21,22			35				35		30							
23,24			40				30		30							
25		20					50		30							
26		25					50		25							
27		30					45		25							
28		40					35		25							
29		60					20		20							
30		70					15		15							

Table A17. Cross-walking table for ESA-CCI LC class 110 - Mosaic herbaceous cover (>50%) / tree and shrub (<50%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6							50				45	5				
7				10	10		20		60							
8				10	20		10		60							
9				25	25				50							
10				30	30				40							
11,12				35	35				30							
13				15	15				70							
14,15				20	10				70							
16				30				10	60							
17,18				35				15	50							
19			15				15		70							
20			10				20		70							
21			20				10		70							
22			30				10		60							
23,24			35				15		50							
25		15					15		70							
26		20					10		70							
27		25					15		60							
28		30						10	60							
29		40						10	50							
30		50						10	40							

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

Table A18. Cross-walking table for ESA-CCI LC class 120 - Shrubland

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6							40				55	5				
7-12							10	50	40							
13							70		30							
14							40	30	30							
15							20	60	20							
16							20	70	10							
17,18							10	80	10							
19							10		90							
20							50		50							
21							90		10							
22							80	10	10							
23,24							100									
25							10	10	80							
26,27							20	60	20							
28							10	70	20							
29,30							10	80	10							

Table A19. Cross-walking table for ESA-CCI LC class 121 - Evergreen shrubland and LC class 122 - Deciduous Shrubland

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6							40				55	5				
7-12							60		40							
13,14							70		30							
15							80		20							
16-18							90		10							
19							10		90							
20							50		50							
21,22							90		10							
23,24							100									
25							20		80							
26-28							80		20							
29,30							90		10							

Table A20. Cross-walking table for ESA-CCI LC class 122 - Evergreen shrubland and LC class 122 - Deciduous Shrubland

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6								40			55	5				
7-12								60	40							
13,14								70	30							
15								80	20							
16-18								90	10							
19								10	90							
20								50	50							
21,22								90	10							
23,24								100								
25								80	20							
26-28								80	20							
29,30								90	10							

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A21. Cross-walking table for ESA-CCI LC class 130 - Grassland

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											90	10				
7-13									100							
14							5		95							
15								7.5	92.5							
16								10	90							
17								12.5	87.5							
18								15	85							
19									100							
20,21							5		95							
22							7.5		92.5							
23,24							10		90							
25									100							
26							5		95							
27							5	5	90							
28								10	90							
29								12.5	87.5							
30								15	85							

Table A22. Cross-walking table for ESA-CCI LC class 140 - Lichens and mosses

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											90	10				
7-30									100							

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

Table A23. Cross-walking table for ESA-CCI LC class 150 - Sparse vegetation (tree, shrub, herbaceouscover)(<15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											50	10				40
7-12							10		40							50
13				5	5		5		35							50
14				5	5		10		30							50
15				5	5			10	30							50
16				5	5			20	20							50
17,18				10	10			20	10							50
19							5		45							50
20,21			5				5		40							50
22			5				10		35							50
23			10				10		30							50
24			15				15		20							50
25							5	5	40							50
26,27		10					5	5	30							50
28,29		10						20	20							50
30	10							20	20							50

Table A24. Cross-walking table for ESA-CCI LC class 151 - Sparse tree (<15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											50	10				40
7-12							10		40							50
13				5	5				40							50
14,15				5	10				35							50
16				10	5				35							50
17,18				10	10				30							50
19-21			5						45							50
22			10						40							50
23			15						35							50
24			20						30							50
25		10							40							50
26-29		15							35							50
30	15								35							50

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A25. Cross-walking table for ESA-CCI LC class 152 - Sparse shrub (<15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1							5				45	10				40
2-6							10				40	10				40
7-10							10		40							50
11,12							20		30							50
13,14							10		40							50
15,16								15	35							50
17,18								20	30							50
19							5		45							50
20,21							10		40							50
22,23							15		35							50
24							20		30							50
25							5		45							50
26							10		40							50
27							7.5	7.5	35							50
28,29								15	35							50
30								20	30							50

Table A26. Cross-walking table for ESA-CCI LC class 153 - Sparse herbaceous cover (<15%)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											40	10				50
7-30									50							50

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Table A27. Cross-walking table for ESA-CCI LC class 160 - Tree cover, flooded, fresh or brakish water

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6				10							45	45				
7-18				70								30				
19-24			70									30				
25-30	35	35										30				

Table A28. Cross-walking table for ESA-CCI LC class 170 - Tree cover, flooded, saline water

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6					40		30				10	20				
7-12				20				60				20				
13-18				30				50				20				
19-24			60				10	10				20				
25-30	80											20				

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

Table A29. Cross-walking table for ESA-CCI LC class 180 - Shrub or herbaceous cover, flooded, fresh / saline / brakish water

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-6											95	5				
7								10				90				
8							15	15	20			50				
9							20	20	20			40				
10-12							20	20	20			40				
13								20	20			60				
14								25	25			50				
15								30	30			40				
16								35	35			30				
17,18								45	15			40				
19,20							30	40	30							
21,22							40	40	20							
23							40	50	10							
24							30	60	10							
25							30	30	40							
26							30	40	30							
27							40	40	20							
28							40	50	10							
29							70	30								
30							90	10								

Table A30. Cross-walking table for ESA-CCI LC class 190 - Urban

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30															100	

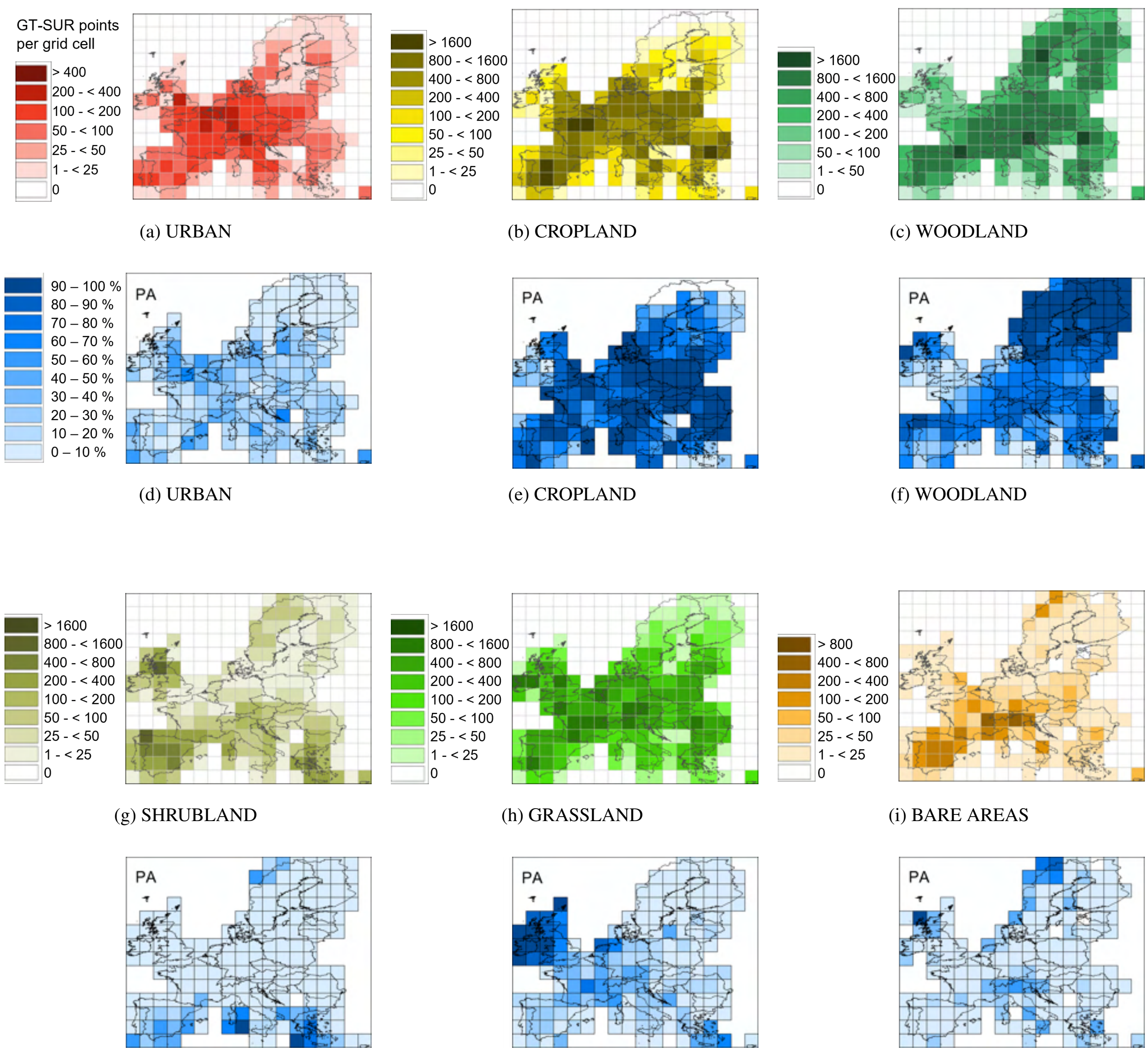
Table A31. Cross-walking table for ESA-CCI LC class 200 - Bare areas, LC class 201 - Consolidated bare areas and LC class 202 - Unconsolidated bare areas.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	Tree						**Shrub**		**Grass**		**Special vegetation**		**Crops**		**Non-vegetated**	
Holdridge Life Zone	tropical broadleaf evergreen	tropical broadleaf deciduous	temperate broadleaf evergreen	temperate broadleaf deciduous	evergreen coniferous	deciduous coniferous	evergreen	deciduous	C3	C4	Tundra	Swamps	crops		urban	bare ground
1-30																100

