A spatiotemporal ensemble machine learning framework for generating land use / land cover time-series maps for Europe (2000 – 2019) based on LUCAS, CORINE and GLAD Landsat

Martijn Witjes (martijn.witjes@opengeohub.org)
OpenGeoHub, Wageningen, The Netherlands

Leandro Parente
OpenGeoHub, Wageningen, The Netherlands

Chris J. van Diemen
OpenGeoHub, Wageningen, The Netherlands

Tomislav Hengl
OpenGeoHub, Wageningen, The Netherlands

Martin Landa
Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

Lukas Brodsky
Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

Lena Halounova
Department of Geomatics, Faculty of Civil Engineering, Czech Technical University of Prague, Prague, Czech Republic

Josip Krizan
MultiOne, Zagreb, Croatia

Luka Antonic
MultiOne, Zagreb, Croatia

Codrina M Ilie
Technical University of Civil Engineering Bucharest, Bucharest, Romania

Vasile Craciunescu
National Meteorological Administration of Romania, Bucharest, Romania

Milan Kilibarda
Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia

Ognjen Antonijevic
Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia

**Luka Glusica**
GILAB, Belgrade, Serbia

---

**Method Article**

**Keywords:** landsat, spatial analysis, spatiotemporal, ensemble, machine learning, probability, uncertainty, land use, land cover, big data, environmental monitoring

**DOI:** [https://doi.org/10.21203/rs.3.rs-561383/v1](https://doi.org/10.21203/rs.3.rs-561383/v1)

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
A seamless spatiotemporal machine learning framework for automated prediction, uncertainty assessment, and analysis of land use / land cover (LULC) dynamics is presented. The framework includes: (1) harmonization and preprocessing of high-resolution spatial and spatiotemporal covariate datasets (GLAD Landsat, NPP/VIIRS) including 5 million harmonized LUCAS and CORINE Land Cover-derived training samples, (2) model building based on spatial k-fold cross-validation and hyper-parameter optimization, (3) prediction of the most probable class, class probabilities and uncertainty per pixel, (4) LULC change analysis on time-series of produced maps. The spatiotemporal ensemble model was fitted by combining random forest, gradient boosted trees, and artificial neural network, with logistic regressor as meta-learner. The results show that the most important covariates for mapping LULC in Europe are: seasonal aggregates of Landsat green and near-infrared bands, multiple Landsat-derived spectral indices, and elevation. Spatial cross-validation of the model indicates consistent performance across multiple years with 62%, 70%, and 87% accuracy when predicting 33 (level-3), 14 (level-2), and 5 classes (level-1); with artificial surface classes such as 'airports' and 'railroads' showing the lowest match with validation points. The spatiotemporal model outperforms spatial models on known-year classification by 2.7% and unknown-year
classification by 3.5%. Results of the accuracy assessment using 48,365 independent test samples shows 87% match with the validation points. Results of time-series analysis (time-series of LULC probabilities and NDVI images) suggest gradual deforestation trends in large parts of Sweden, the Alps, and Scotland. An advantage of using spatiotemporal ML is that the fitted model can be used to predict LULC in years that were not included in its training dataset, allowing generalization to past and future periods, e.g. to predict land cover for years prior to 2000 and beyond 2020. The generated land cover time-series data stack (ODSE-LULC), including the training points, is publicly available via the Open Data Science (ODS)-Europe Viewer.
A Spatiotemporal Ensemble Machine Learning Framework for Generating Land Use / Land Cover Time-series Maps for Europe (2000–2019) based on LUCAS, CORINE and GLAD Landsat

Martijn Witjes¹, Leandro Parente¹, Chris van Diemen¹, Tomislav Hengl¹, Martin Landa², Lukáš Brodský², Lena Halounová², Josip Križan³, Luka Antonić³, Codrina Maria Ilie⁴, Vasile Craciunescu⁴, Milan Kilibarda⁵, Ognjen Antonijević⁵, and Luka Glušica⁶

¹OpenGeoHub, Wageningen, the Netherlands
²Department of Geomatics, Faculty of Civil Engineering, CTU in Prague, Czech Republic
³MultiOne, Zagreb, Croatia
⁴Terrasigna, Romania
⁵Department of Geodesy and Geoinformatics, Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia
⁶GiLAB, Belgrade, Serbia