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

Appendix B



(a) URBAN (b) CROPLAND (c) WOODLAND

(d) URBAN (e) CROPLAND (f) WOODLAND

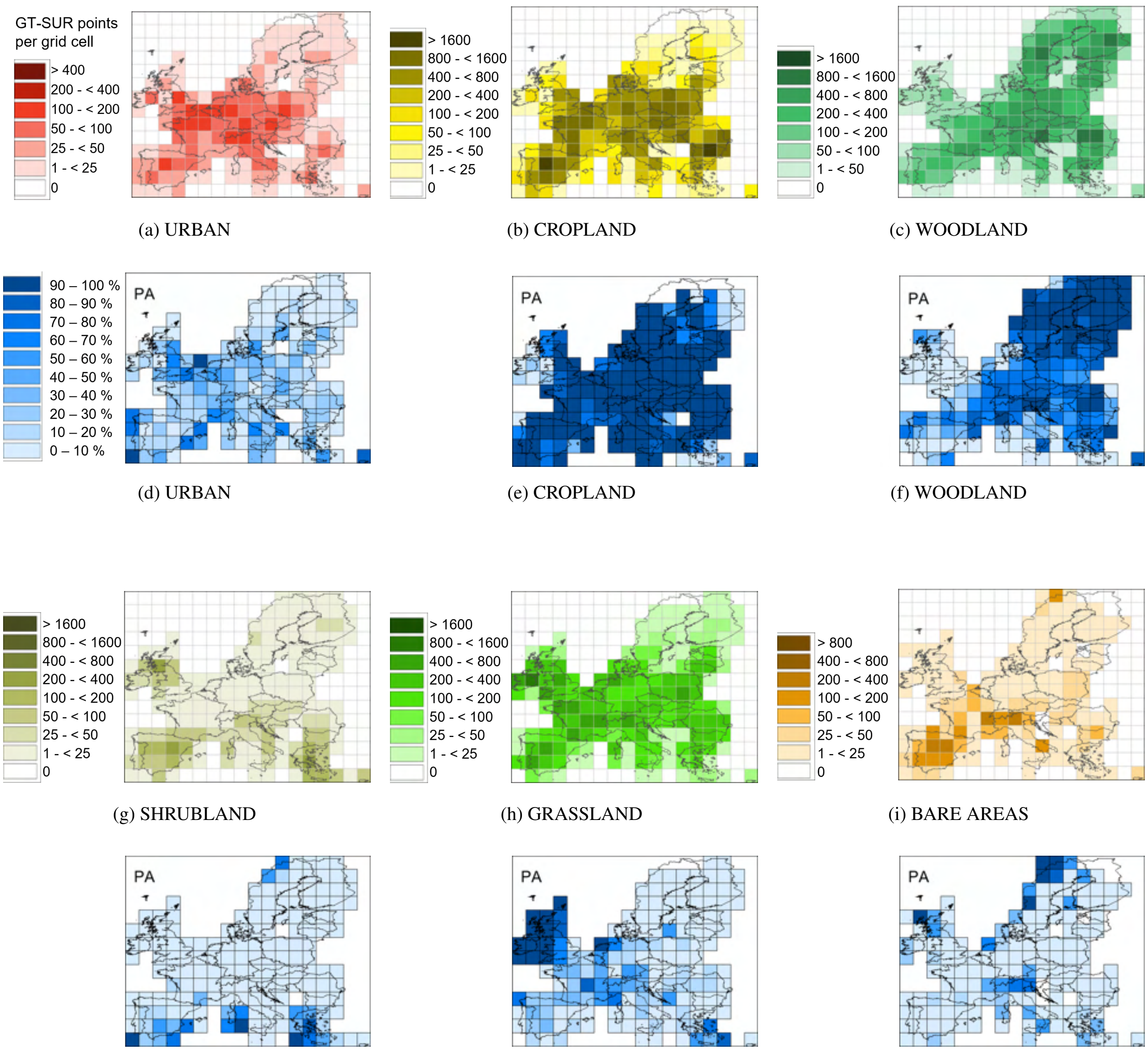
(g) SHRUBLAND (h) GRASSLAND (i) BARE AREAS

(j) SHRUBLAND (k) GRASSLAND (l) BARE AREAS

Figure B1. Total count of GT-SUR points per 2.5° grid cell (a-c; g-i) and producer's accuracy for the individual LULC groups (d-f;j-l) for filter set 5 (dominant LULC group occupies > 50% per LANDMATE PFT grid cell)

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions



(a) URBAN (b) CROPLAND (c) WOODLAND

(d) URBAN (e) CROPLAND (f) WOODLAND

(g) SHRUBLAND (h) GRASSLAND (i) BARE AREAS

(j) SHRUBLAND (k) GRASSLAND (l) BARE AREAS

Figure B2. Total count of GT-SUR points per 2.5° grid cell (a-c; g-i) and producer's accuracy for the individual LULC groups (d-f;j-l) for filter set 7 (dominant LULC group occupies > 70% per LANDMATE PFT grid cell)

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

	1	2	3	4	5	6	7	SUM	UA
1	3234	806	1063	178	1769	120	407	7577	42.68
2	6625	67374	22298	5444	28559	4185	2485	136970	49.19
3	2414	5081	88064	8989	12818	1527	5544	124437	70.77
4	624	5316	4637	5498	1789	439	1487	19790	27.78
5	1411	4515	8063	6082	20763	1767	1643	44244	46.93
6	82	199	200	830	567	1810	460	4148	43.64
7	3	4	49	277	276	530	314	1453	21.61
SUM	14393	83295	124374	27298	66541	10378	12340		
PA	22.47	80.887	70.81	20.14	31.20	17.44	2.54	OA:	55.24

Table B1. Confusion matrix for LANDMATE PFT filter set 1 - Dominant LULC group occupies a minimum of 10 % of a LANDMATE PFT grid cell

	1	2	3	4	5	6	7	SUM	UA
1	3234	806	1063	178	1769	120	407	7577	42.68
2	6625	67374	22298	5444	28559	4185	2485	136970	49.19
3	2414	5081	88064	8989	12818	1527	5544	124437	70.77
4	624	5316	4637	5498	1789	439	1487	19790	27.78
5	1411	4515	8063	6082	20763	1767	1643	44244	46.93
6	82	199	200	830	567	1810	460	4148	43.64
7	3	4	49	277	276	530	314	1453	21.61
SUM	14393	83295	124374	27298	66541	10378	12340		
PA	22.47	80.887	70.81	20.14	31.20	17.44	2.54	OA:	55.24

Table B2. Confusion matrix for LANDMATE PFT filter set 2 - Dominant LULC group occupies a minimum of 20 % of a LANDMATE PFT grid cell

	1	2	3	4	5	6	7	SUM	UA
1	3221	793	1041	174	1748	117	404	7498	42.96
2	6596	67323	22210	5395	28488	4168	2457	136637	49.27
3	2377	5034	87838	8903	12750	1511	5483	123896	70.90
4	615	5280	4484	5363	1748	425	1401	19316	27.76
5	1401	4485	7961	5983	20716	1754	1559	43859	47.23
6	78	187	186	798	552	1799	452	4052	44.40
7	3	4	47	276	275	530	310	1445	21.45
SUM	14291	83106	123767	26892	66277	10304	12066		
PA	22.54	81.01	70.97	19.94	31.26	17.46	2.57	OA:	55.41

Table B3. Confusion matrix for LANDMATE PFT filter set 3 - Dominant LULC group occupies a minimum of 30 % of a LANDMATE PFT grid cell

	1	2	3	4	5	6	7	SUM	UA
1	3079	715	904	152	1597	109	364	6920	44.49
2	6263	66184	20069	4795	27209	4034	2304	130858	50.58
3	2061	4045	83073	7509	11168	1274	5030	114160	72.77
4	501	4813	3013	4235	1392	329	742	15025	28.19
5	1238	4031	6748	5091	19572	1571	1219	39470	49.59
6	54	123	122	606	469	1681	425	3480	48.30
7	2	2	40	254	258	517	252	1325	19.02
SUM	13198	79913	113969	22642	61665	9515	10336		
PA	23.33	82.82	72.89	18.70	31.74	17.67	2.44	OA:	57.22

Table B4. Confusion matrix for LANDMATE PFT filter set 4 - Dominant LULC group occupies a minimum of 40 % of a LANDMATE PFT grid cell

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

	1	2	3	4	5	6	7	SUM	UA
1	2632	499	676	117	1218	84	292	5518	47.70
2	5482	62499	15269	3772	23519	3737	1913	116191	53.79
3	1510	2215	71799	5277	7767	853	4284	93705	76.62
4	362	3865	1752	2689	915	206	350	10139	26.52
5	933	2992	4373	3605	16306	1227	893	30329	53.76
6	31	61	62	292	321	1375	392	2534	54.26
7	1	0	29	110	214	233	70	657	10.65
SUM	10951	72131	93960	15862	50260	7715	8194		
PA	24.03	86.65	76.41	16.95	32.44	17.82	0.85	OA:	60.74

Table B5. Confusion matrix for LANDMATE PFT filter set 5 - Dominant LULC group occupies a minimum of 50 % of a LANDMATE PFT grid cell

	1	2	3	4	5	6	7	SUM	UA
1	2123	284	464	85	844	67	231	4098	51.81
2	4436	56963	10802	2887	19016	3314	1556	98974	57.55
3	1025	978	57212	2949	4699	488	3345	70696	80.93
4	194	2459	967	1713	518	122	240	6213	27.57
5	628	1847	2584	2333	12497	798	630	21317	58.62
6	14	27	34	104	181	1022	339	1721	59.38
7	1	0	18	40	153	87	25	324	7.72
SUM	8421	62558	72081	10111	37908	5898	6366		
PA	25.21	91.06	79.37	16.94	32.97	17.33	0.39	OA:	64.70

Table B6. Confusion matrix for LANDMATE PFT filter set 6 - Dominant LULC group occupies a minimum of 60 % of a LANDMATE PFT grid cell

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

	1	2	3	4	5	6	7	SUM	UA
1	1684	167	311	53	568	44	185	3012	55.91
2	3288	49624	7217	2088	14351	2840	1145	80553	61.60
3	414	255	30158	806	1745	177	1910	35465	85.04
4	40	793	458	988	191	42	160	2672	36.98
5	410	1053	1363	1415	9113	478	425	14257	63.92
6	5	11	15	61	104	768	302	1266	60.66
7	1	0	9	19	99	50	9	187	4.81
SUM	5842	51903	39531	5430	26171	4399	4136		
PA	28.83	95.61	76.29	18.20	34.82	17.46	0.22	OA:	67.20

Table B7. Confusion matrix for LANDMATE PFT filter set 7 - Dominant LULC group occupies a minimum of 70 % of a LANDMATE PFT grid cell

	1	2	3	4	5	6	7	SUM	UA
1	1261	83	208	29	369	32	138	2120	59.48
2	2009	38997	4002	1296	9321	2239	745	58609	66.54
3	32	21	3201	54	195	8	108	3619	88.45
4	10	74	198	442	51	9	106	890	49.66
5	241	518	640	691	5957	240	229	8516	69.95
6	3	5	10	39	62	533	268	920	57.93
7	1	0	6	8	53	17	6	91	6.59
SUM	3557	39698	8265	2559	16008	3078	1600		
PA	35.45	98.23	38.73	17.27	37.21	17.32	0.38	OA:	67.41

Table B8. Confusion matrix for LANDMATE PFT filter set 8 - Dominant LULC group occupies a minimum of 80 % of a LANDMATE PFT grid cell

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

	1	2	3	4	5	6	7	SUM	UA
1	808	44	111	14	207	16	89	1289	62.68
2	592	17167	877	414	2601	1043	269	22963	74.76
3	1	1	47	1	1	0	14	65	72.31
4	2	7	28	74	11	1	10	133	55.64
5	40	81	108	181	1358	83	58	1909	71.14
6	3	1	7	20	28	338	230	627	53.91
7	0	0	1	2	2	1	1	7	14.29
SUM	1446	17301	1179	706	4208	1482	671		
PA	55.88	99.23	3.99	10.48	32.27	22.81	0.15	OA:	73.33

Table B9. Confusion matrix for LANDMATE PFT filter set 9 - Dominant LULC group occupies a minimum of 90 % of a LANDMATE PFT grid cell

	1	2	3	4	5	6	7	SUM	UA
1	252	10	28	0	40	8	51	389	64.78
2	22	565	16	7	52	14	20	696	81.18
3	0	0	0	0	0	0	0	0	/
4	0	0	0	0	0	0	0	0	/
5	0	1	4	14	48	6	1	74	64.86
6	2	0	4	7	9	112	156	290	38.62
7	0	0	0	0	0	0	0	0	/
SUM	276	576	52	28	149	140	228		
PA	91.30	98.09	0.00	0.00	32.21	80.00	0.00	OA:	67.43

Table B10. Confusion matrix for LANDMATE PFT filter set 10 - Dominant LULC group occupies 100 % of a LANDMATE PFT grid cell

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Author contributions. VR conceptualized the paper outline and objective with the support of DR, PH and JB. VR and PH developed the cross-walking procedure and the corresponding cross-walking tables. PH developed the required translation software for the cross-walking procedure. VR developed the accuracy assessment design for the LANDMATE PFT map supported by BB. VR conducted the accuracy assessment and the visualization of results. VR wrote the original draft of the paper, VR, PH, DR, JB, and BB reviewed and edited the draft. VR wrote the final paper

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was financed within the framework of the Helmholtz Institute for Climate Service Science (HICSS), a co-operation between Climate Service Center Germany (GERICS) and Universität Hamburg, Germany and conducted as part of the project LANDMATE (Modelling human LAND surface modifications and its feedbacks on local and regional cliMATE). We acknowledge the support of LUCAS by WCRP-CORDEX as a Flagship Pilot Study. We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). We thank the European Space Agency (ESA) for making the Land cover products publicly available. Special thanks go to the FPS LUCAS partners for providing useful comments in order to improve the dataset.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

References

Alkama, R. and Cescatti, A.: Biophysical climate impacts of recent changes in global forest cover, Science, 351, 600–604, 2016.