Corresponding author:
Martijn Witjes¹

Email address: martijn.witjes@opengeohub.org

ABSTRACT
A seamless spatiotemporal machine learning framework for automated prediction, uncertainty assessment, and analysis of Land Cover / Land Use dynamics is presented. The framework includes: (1) harmonization and preprocessing of high-resolution spatial and spatiotemporal covariate datasets (GLAD Landsat, NPP/VIIRS) including 5 million harmonized LUCAS and CORINE Land Cover-derived training samples, (2) model building based on spatial k-fold cross-validation and hyper-parameter optimization, (3) prediction of the most probable class, class probabilities and uncertainty per pixel, (4) LULC change analysis on time-series of produced maps. The spatiotemporal ensemble model was fitted by combining random forest, gradient boosted trees, and artificial neural network, with logistic regressor as meta-learner. The results show that the most important covariates for mapping LULC in Europe are: seasonal aggregates of Landsat green and near-infrared bands, multiple Landsat-derived spectral indices, and elevation. Spatial cross-validation of the model indicates consistent performance across multiple years with 62%, 70%, and 87% accuracy when predicting 33 (level-3), 14 (level-2), and 5 classes (level-1); with artificial surface classes such as “airports” and “railroads” showing the lowest match with validation points. The spatiotemporal model outperforms spatial models on known-year classification by 2.7% and unknown-year classification by 3.5%. Results of the accuracy assessment using 48,365 independent test samples shows 87% match with the validation points. Results of time-series analysis (time-series of LULC probabilities and NDVI images) suggest gradual deforestation trends in large parts of Sweden, the Alps, and Scotland. An advantage of using spatiotemporal ML is that the fitted model can be used to predict LULC in years that were not included in its training dataset, allowing generalization to past and future periods, e.g. to predict land cover for years prior to 2000 and beyond 2020. The generated land cover time-series data stack (ODSE-LULC), including the training points, is publicly available via the Open Data Science (ODS)-Europe Viewer.

Submitted to PeerJ on 7th of May 2021

INTRODUCTION

Anthropogenic land cover change has influenced global climate since the Paleolithic (Kaplan et al., 2011) and continues to be a major driver of regional (Pielke Sr et al., 2002) and global (Houghton et al., 2012) climate change. Furthermore, it is the single largest cause of global biodiversity loss (Sala et al., 2000), and has quantifiable consequences for the availability and quality of natural resources, water, and air (Foley et al., 2005). Key applications of land cover change maps are to inform policy (Duveiller et al., 2020), analyse land-based emissions (Hong et al., 2021) and/or help estimate local climate extremes (Sy and Quesada, 2020). Quantifying land cover dynamics is often crucial for policy-making at regional and global levels (Liu et al., 2020b; Trisurat et al., 2019; Shumba et al., 2020).

Land cover has been traditionally first mapped by doing visual interpretation of aerial photographs; later on by automating classification of multispectral remotely sensed data / with semi-supervised or fully-supervised methods (Townshend et al., 2012; Feranec et al., 2016; Liu et al., 2021). There are currently multiple global (Feng and Bai, 2019; Buchhorn et al., 2020) and regional (Homer et al., 2007; Batista e Silva et al., 2013; Pflugmacher et al., 2019; Malinowski et al., 2020) land cover products based on using Machine Learning and offering predictions (or their refinements) at high spatial resolutions for the whole of continental Europe (Table 1). The increasing number of land cover applications and datasets in Europe can largely be attributed to (1) the extensive Land use and Coverage Area frame Survey...
(LUCAS) *in-situ* point data being publicly available for research (at [https://land.copernicus.eu/imagery-in-situ/lucas](https://land.copernicus.eu/imagery-in-situ/lucas)), and (2) NASA’s Landsat and ESA’s Sentinel multispectral images being increasingly available for spatial analysis ([Szantoi et al., 2020; Liu et al., 2021](https://land.copernicus.eu/imagery-in-situ/lucas)).

Not all land cover prediction algorithms and systems, however, perform equally. [Vilar et al. (2019)](https://land.copernicus.eu/imagery-in-situ/lucas) have done extensive evaluation of accuracy of the Coordination of Information on the Environment (CORINE) Land Cover (CLC) products for period 2011–2012 using the LUCAS data and found that agreement with LUCAS was slightly higher for Climate Change Initiative — Land Cover (CCI-LC) (59%; 18 classes) than for CLC (56%; 44 classes). [Gao et al. (2020)](https://land.copernicus.eu/imagery-in-situ/lucas) has evaluated accuracy of the global 30 m resolution products GlobeLand30 with 10 classes ([Chen et al., 2015](https://land.copernicus.eu/imagery-in-situ/lucas)), and Global Land Cover with Fine Classification System at 30 m (GLC FCS30) with 18 classes ([Zhang et al., 2020](https://land.copernicus.eu/imagery-in-situ/lucas)) using the LUCAS point data and concluded that the GlobeLand30-2010 product agrees with LUCAS points up to 89%, while GLC FCS30-2015 agrees up to 85%. The large difference in the agreement reported by [Vilar et al. (2019)](https://land.copernicus.eu/imagery-in-situ/lucas) and [Chen et al. (2015](https://land.copernicus.eu/imagery-in-situ/lucas)) can be attributed to the number of classes in the two studies: the absolute accuracy linearly drops with the number of classes ([Herold et al., 2008; Van et al., 2019](https://land.copernicus.eu/imagery-in-situ/lucas)), and usually the accuracy results for 6–10 classes vs 40 classes can be up to 50% better.