Bégué, A., Arvor, D., Bellon, B., Betbeder, J., De Abelleyra, D., PD Ferraz, R., Lebourgeois, V., Lelong, C., Simões, M., and R Verón, S.: Remote sensing and cropping practices: A review, Remote Sensing, 10, 99, 2018.

Belda, M., Halenka, T., Huszar, P., Karlicky, J., and Nováková, T.: Do we need urban parameterization in high resolution regional climate simulations?, in: AGU Fall Meeting Abstracts, vol. 2018, pp. A21L–2878, 2018.

Bojinski, S., Verstraete, M., Peterson, T. C., Richter, C., Simmons, A., and Zemp, M.: The concept of essential climate variables in support of climate research, applications, and policy, Bulletin of the American Meteorological Society, 95, 1431–1443, 2014.

Bonan, G. B.: Forests and climate change: forcings, feedbacks, and the climate benefits of forests, science, 320, 1444–1449, 2008.

Bonan, G. B., Levis, S., Kergoat, L., and Oleson, K. W.: Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models, Global Biogeochemical Cycles, 16, 5–1, 2002.

Bontemps, S., Defourny, P., Radoux, J., Van Bogaert, E., Lamarche, C., Achard, F., Mayaux, P., Boettcher, M., Brockmann, C., Kirches, G., et al.: Consistent global land cover maps for climate modelling communities: current achievements of the ESA's land cover CCI, in: Proceedings of the ESA living planet symposium, Edimburgh, pp. 9–13, 2013.

Box, E. O.: Plant functional types and climate at the global scale, Journal of Vegetation Science, 7, 309–320, 1996.

Bright, R. M., Zhao, K., Jackson, R. B., and Cherubini, F.: Quantifying surface albedo and other direct biogeophysical climate forcings of forestry activities, Global Change Biology, 21, 3246–3266, 2015.

Burke, M. and Emerick, K.: Adaptation to climate change: Evidence from US agriculture, American Economic Journal: Economic Policy, 8, 106–40, 2016.

Chapin Iii, F., McGuire, A., Randerson, J., Pielke, R., Baldocchi, D., Hobbie, S. E., Roulet, N., Eugster, W., Kasischke, E., Rastetter, E., et al.: Arctic and boreal ecosystems of western North America as components of the climate system, Global Change Biology, 6, 211–223, 2000.

Chen, X., Zhang, X.-S., and Li, B.-L.: The possible response of life zones in China under global climate change, Global and Planetary Change, 38, 327–337, 2003.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Böhner, J.: System for automated geoscientific analyses (SAGA) v. 2.1. 4., Geoscientific Model Development Discussions, 8, 2015.

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J., and Jones, P. D.: An ensemble version of the E-OBS temperature and precipitation data sets, Journal of Geophysical Research: Atmospheres, 123, 9391–9409, 2018.

Daly, C., Helmer, E. H., and Quiñones, M.: Mapping the climate of puerto rico, vieques and culebra, International Journal of Climatology: A Journal of the Royal Meteorological Society, 23, 1359–1381, 2003.

Daniel, M., Lemonsu, A., Déqué, M., Somot, S., Alias, A., and Masson, V.: Benefits of explicit urban parameterization in regional climate modeling to study climate and city interactions, Climate Dynamics, 52, 2745–2764, 2019.

Davin, E. L., Rechid, D., Breil, M., Cardoso, R. M., Coppola, E., Hoffmann, P., Jach, L. L., Katragkou, E., de Noblet-Ducoudré, N., Radtke, K., et al.: Biogeophysical impacts of forestation in Europe: first results from the LUCAS (Land Use and Climate Across Scales) regional climate model intercomparison, Earth System Dynamics, 11, 183–200, 2020.

Di Gregorio, A.: Land cover classification system: classification concepts and user manual: LCCS, vol. 2, Food & Agriculture Org., 2005.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Dierckx, W., Sterckx, S., Benhadj, I., Livens, S., Duhoux, G., Van Achteren, T., Francois, M., Mellab, K., and Saint, G.: PROBA-V mission for global vegetation monitoring: standard products and image quality, International Journal of Remote Sensing, 35, 2589–2614, 2014.

Donlon, C., Berruti, B., Buongiorno, A., Ferreira, M.-H., Féménias, P., Frerick, J., Goryl, P., Klein, U., Laur, H., Mavrocordatos, C., et al.: The global monitoring for environment and security (GMES) sentinel-3 mission, Remote Sensing of Environment, 120, 37–57, 2012.

d'Andrimont, R., Yordanov, M., Martinez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Joebges, C., Lemoine, G., et al.: Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union, Scientific Data, 7, 1–15, 2020.

ESA: Land Cover CCI Product User Guide Version 2, Tech. rep., European Space Agency, maps.elie.ucl.ac.be/CCI/viewer/download/ESACCI-LC-Ph2-PUGv2_2.0.pdf, 2017.

ESA: Land cover CCI product user guide version 2, Tech. Report, pp. p–105, 2017.

ESA, E. A. P.: Available online: http://envisat. esa. int/handbooks/asar, CNTR. html (accessed on 27 January 2020), 2002.

Foody, G. M.: Status of land cover classification accuracy assessment, Remote sensing of environment, 80, 185–201, 2002.

Ge, J., Qi, J., Lofgren, B. M., Moore, N., Torbick, N., and Olson, J. M.: Impacts of land use/cover classification accuracy on regional climate simulations, Journal of Geophysical Research: Atmospheres, 112, 2007.

Harris, I., Jones, P., Osborn, T., and Lister, D.: Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset, International Journal of Climatology, 34, 623–642, https://doi.org/10.1002/joc.3711, 2014.

Hartley, A., MacBean, N., Georgievski, G., and Bontemps, S.: Uncertainty in plant functional type distributions and its impact on land surface models, Remote Sensing of Environment, 203, 71–89, 2017.

Hastings, D. A. and Emery, W. J.: The advanced very high resolution radiometer (AVHRR)-A brief reference guide, Photogrammetric Engineering and Remote Sensing, 58, 1183–1188, 1992.

Hoffmann, P., Katzfey, J., McGregor, J., and Thatcher, M.: Bias and variance correction of sea surface temperatures used for dynamical downscaling, Journal of Geophysical Research: Atmospheres, 121, 12–877, 2016.

Hoffmann, P., Reinhart, V., Rechid, D., de Noblet-Ducoudré, N., Davin, E., Asmus, C., Bechtel, B., Böhner, J., Katrakgou, E., and Luyssaert, S.: High-resolution land-use land-cover change data forregional climate modelling applications over Europe - Part 2: Historical and future changes, ESSD, submitted.

Holdridge, L. R. et al.: Life zone ecology., Life zone ecology., 1967.

Hua, T., Zhao, W., Liu, Y., Wang, S., and Yang, S.: Spatial consistency assessments for global land-cover datasets: A comparison among GLC2000, CCI LC, MCD12, GLOBCOVER and GLCNMO, Remote Sensing, 10, 1846, 2018.

Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R., Feddema, J., Fischer, G., Fisk, J., Hibbard, K., Houghton, R., Janetos, A., et al.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, Climatic change, 109, 117, 2011.

Jung, M., Henkel, K., Herold, M., and Churkina, G.: Exploiting synergies of global land cover products for carbon cycle modeling, Remote Sensing of Environment, 101, 534–553, 2006.

Karthikeyan, L., Chawla, I., and Mishra, A. K.: A review of remote sensing applications in agriculture for food security: Crop growth and yield, irrigation, and crop losses, Journal of Hydrology, 586, 124 905, 2020.

Khatun, K., Imbach, P., and Zamora, J.: An assessment of climate change impacts on the tropical forests of Central America using the Holdridge Life Zone (HLZ) land classification system, iForest-Biogeosciences and Forestry, 6, 183, 2013.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Open Access Earth System Science Data Discussions

Kueppers, L. M., Snyder, M. A., and Sloan, L. C.: Irrigation cooling effect: Regional climate forcing by land-use change, Geophysical Research Letters, 34, 2007.

Lattanzi, F. A.: C3/C4 grasslands and climate change, in: Grassland Science in Europe, pp. 3–13, 2010.

Lavorel, S., Díaz, S., Cornelissen, J. H. C., Garnier, E., Harrison, S. P., McIntyre, S., Pausas, J. G., Pérez-Harguindeguy, N., Roumet, C., and Urcelay, C.: Plant functional types: are we getting any closer to the Holy Grail?, in: Terrestrial ecosystems in a changing world, pp. 149–164, Springer, 2007.

Lawrence, D. and Vandecar, K.: Effects of tropical deforestation on climate and agriculture, Nature climate change, 5, 27–36, 2015.

Li, W., MacBean, N., Ciais, P., Defourny, P., Lamarche, C., Bontemps, S., Houghton, R. A., and Peng, S.: Gross and net land cover changes in the main plant functional types derived from the annual ESA CCI land cover maps (1992–2015), 2018.

Lobell, D., Bala, G., and Duffy, P.: Biogeophysical impacts of cropland management changes on climate, Geophysical Research Letters, 33, 2006.

Lu, Y. and Kueppers, L. M.: Surface energy partitioning over four dominant vegetation types across the United States in a coupled regional climate model (Weather Research and Forecasting Model 3–Community Land Model 3.5), Journal of Geophysical Research: Atmospheres, 117, 2012.

Lugo, A. E., Brown, S. L., Dodson, R., Smith, T. S., and Shugart, H. H.: The Holdridge life zones of the conterminous United States in relation to ecosystem mapping, Journal of biogeography, 26, 1025–1038, 1999.

Mahmood, R., Pielke Sr, R. A., Hubbard, K. G., Niyogi, D., Dirmeyer, P. A., McAlpine, C., Carleton, A. M., Hale, R., Gameda, S., Beltrán-Przekurat, A., et al.: Land cover changes and their biogeophysical effects on climate, International journal of climatology, 34, 929–953, 2014.

Maisongrande, P., Duchemin, B., and Dedieu, G.: VEGETATION/SPOT: an operational mission for the Earth monitoring; presentation of new standard products, International Journal of Remote Sensing, 25, 9–14, 2004.

Marconcini, M., Metz-Marconcini, A., Üreyen, S., Palacios-Lopez, D., Hanke, W., Bachofer, F., Zeidler, J., Esch, T., Gorelick, N., Kakarla, A., et al.: Outlining where humans live, the World Settlement Footprint 2015, Scientific Data, 7, 1–14, 2020.

Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A.: Good practices for estimating area and assessing accuracy of land change, Remote Sensing of Environment, 148, 42–57, 2014.

Ottosen, T.-B., Lommen, S. T., and Skjøth, C. A.: Remote sensing of cropping practice in Northern Italy using time-series from Sentinel-2, Computers and Electronics in Agriculture, 157, 232–238, 2019.

Pau, S., Edwards, E. J., and Still, C. J.: Improving our understanding of environmental controls on the distribution of C3 and C4 grasses, Global Change Biology, 19, 184–196, 2013.

Perugini, L., Caporaso, L., Marconi, S., Cescatti, A., Quesada, B., de Noblet-Ducoudre, N., House, J. I., and Arneth, A.: Biophysical effects on temperature and precipitation due to land cover change, Environmental Research Letters, 12, 053 002, 2017.

Poulter, B., Ciais, P., Hodson, E., Lischke, H., Maignan, F., Plummer, S., and Zimmermann, N.: Plant functional type mapping for earth system models, Geoscientific Model Development, 4, 993, 2011.

Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S., Boettcher, M., Brockmann, C., Defourny, P., et al.: Plant functional type classification for earth system models: results from the European Space Agency's Land Cover Climate Change Initiative, Geoscientific Model Development, 8, 2315–2328, 2015.

Rechid, D., Davin, E., de Noblet-Ducoudré, N., and Katragkou, E.: CORDEX Flagship Pilot Study" LUCAS-Land Use & Climate Across Scales"-a new initiative on coordinated regional land use change and climate experiments for Europe, EGUGA, p. 13172, 2017.

Reinhart, V., Fonte, C. C., Hoffmann, P., Bechtel, B., Rechid, D., and Böhner, J.: Comparison of ESA climate change initiative land cover to CORINE land cover over Eastern Europe and the Baltic States from a regional climate modeling perspective, International Journal of Applied Earth Observation and Geoinformation, 94, 102 221, 2021a.

Reinhart, V., Hoffmann, P., and Rechid, D.: LANDMATE PFT land cover dataset for Europe 2015 (Version 1.0), https://doi.org/10.26050/WDCC/LM_PFT_LandCov_EUR2015_v1.0, 2021b.

Richardson, A. D., Keenan, T. F., Migliavacca, M., Ryu, Y., Sonnentag, O., and Toomey, M.: Climate change, phenology, and phenological control of vegetation feedbacks to the climate system, Agricultural and Forest Meteorology, 169, 156–173, 2013.

Rufin, P., Frantz, D., Ernst, S., Rabe, A., Griffiths, P., Özdoğan, M., and Hostert, P.: Mapping cropping practices on a national scale using intra-annual landsat time series binning, Remote Sensing, 11, 232, 2019.

Saad, R., Koellner, T., and Margni, M.: Land use impacts on freshwater regulation, erosion regulation, and water purification: a spatial approach for a global scale level, The International Journal of Life Cycle Assessment, 18, 1253–1264, 2013.

Santos-Alamillos, F., Pozo-Vázquez, D., Ruiz-Arias, J., and Tovar-Pescador, J.: Influence of land-use misrepresentation on the accuracy of WRF wind estimates: Evaluation of GLCC and CORINE land-use maps in southern Spain, Atmospheric Research, 157, 17–28, 2015.

Sertel, E., Robock, A., and Ormeci, C.: Impacts of land cover data quality on regional climate simulations, International Journal of Climatology, 30, 1942–1953, 2010.

Siebert, S., Döll, P., Hoogeveen, J., Faures, J.-M., Frenken, K., and Feick, S.: Development and validation of the global map of irrigation areas, Hydrology and Earth System Sciences, 9, 535–547, 2005.

Skov, F. and Svenning, J.-C.: Potential impact of climatic change on the distribution of forest herbs in Europe, Ecography, 27, 366–380, 2004.

Stehman, S. V.: Sampling designs for accuracy assessment of land cover, International Journal of Remote Sensing, 30, 5243–5272, 2009.

Sulla-Menashe, D. and Friedl, M. A.: User guide to collection 6 MODIS land cover (MCD12Q1 and MCD12C1) product, USGS: Reston, VA, USA, pp. 1–18, 2018.

Szelepcsényi, Z., Breuer, H., and Sümegi, P.: The climate of Carpathian Region in the 20th century based on the original and modified Holdridge life zone system, Central European Journal of Geosciences, 6, 293–307, 2014.

Szelepcsényi, Z., Breuer, H., Kis, A., Pongrácz, R., and Sümegi, P.: Assessment of projected climate change in the Carpathian Region using the Holdridge life zone system, Theoretical and applied climatology, 131, 593–610, 2018.

Tatli, H. and Dalfes, H. N.: Defining Holdridge's life zones over Turkey, International Journal of Climatology, 36, 3864–3872, 2016.