Overall, the land cover mapping results in Europe match the results of [Calderón-Loor et al. (2021)](https://land.copernicus.eu/imagery-in-situ/lucas) who estimated 90% accuracy on 6 classes with 7 years (from 1985 to 2015) of Landsat data of Australia. [Tsendbazar et al. (2018)](https://land.copernicus.eu/imagery-in-situ/lucas) reports similar accuracy levels for Africa. Likewise, [Liu et al. (2020a)](https://land.copernicus.eu/imagery-in-situ/lucas) reports 83% accuracy on 7 classes with 34 years of Global Land Surface Satellite (GLASS) data. USA one of the first continent-scale countries to produce and distribute the National Land Cover Database at 30 m ([https://www.mrlc.gov/](https://www.mrlc.gov/); years 2001, 2004, 2006, 2008, 2011, 2013, 2016, 2018) ([Homer et al., 2020](https://land.copernicus.eu/imagery-in-situ/lucas)), which focuses on 16 classes and reports accuracy of at least 80%.

[Inglada et al. (2017)](https://land.copernicus.eu/imagery-in-situ/lucas) report a kappa of 0.86 for mapping 17 classes of land cover for France for year 2014. The most-up-to-date land cover products for Europe by ([Malinowski et al., 2020](https://land.copernicus.eu/imagery-in-situ/lucas)) report mapping accuracy of 86% based on predicting 13 classes with 2017 Sentinel-2 data. The ESA’s WorldCover project ([https://esa-worldcover.org/](https://esa-worldcover.org/)) is another global 10 m spatial resolution land cover product that aims at consistent mapping accuracy of at least 75%.

### Table 1. Inventory and comparison of existing land cover data products at finer spatial resolutions (≤300 m) available for the continental Europe.

| Product / reference | Time span | Spatial resolution | Mapping accuracy | Classification system | Uncertainty / Probability |
|---------------------|-----------|--------------------|------------------|-----------------------|--------------------------|
| CLC                 | 1990, 2000, 2006, 2012, 2018 | 100 m (25 ha) | ≤85% | 44 classes | N / N |
| ESA CCI-LC (Batista e Silva et al., 2013) | 2006 | 300 m | 22 classes | N / N |
| S2GLC (Malinowski et al., 2020) | 2017 | 10 m | 15 classes | |
| Pfugmacher et al. (2019) | 2014-2016 | 30 m | 75% | 12 classes | N / N |
| GLC FCS30 (Zhang et al., 2020) | 2015, 2020 | 30-m | | |
| Buchhorn et al. (2020) | 2015, 2016, 2017, 2018 | 100 m | ?? | N / Y |
| ESA WorldCover | 2020 | 10 m | >75% | 12+ | N / N |
| ELC10 (Venter and Sydenham, 2021) | 2020 | 10 m | >90% | 8 classes | N / N |
| ODSE-LULC (our product) | 2000, 2001, …, 2019 | 30 m | 33 classes | Y / Y |

Based on these works, it can be said that the state-of-the-art land cover mapping projects primarily aim at:
(a) automating process as much as possible so that land cover maps can be produced almost on monthly
or even daily revisit times,

(b) using multi-source Earth Observation data, with especial focus on combining power of the Sentinel-1
and 2 data (Venter and Sydenham, 2021),

(c) producing higher and higher spatial resolution and thematic detail data.

Although the modern approaches to land cover mapping listed in Table 1 report relatively high levels
of accuracy, we recognize several limitations of the general approach:

• The focus of common land cover classification products is often only on hard classes (the most
probable class); per-pixel uncertainty in predictions is often either not reported or not derived at
all. Mapping accuracy is provided as a general number (average performance) for the whole area,
although in practice prediction accuracy often varies from class to class.

• Most of policy-related institutions require time-series land cover data products which are compatible
with legacy products such as CLC and CCI-LC, while most of research focuses on producing general
land cover maps for recent years only. It appears that many land cover mapping missions, especially
the global missions, tend to sacrifice complexity of the target legend for the sake of higher nominal
accuracy.

• In the case of the land cover dynamics mapping, usually no further analysis is provided to help
detect and quantify trends and eventually understand main drivers of land cover change.

Land cover data with higher thematic resolution have shown to help improve the performance of
subsequent change detection (Buyantuyev and Wu, 2007), as well as the performance and level of detail
of modeling land cover trends (Conway, 2009) and other environmental phenomena (Castilla et al., 2009;
Zhou et al., 2014). Increasing thematic resolution while limiting the prediction to one trained classifier,
however, poses several challenges: (1) training a single model on multi-year data requires extensive data
harmonization efforts, and (2) the exponential increase of possible change types with each additional
predicted class complicates the manual creation of post-classification temporal consistency rules.

With an increasing spatial resolution and increasing extent of EO images, the gap between historic
land cover maps and current 10 m resolution products is growing. This makes it difficult to identify key
processes of land cover change over large areas (Veldkamp and Lambin, 2001; Vilar et al., 2019). Hence,
a balanced and consistent approach is needed that can take into account both accuracy gains due to spatial
resolution, and applicability for time-series analysis / change detection for longer periods of time.

In this paper we describe a complete seamless framework for spatiotemporal prediction, uncertainty
assessment and analysis of the land cover dynamics using long time-series (20+ years) in a High Per-
formance Computing framework. We present results of modeling and predicting land cover classes for
continental Europe using spatiotemporal Machine Learning at 30 m spatial resolution. We fit and use a
single model for the whole spacetime cube of interest. This allows us to both continue predicting land
cover for subsequent years, and to try to back-track land cover status even prior to the year 2000, without
a need to collect additional training (point) data.
We provide, in addition, results of internal accuracy assessment, based on 5-fold spatial cross-validation with refitting (Roberts et al., 2017; Lovelace et al., 2019), a comparison of spatial vs spatiotemporal models, and also results of time-series analysis on the whole data-cube (2000–2019). We use, as much as possible, a consistent methodology, which implies:

1. Using consistent training data based on consistent sampling methodology and sampling intensity over the complete spacetime cube of interest (LUCAS; d’Andrimont et al. (2020));

2. Using consistent / harmonized Earth Observation images based on the Global Land Analysis and Discovery (GLAD) Analysis Ready Data (ARD) Landsat product (Potapov et al., 2020), Night Light images NPP/VIIRS (Román et al., 2018) and similar;

3. Providing consistent statistical analysis per every pixel of the space-time cube and per each probability;

Our modeling framework comes at high costs however: the data we have produced is about 50–100 times larger in size than common land cover products with the total size of about 20 TiB (Cloud-Optimized GeoTIFFs). These data is both more complex to analyze and to visualize. To deal with the data size, we ran all processing in a fully automated and fully optimized High Performance Computing (HPC) framework. We refer to the dataset we have produced as Open Data Science Europe — Land Use / Land Cover or short ODSE-LULC.

In the following section we describe how we prepared data, fitted models, tested spatial vs spatiotemporal models, and fitted pixel-wise space-time regressions for NDVI and probability time-series. We then report the results and discuss advantages and limitations of spatiotemporal Ensemble Machine Learning (EML), and suggest what we consider could be next development directions and challenges.

**MATERIALS AND METHODS**

**Spatiotemporal Machine Learning**

The annual land cover product for continental Europe was generated using spatiotemporal EML. The general model used to predict land cover was of the form:

\[ Y(\lambda, \phi, t) = f[X(\lambda, \phi, t), X'(\lambda, \phi), \epsilon'] \]  

(1)

where \( \lambda, \phi \) are the longitude and latitude, \( t \) is time of observation, \( m \) is the trend, \( X' \) are the “static” covariates that are assumed constant through time (such as elevation and derived variables), and \( \epsilon' \) is the stochastic component that is added to the prediction errors. The model in Eq. (1) implies that all training points are overlaid and matched either in spacetime (time-series of rasters) or space only (single list of covariates).

In this work, for dynamically changing covariates we used harmonized and gap-filled Landsat bands (Blue, Green, Red, NIR, SWIR1, SWIR2 and Thermal) derived for each of the four season (hence 20×4 images) and Suomi-NPP Visible Infrared Imaging Radiometer Suite (VIIRS) night light images.
down-scaled from 500 m to 30 m using cubic-splines. For static covariates, we used continental EU

DTM derivatives elevation and slope in percent (Hengl et al., 2021), and the 30+ year probability of water occurrence (Pekel et al., 2016).