Tatli, H. and Dalfes, H. N.: Analysis of temporal diversity of precipitation along with biodiversity of Holdridge life zones, Theoretical and Applied Climatology, 144, 391–400, 2021.

Thompson, C., Beringer, J., Chapin III, F. S., and McGuire, A. D.: Structural complexity and land-surface energy exchange along a gradient from arctic tundra to boreal forest, Journal of Vegetation Science, 15, 397–406, 2004.

Vilar, L., Garrido, J., Echavarría, P., Martinez-Vega, J., and Martín, M. P.: Comparative analysis of CORINE and climate change initiative land cover maps in Europe: Implications for wildfire occurrence estimation at regional and local scales, International Journal of Applied Earth Observation and Geoinformation, 78, 102–117, 2019.

Wei, Y., Liu, S., Huntzinger, D., Michalak, A., Viovy, N., Post, W., Schwalm, C., Schaefer, K., Jacobson, A., LU, C., Tian, H., Ricciuto, D., Cook, R., Mao, J., and Shi, X.: NACP MsTMIP: Global and North American Driver Data for Multi-Model Intercomparison, https://doi.org/10.3334/ORNLDAAC/1220, 2014.

https://doi.org/10.5194/essd-2021-251
Preprint. Discussion started: 10 August 2021

Wilhelm, C., Rechid, D., and Jacob, D.: Interactive coupling of regional atmosphere with biosphere in the new generation regional climate system model REMO-iMOVE, Geoscientific Model Development, 7, 1093–1114, https://doi.org/10.5194/gmd-7-1093-2014, 2014.

Winter, J. M., Pal, J. S., and Eltahir, E. A.: Coupling of integrated biosphere simulator to regional climate model version 3, Journal of Climate, 22, 2743–2757, 2009.

Wulder, M. A., Franklin, S. E., White, J. C., Linke, J., and Magnussen, S.: An accuracy assessment framework for large-area land cover classification products derived from medium-resolution satellite data, International Journal of Remote Sensing, 27, 663–683, 2006.

Wullschleger, S. D., Epstein, H. E., Box, E. O., Euskirchen, E. S., Goswami, S., Iversen, C. M., Kattge, J., Norby, R. J., van Bodegom, P. M., and Xu, X.: Plant functional types in Earth system models: past experiences and future directions for application of dynamic vegetation models in high-latitude ecosystems, Annals of botany, 114, 1–16, 2014.

Yang, Y., Xiao, P., Feng, X., and Li, H.: Accuracy assessment of seven global land cover datasets over China, ISPRS Journal of Photogrammetry and Remote Sensing, 125, 156–173, 2017.

Yue, T., Liu, J., Jørgensen, S. E., Gao, Z., Zhang, S., and Deng, X.: Changes of Holdridge life zone diversity in all of China over half a century, Ecological Modelling, 144, 153–162, 2001.

Yue, T. X., Fan, Z. M., Liu, J. Y., and Wei, B. X.: Scenarios of major terrestrial ecosystems in China, ecological modelling, 199, 363–376, 2006.

B. Appendix: Additional figures

Table B.1: Confusion matrix for ESA CCI PFT filter set 1 - Dominant LULC group occupies a minimum of 10 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	3134	749	1072	172	1818	107	236	7288	43.00
WOODLAND	7289	72447	25082	6421	31898	4423	2117	149677	48.40
CROPLAND	2196	4340	84899	7557	11789	1324	3475	115580	73.45
SHRUBLAND	206	1508	2058	2938	653	127	65	7555	38.89
GRASSLAND	1196	3466	8820	7659	17894	1033	2167	42235	42.37
BARE AREAS	82	415	581	1776	1673	3068	335	7930	38.69
OTHER	290	370	1871	784	822	310	5414	9861	54.90
SUM	14393	83295	124383	27307	66547	10392	13809		
PA	21.77	86.98	68.26	10.76	26.89	29.52	39.21	340126	55.80

Table B.2: Confusion matrix for ESA CCI PFT filter set 2 - Dominant LULC group occupies a minimum of 20 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	3134	749	1072	172	1818	107	236	7288	10.28
WOODLAND	7289	72447	25082	6421	31898	4423	2117	149677	48.40
CROPLAND	2196	4340	84899	7557	11789	1324	3475	115580	73.45
SHRUBLAND	206	1507	2058	2938	653	127	65	7554	38.89
GRASSLAND	1196	3466	8806	7638	17863	1014	2160	42143	42.39
BARE AREAS	80	409	573	1686	1595	2512	237	7092	35.42
OTHER	289	369	1870	769	820	270	5255	9642	54.50
SUM	14390	83287	124360	27181	66436	9777	13545		
PA	21.78	86.98	68.27	10.81	26.89	25.69	38.80	338976	55.07

Table B.3: Confusion matrix for ESA CCI PFT filter set 3 - Dominant LULC group occupies a minimum of 30 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	3128	747	1066	170	1816	107	235	7269	43.03
WOODLAND	7288	72438	25069	6409	31885	4421	2115	149625	48.41
CROPLAND	2191	4335	84888	7555	11785	1322	3474	115550	73.46
SHRUBLAND	201	1495	2024	2895	641	124	64	7444	38.89
GRASSLAND	1181	3412	8627	7371	17759	972	1944	41266	43.04
BARE AREAS	64	358	523	1406	1476	2192	171	6190	35.41
OTHER	278	366	1845	722	796	230	5051	9288	54.38
SUM	14331	83151	124042	26528	66158	9368	13054		
PA	21.83	87.12	68.43	10.91	26.84	23.40	38.69	336632	55.95

Table B.4: Confusion matrix for ESA CCI PFT filter set 4 - Dominant LULC group occupies a minimum of 40 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	3055	712	997	160	1721	103	219	6967	43.85
WOODLAND	7136	71899	24084	6003	31137	4367	2070	146696	49.01
CROPLAND	2075	3985	83043	7083	11161	1240	3375	111962	74.17
SHRUBLAND	163	1278	1454	2215	448	90	49	5697	38.88
GRASSLAND	1004	2813	6301	5136	15791	746	1241	33032	47.81
BARE AREAS	35	195	336	779	1018	1414	91	3868	36.56
OTHER	237	312	1657	499	668	175	4284	7832	54.70
SUM	13705	81194	117872	21875	61944	8135	11329		
PA	22.29	88.55	70.45	10.13	25.49	17.38	37.81	316054	57.49

Table B.5: Confusion matrix for ESA CCI PFT filter set 5 - Dominant LULC group occupies a minimum of 50 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	2797	608	810	125	1428	86	192	6046	46.26
WOODLAND	6576	69227	20372	4938	28054	4109	1859	135135	51.23
CROPLAND	1628	2688	73489	5238	8567	881	2939	95430	77.01
SHRUBLAND	96	825	700	1307	248	47	25	3248	40.24
GRASSLAND	649	1823	3362	2719	11110	388	742	20793	53.43
BARE AREAS	16	72	117	268	425	590	34	1522	38.76
OTHER	153	224	1336	313	476	111	3396	6009	56.52
SUM	11915	75467	100186	14908	50308	6212	9187		
PA	23.47	91.73	73.35	8.77	22.08	9.50	36.97	268183	60.38

Table B.6: Confusion matrix for ESA CCI PFT filter set 6 - Dominant LULC group occupies a minimum of 60 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	2319	423	576	87	1075	63	137	4680	49.55
WOODLAND	5575	63626	15427	3768	23450	3741	1526	117113	54.33
CROPLAND	1039	1292	56321	2928	4730	493	2073	68876	81.77
SHRUBLAND	35	368	269	652	95	24	13	1456	44.78
GRASSLAND	391	1061	1547	1244	7077	159	421	11900	59.47
BARE AREAS	4	19	29	79	105	181	11	428	42.29
OTHER	91	142	938	178	305	66	2355	4075	57.79
SUM	9454	66931	75107	8936	36837	4727	6536		
PA	24.53	95.06	74.99	7.30	19.21	3.83	36.03	208528	63.56