As an additional space-time varying covariate we used the geometric Earth surface minimum and maximum temperature, which can be defined universally anywhere on globe by using (Kilibarda et al., 2014):

\[
\begin{align*}
t_{\text{min}} &= 24.2 \cdot \cos \phi - 15.7 \cdot (1 - \cos \theta) \cdot \sin |\phi| - 0.6 \cdot \frac{z}{100} \\
t_{\text{max}} &= 37 \cdot \cos \phi - 15.4 \cdot (1 - \cos \theta) \cdot \sin |\phi| - 0.6 \cdot \frac{z}{100}
\end{align*}
\]

where \( \theta \) is derived as:

\[
\theta = (\text{day} - 18) \cdot \frac{2\pi}{365} + 2^{1 - \text{sgn}(\phi)} \cdot \pi.
\]

where \( \text{day} \) is the day of year, \( \phi \) is the latitude, the number 18 represents the coldest day in the northern and warmest day in the southern hemisphere, \( z \) is the elevation in meter, 0.6 is the vertical temperature gradient per 100 m, and sgn denotes the signum function that extracts the sign of a real number. In R syntax the function from Eqs.(2–4) can be implemented as:

```r
temp.from.geom <- function(f, day, a=30.419375, 
                            b=-15.539232, elev=0, t.grad=0.6) {
  f = ifelse(f == 0, 1e-10, f) 
  costeta = cos((day-18)*pi/182.5 + 2*(1-sgn(f)) * pi) 
  cosfi = cos(f*pi/180) 
  A = cosfi 
  B = (1-costeta) * abs(sin(f*pi/180)) 
  x = a*A + b*B - t.grad * elev / 100 
  return(x) 
}
```

```r
temp.from.geom(f=52, day=120) # => 8.73
```

The geometric minimum and maximum temperature can be considered "geographical covariates" because they are basically geometric transformations of latitude and day of the year. We used them because we assume that they can help Machine Learning algorithms distinguish between land cover classes under distant latitudes e.g. coniferous forest in Greece and Norway.

A detailed overview of the workflow used to fit models and produce predictions of land cover is presented in Fig. 1. It was implemented in Python and R programming languages, and is publicly available, under an Apache-2 license, through the eumap library. The eumap package builds upon scikit learn (Pedregosa et al., 2011; Géron, 2019); with StackingClassifier as the key function used to produce
For modeling we used an ensemble of three learners:

1. Random Forest (Breiman, 2001);
2. Gradient-boosted trees (Chen and Guestrin, 2016);
3. Artificial Neural Network (McCulloch and Pitts, 1943);

These were selected among initial 10 learners which we tested on sample data first. We fine-tune and
optimize the hyperparameters for three ML models by minimizing the log-loss metric derived from a
5–fold spatial cross validation (Lovelace et al., 2019), based on a 30×30 km tilling system. We then fit
the ensemble model (meta-learner) using the best hyperparameters attached to each learner and used the
logistic regression classifier (Defazio et al., 2014) as the final meta-learner.

Although EML model fitting and prediction is at the order of magnitude more computational than
using a single learner (as in Malinowski et al. (2020) and/or Venter and Sydenham (2021)), it has shown
to bring several advantages: (1) it typically helps increase accuracy (Seni and Elder, 2010; Zhang and Ma,
2012), (2) it can be used to get a more reliable model-free estimate of the prediction uncertainty.

After the model fine-tuning and feature selection using scikit learn, we generated a single ensemble
model that was then used to predict for every pixel for each of the 20 years:

1. Probability per class (33 land cover classes);
2. the most probable land cover class;
3. probability error (uncertainty) per class derived as the standard deviation of three predicted proba-
bilities for each pixel.

Based on the predictions of the land cover classes, probabilities and input Landsat images, we also
derive for each pixel:

1. Land cover change class per year and for the 2000–2019 range;
2. slope of change fitted using logistic regression on the probabilities for key classes;
3. slope of change fitted using logistic regression for NDVI seasonal time-series.

All the output predictions (20 dominant land cover class, 660 per-class probabilities, and 660 per-class
uncertainties) were predicted first per tile, then exported as Cloud Optimized Geotiffs (COGs) files
and are publicly available through the Open Data Science Europe (ODS-Europe) Viewer, the S3 Cloud
Object Service, and from http://doi.org/10.5281/zenodo.4725429. The classification matrix
with all training points and values of covariates is available from http://doi.org/10.5281/zenodo.4740691.
Figure 1. General workflow used to prepare point data and covariate layers, fit models and generate annual land cover products (2000–2019). Components of the workflows are described in detail via the GitLab repository of the GeoHarmonizer project (https://gitlab.com/geoharmonizer_inea/).
Target land cover classification system

The target land cover nomenclature was designed based on CLC nomenclature (Bossard et al., 2000) and is available in Table 2. CLC is probably the most comprehensive and detailed European land cover product to date. The CLC program was established in 1985 by the European Commission (EC) to provide geographically harmonized information concerning the environment on the continent. The original CLC dataset is mapped in 44 classes with a minimum mapping unit of 25 ha for areal phenomena and 10 ha for changes. CLC mapping relies on harmonized protocol and guidelines that are shared for country-wise visual photo-interpretation.