Table B.7: Confusion matrix for ESA CCI PFT filter set 7 - Dominant LULC group occupies a minimum of 70 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	1784	246	364	51	702	44	88	3279	54.41
WOODLAND	4444	56713	10728	2762	18637	3293	1160	97737	58.03
CROPLAND	228	266	17671	574	932	81	430	20182	87.56
SHRUBLAND	10	81	74	260	32	5	3	465	55.91
GRASSLAND	221	509	676	493	4379	62	201	6541	66.95
BARE AREAS	1	5	4	27	19	34	5	95	35.79
OTHER	34	68	226	66	140	26	920	1480	62.16
SUM	6722	57888	29743	4233	24841	3545	2807		
PA	26.54	97.97	59.41	6.14	17.63	0.96	32.78	129779	63.00

Table B.8: Confusion matrix for ESA CCI PFT filter set 8 - Dominant LULC group occupies a minimum of 80 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	1257	127	217	22	422	24	47	2116	59.40
WOODLAND	3211	48638	6706	1923	13699	2769	807	77753	62.55
CROPLAND	4	19	302	8	23	0	3	359	84.12
SHRUBLAND	1	8	9	49	4	0	1	72	68.06
GRASSLAND	80	174	227	119	1940	18	66	2624	73.93
BARE AREAS	0	3	0	3	0	1	0	7	14.29
OTHER	9	22	15	19	39	6	220	330	66.67
SUM	4562	48991	7476	2143	16127	2818	1144		
PA	27.55	99.28	4.04	2.29	12.03	0.04	19.23	83261	62.94

Table B.9: Confusion matrix for ESA CCI PFT filter set 9 - Dominant LULC group occupies a minimum of 90 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	781	43	113	10	205	13	25	1190	65.63
WOODLAND	1877	37309	3612	1176	8611	2105	441	55131	67.67
CROPLAND	0	0	1	0	1	0	0	2	50.00
SHRUBLAND	0	0	0	2	0	0	0	2	100.00
GRASSLAND	0	1	0	1	13	1	0	16	81.25
BARE AREAS	3	1	1	3	3	1	29	41	2.44
OTHER	0	0	1	2	2	1	1	7	14.29
SUM	2661	37354	3728	1194	8835	2121	496		
PA	29.35	99.88	0.03	0.17	0.15	0.05	0.20	56389	67.58

Table B.10: Confusion matrix for ESA CCI PFT filter set 3 - Dominant LULC group occupies 100 % of a ESA CCI PFT grid cell

	URBAN	WOODLAND	CROPLAND	SHRUBLAND	GRASSLAND	BARE AREAS	OTHER	SUM	UA
URBAN	302	14	36	3	59	6	9	429	70.40
WOODLAND	663	19548	1096	463	3183	1135	166	26254	74.46
CROPLAND	0	0	0	0	0	0	0	0	/
SHRUBLAND	0	0	0	0	0	0	0	0	/
GRASSLAND	0	0	0	0	0	0	0	0	/
BARE AREAS	0	0	0	0	0	0	0	0	/
OTHER	0	0	0	0	0	0	0	0	/
SUM	965	19562	1132	466	3242	1141	175		
PA	31.30	99.93	0.00	0.00	0.00	0.00	0.00	26683	74.39

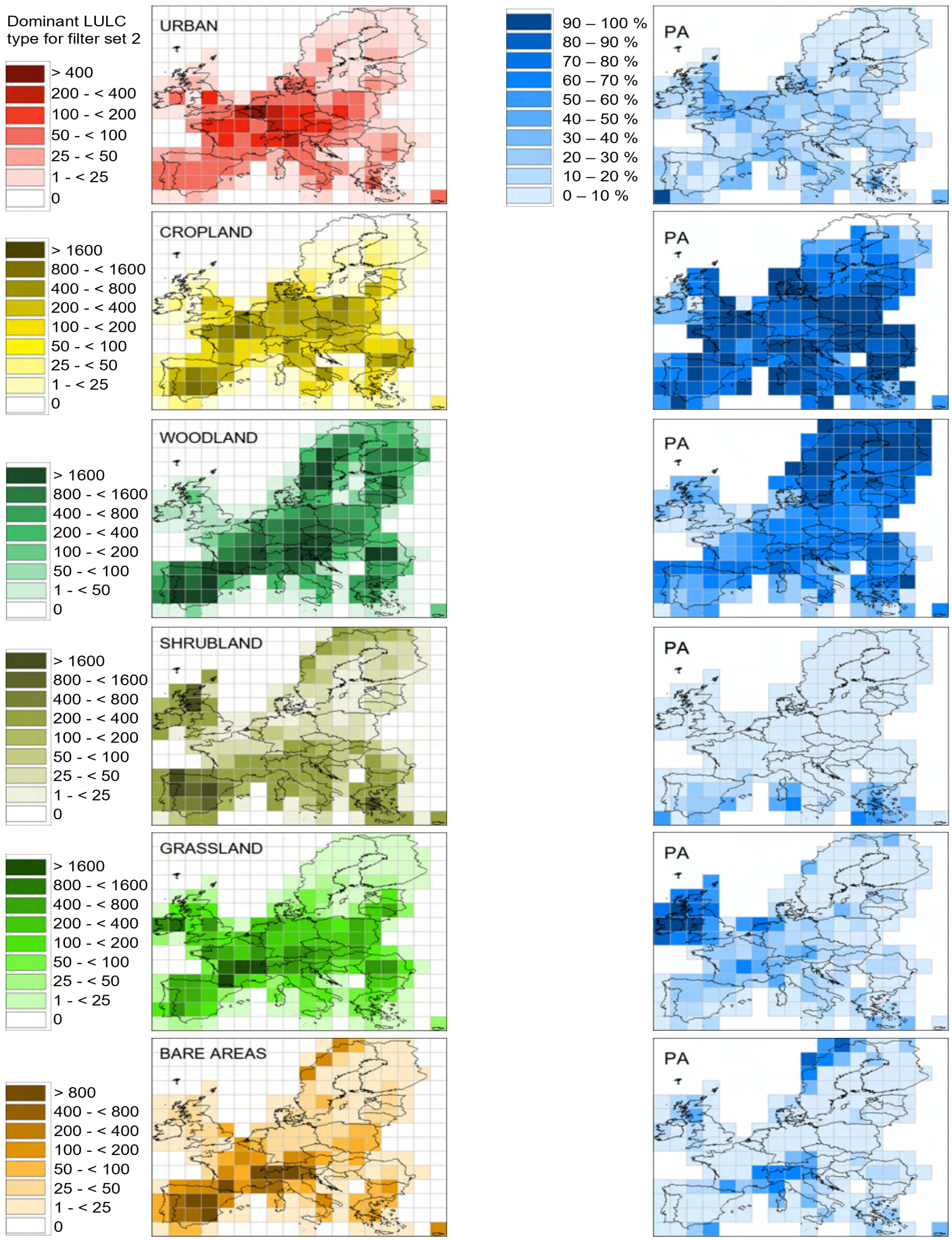


Figure B.1: Total count of ESA CCI PFT cells where the dominant LULC types cover 20 % or more per 2.5° grid cell (left column) and producer's accuracy for the individual LULC groups for the filter set 2 (right column)