Note that CLC nomenclature is complex with many classes being both land use and land cover and having nested or mixed relationships. For example, “airports” is a land use category that can include grasslands and roads. Similar confusions may appear between class pairs such as construction sites with bare rock; beaches, dunes, sands with sparsely vegetated areas; vineyards with fruit trees and olive trees; pastures with natural grasslands.

The ODSE-LULC nomenclature is based on the CLC legend, although a number of changes was applied. First, we removed heterogeneous and mixed classes defined for polygon mapping (sport and leisure facilities, complex cultivation patterns, land principally occupied by agriculture, agro-forestry areas, etc.). Next, we marked a set of CLC classes, which are potentially problematic for pixel-wise classification models and may require a contextual classification approach, for instance: Road and rail network, port areas, airports, water courses and water bodies, inland wetlands and maritime wetlands, coastal lagoons and estuaries. We did not remove these classes beforehand to avoid overly reducing our thematic resolution, and to avoid our expectations to introduce bias in our approach. The final legend in Table 2 can thus be considered a necessary compromise between being as unbiased as possible towards the capabilities of our approach, maintaining compatibility to the CLC legend, and limiting computational costs.

Covariates

All covariates used by our model are derived from remotely sensed earth observation data from multiple sources, the largest share being derived from Landsat imagery. This was obtained by downloading the Landsat ARD, provided by GLAD (Potapov et al., 2020), for the years 1999 to 2020 and for the entire extent of continental Europe (see eumap landmask (Hengl et al., 2021)). This imagery archive was screened to remove the cloud and cloud shadow pixels, maintaining only the quality assessment-QA values labeled as clear-sky according to GLAD. Second, we aggregated the individual images by season according to three different quantiles (25th, 50th and 75th) and the following calendar dates for all period:

- Winter: December 2 of previous year until March 20 of current year,
- Spring: March 21 until June 24 of current year,
- Summer: June 25 until September 12 of current year,
- Fall: September 13 until December 1 of current year,
Table 2. The ODSE-LULC land cover legend used based on CLC (Bossard et al., 2000), and number of LUCAS / CLC training samples. The distribution of training samples is shown in Fig. 3.