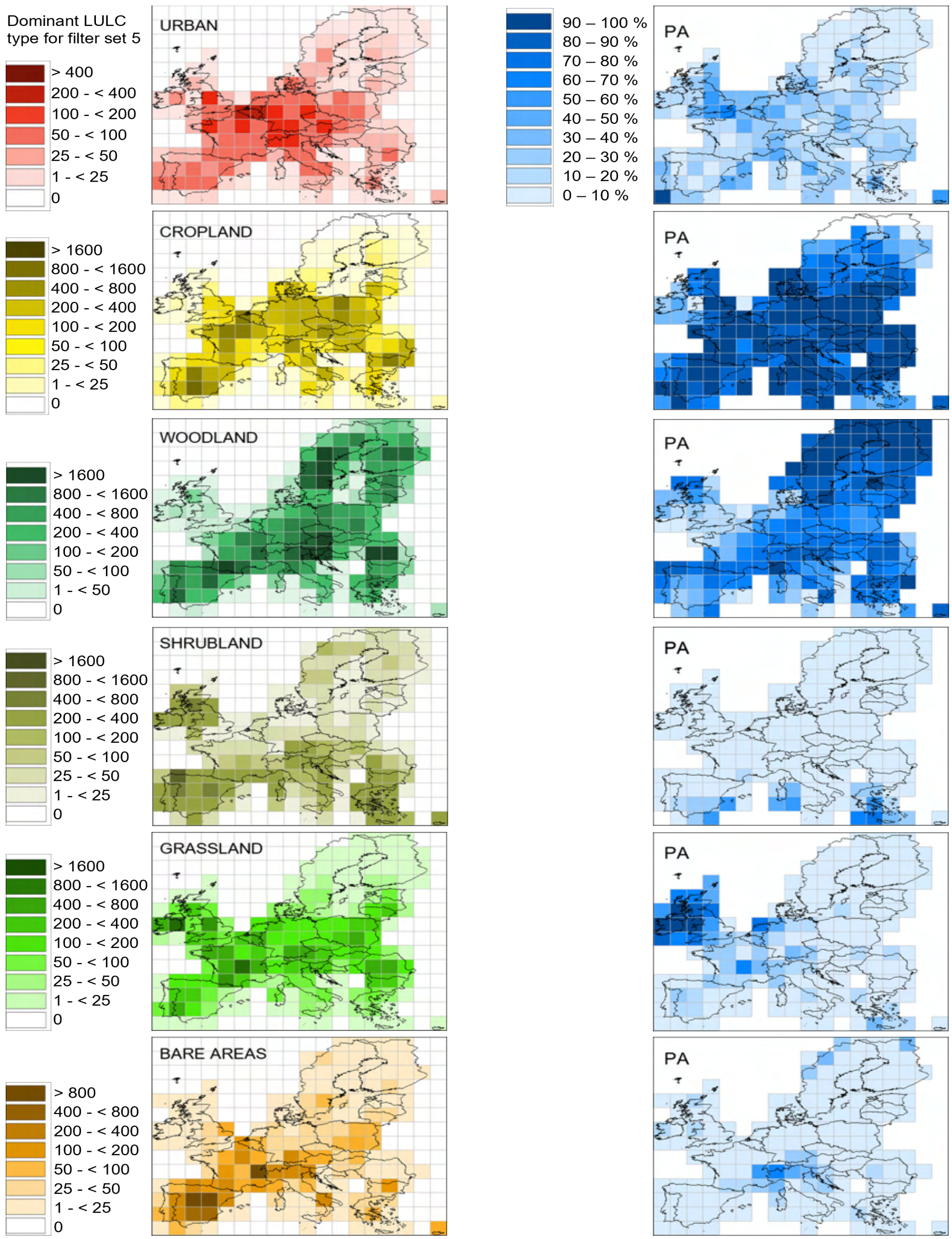


Figure B.2: Total count of ESA CCI PFT cells where the dominant LULC types cover 50 % or more per 2.5° grid cell (left column) and producer's accuracy for the individual LULC groups for the filter set 5 (right column)

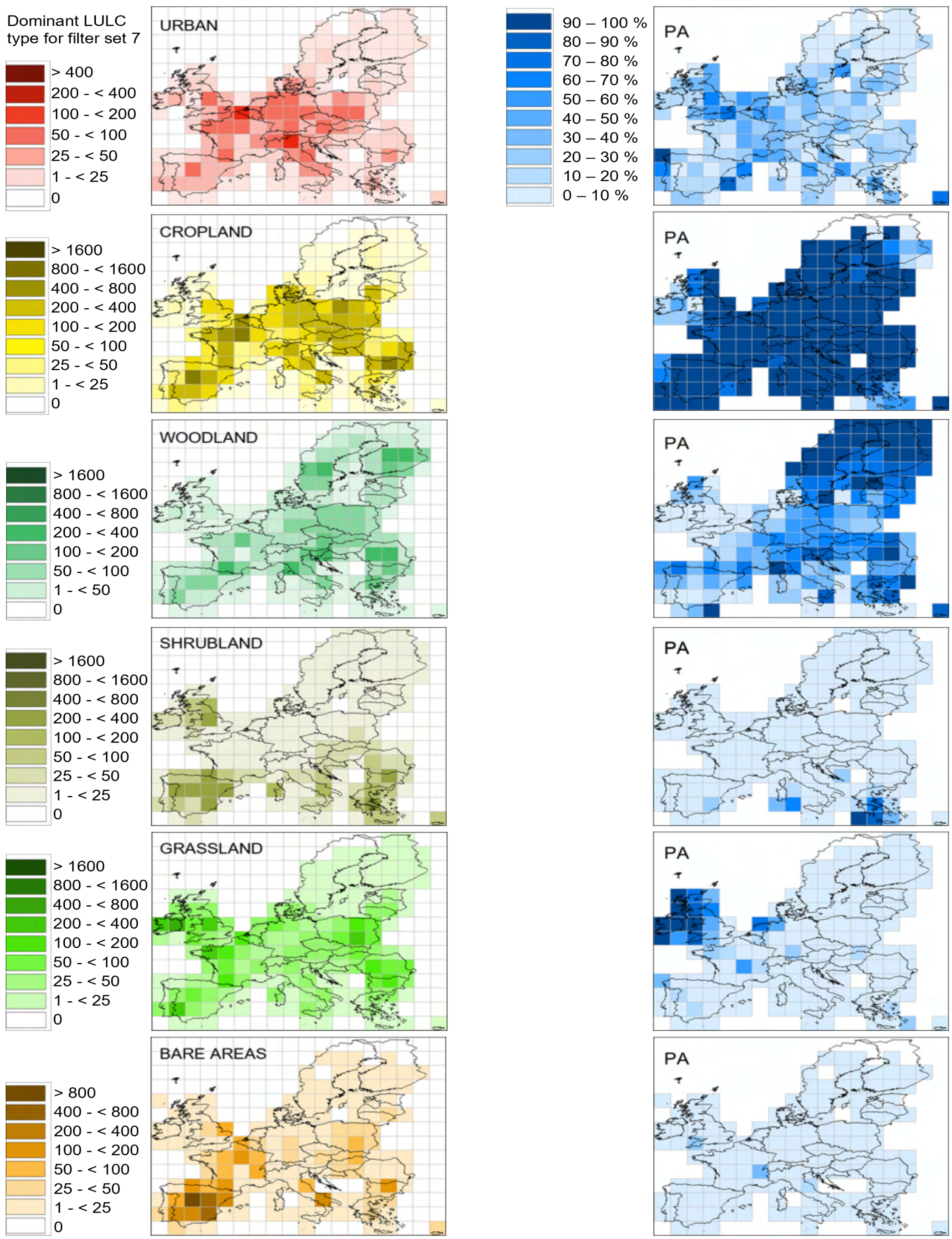


Figure B.3: Total count of ESA CCI PFT cells where the dominant LULC types cover 70 % or more per 2.5° grid cell (left column) and producer's accuracy for the individual LULC groups for the filter set 7 (right column)

Eidesstattliche Versicherung | Declaration on Oath

Hiermit erkläre ich an Eides statt, dass ich die vorliegende Dissertationsschrift selbst verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.
|
I hereby declare upon oath that I have written the present dissertation independently and have not used further resources and aids than those stated.

Hamburg, den 24.11.2021 Vanessa Reinhart

Ort, den | City, date Unterschrift | Signature