| Class name                        | Class description                                                                 | Number of samples |
|-----------------------------------|-----------------------------------------------------------------------------------|-------------------|
|                                   |                                     | LUCAS | CLC | Total |
| 111: Urban fabric                 | The aggregated continuous and discontinuous urban fabric class dominated by urban structures, where impermeable features cover 30–100% of the land. | 35.613 | 701.402 | 737.015 |
| 122: Road and rail networks       | Motorways and railways, including associated installations.                         | 27.892 | 13.783 | 41.675 |
| 123: Port areas                   | Infrastructure of port areas, including quays, dockyards and marinas.               | -     | 4.265  | 4.265  |
| 124: Airports                     | Airports installations: runways, buildings and associated land.                     | 109    | 6.682  | 6.791  |
| 131: Mineral extraction sites     | Areas of open-pit extraction of construction materials (sandpits, quarries) or other minerals (open-cast mines). | 14.909 | 40.011  | 54.980 |
| 132: Dump sites                   | Public, industrial or mine dump sites.                                             | 29     | 6.207  | 6.236  |
| 133: Construction sites           | Spaces under construction development, soil or bedrock excavations, earthworks.     | 532    | 6.954  | 7.486  |
| 141: Urban green                  | Areas with vegetation within urban fabric.                                          | 22.269 | 17.804 | 40.073 |
| 211: Non-irrigated arable land    | Cultivated land parcels under rain-fed agricultural use for annually harvested non-permanent crops, normally under a crop rotation system. | 318.971 | 755.454 | 1,074.425 |
| 212: Permanently irrigated arable land | Cultivated land parcels under agricultural use for arable crops that are permanently or periodically irrigated. | -     | 44.770 | 44.770 |
| 213: Rice fields                  | Cultivated land parcels prepared for rice production, consisting of periodically flooded flat surfaces with irrigation channels. | 1.773  | 3.734  | 5.507  |
| 221: Vineyards                    | Areas planted with vines.                                                          | 15.080 | 69.215 | 84.295 |
| 222: Fruit trees and berry plantations | Cultivated parcels planted with fruit trees and shrubs, including nuts, intended for fruit production. | 18.566 | 63.667 | 82.233 |
| 223: Olive groves                 | Cultivated areas planted with olive trees, including mixed occurrence of vines on the same parcel. | 19.381 | 51.402 | 70.783 |
| 231: Pastures                     | Meadows with dispersed trees and shrubs occupying up to 50% of surface characterized by rich floristic composition. | 204.042 | 704.306 | 908.348 |
| 311: Broad-leaved forest          | Vegetation formation composed principally of trees, including shrub and bush understory, where broad-leaved species predominate. | 185.954 | 751.158 | 937.112 |
| 312: Coniferous forest            | Vegetation formation composed principally of trees, including shrub and bush understory, where coniferous species predominate. | 147.552 | 739.288 | 886.440 |
| 321: Natural grasslands           | Grasslands under no or moderate human influence. Low productivity grasslands. Often in areas of rough, uneven ground, also with rocky areas, or patches of other (semi-)natural vegetation. | 56.440  | 293.825 | 350.265 |
| 322: Moors and heathland          | Vegetation with low and closed cover, dominated by bushes, shrubs (heather, briars, broom, gorse, laburnum etc.) and herbaceous plants, forming a climax stage of development. | 114.228 | 178.600 | 292.828 |
| 323: Sclerophyllous vegetation    | Bushy sclerophyllous vegetation in a climax stage of development, including maquis, maorral and garrigue. | -     | 38.683 | 138.683 |
| 324: Transitional woodland-shrub  | Transitional bushy and herbaceous vegetation with occasional scattered trees. Can represent either woodland degradation or forest regeneration / re-colonization. | -     | 724.874 | 724.874 |
| 331: Beaches, dunes, sands        | Natural un-vegetated expanses of sand or pebble/gravel, in coastal or continental locations, like beaches, dunes, gravel pads. | 9.900  | 15.275 | 25.175 |
| 332: Bare rocks                   | Scrce, cliffs, rock outcrops, including areas of active erosion.                    | 8.138  | 67.173 | 75.311 |
| 333: Sparsely vegetated areas     | Areas with sparse vegetation, covering 10-50% of the surface.                       | -     | 221.421 | 221.421 |
| 334: Burnt areas                  | Areas affected by recent fires.                                                     | -     | 2.175  | 2.175  |
| 335: Glaciers and perpetual snow  | Land covered by ice or permanent snowfields.                                      | 319    | 6.954  | 7.273  |
| 411: Inland wetlands              | Low-lying land usually flooded in winter, partly saturated by water; and wetlands with considerable amount of decomposed moss and vegetation matter. | 17.158 | 256.401 | 273.559 |
| 421: Maritime wetlands            | Vegeated low-lying coastal areas above the high-tide line, susceptible to seawater flooding; salt-pans for salt extraction, and tidal coastal zones. | 1.094  | 15.126 | 16.220 |
| 511: Water courses                | Natural or artificial water courses for water drainage channels.                   | 6.030  | 8.582  | 14.612 |
| 512: Water bodies                 | Natural or artificial water surfaces covered by standing water most of the year.   | 20.038 | 198.979 | 219.017 |
| 521: Coastal lagoons              | Stretches of salt or brackish water in coastal areas which are separated from the sea by a tongue of land or other similar topography. | 1.401  | 2.270  | 3.671  |
| 522: Estuaries                    | The mouth of a river under tidal influence within which the tide ebbs and flows.     | -     | 1.241  | 1.241  |
| 523: Sea and ocean                | Zone seaward of the lowest tide limit.                                             | -     | 595    | 595    |
From more than 73 TiB of input data we produced 84 images (3 quantiles × 4 seasons × 7 Landsat bands) for each year with different occurrences of no-data values due to cloud contamination in all observations of a specific season.

We next impute all missing values in the Landsat temporal composites using the “Temporal Moving Window Median” TMWM algorithm we implemented in python. The algorithm works as follows: we first detect gaps and artifacts in the Landsat data, next we iteratively interpolate values from the temporal neighbours until all pixels have values using the median function. For gap-filling we prioritize observations of: 1—the same season, 2—neighboring seasons and 3—all the year.

The TMWM has been selected among 3–4 alternative algorithms by bench-marking, and to our knowledge provides the best combination of gap-filling accuracy and computational costs. It is publicly available in the eumap library.

In addition to the Landsat data, we also used as covariates:

- Landsat spectral indices: Normalized Difference Vegetation Index (NDVI), Soil Adjusted Vegetation Index (SAVI), Modified Soil Adjusted Vegetation Index (MSAVI), Normalized Difference Moisture Index (NDMI), Landsat Normalized Burn Ratio (NBR), NBR2, and Normalized Difference Water Index (NDWI) derived according to formulas in Table 3;

- Time-series SUOMI NPP VIIRS night lights at 500 m resolution downscaled to 30 m resolution (Hillger et al., 2013);

- Global surface water frequency at 30 m resolution (Pekel et al., 2016);

- Continental EU DTM-based elevation and slope (Hengl et al., 2021);

- Geometric minimum and maximum temperature derived for every pixel using Eq.(2);

The Landsat spectral indices and the geometric minimum and maximum temperatures were calculated on-the-fly during the prediction step, avoiding the overhead to maintain this large amount of data in storage media.
Table 3. Spectral indices derived from the Landsat data and used as additional covariates in the spatiotemporal EML.

| Spectral Index | Equation | Reference |
|---------------|----------|-----------|
| NDVI          | \( \frac{nir - red}{nir + red} \) | (Tucker, 1979) |
| SAVI          | \( \frac{nir - red}{(nir + red + 0.5) \times 1.5} \) | (Huete, 1988) |
| MSAVI         | \( \frac{(2 \times nir + 1) - \sqrt{(2 \times nir + 1)^2 - 8 \times (nir - red)^2}}{2} \) | (Qi et al., 1994) |
| NDWI          | \( \frac{green - swir2}{green + swir2} \) | (Gao, 1996) |
| NBR           | \( \frac{nir - thermal}{nir + thermal} \) | (Key and Benson, 1999) |
| NDMI          | \( \frac{nir - swir1}{nir + swir1} \) | (Jin and Sader, 2005) |
| NBR2          | \( \frac{swir1 - thermal}{swir1 + thermal} \) | (Key and Benson, 2006) |

Training points

Point data preparation

We obtained the training dataset from the geographic location of LUCAS (in-situ source) and the centroid of all CLC polygons (as shown in Fig. 2), harmonized according to the 33 land cover classes (see Table 2) and organized by year, where each unique combination of longitude, latitude and year was considered as a independent sample, resulting in more than 7 million training points.

The LUCAS data from 2006, 2009, 2012, 2015 and 2018, as provided by Eurostat (obtained from: https://ec.europa.eu/eurostat/web/lucas) is the largest and most comprehensive in-situ land cover data set for Europe. The survey has evolved since 2000 and requires harmonisation before it can be used for mapping over several years. We imported data sets from individual years and harmonized these before merging it into one common database with an automated workflow implemented in Python and SQL (Fig. 1). For the multi-year harmonization procedure we first harmonized attribute names, re-coded variables, harmonized point locations, and aggregated the points based on their location in space and time. After these operations, we translated the LUCAS land cover nomenclature to the ODSE-LULC nomenclature, Table 2, according to the method designed by Buck et al. (2015).

We assigned the LUCAS points with a unique land-cover class a confidence rating of 100%, while the CLC points received 85% confidence. These confidence weights were considered in the ML training step and the points were used in the spacetime overlay. For example, a sample from 2018 would retrieve Landsat composites with reference to the same year. The distribution of all training points is shown in Fig. 3.

Filtering

The CLC minimal mapping unit of 25 ha required filtering on the training points before they could be used to represent 30 m resolution LULC, for example, to remove points for “111: urban fabric” located...
**Figure 2.** General workflow for merging training points obtained from LUCAS and CLC.
Figure 3. Training points distribution based on LUCAS, CLC and their combination. Class codes are described in Table 2.
in small patches of urban greenery (<25 ha). For this purpose, we extracted vector data from OSM layers for roads, railways, and buildings (obtained from https://download.geofabrik.de/), filtering them according to their crowd source-derived names and the following steps:

- Extract all possible values of the most descriptive variable in the OSM data;
- Sum the number of occurrences and the total area covered by vector features of each type over all countries;
- Assign each type to a category until at least 99% of vector features are categorized.

After this categorization step, the OSM layers were first rasterized to 10 m resolution by assigning cells located in vector features the value 100. These rasters were then averaged to 30 m to create a 0—100 density layer for the three feature types. The building density raster was then combined with Copernicus High Resolution Layers (HRL) (obtained from https://land.copernicus.eu/pan-european/high-resolution-layers) to supplement its coverage where crowd-sourced OSM information was unavailable.

In the OSM, buildings are digitized at highest level of detail but unfortunately is incomplete for whole of Europe. To improve the coverage of the built-up areas, we first fill the non-mapped areas in OSM with the Impervious Built-up 2018 pixel values, which was averaged to 30 m and rescaled to 101—200 values. We also use HRL products to filter other classes: Table 4 shows the exact conditions points of specific LULC classes needed to meet in order to be retained in our dataset. This is a similar procedure to the one used by Inglada et al. (2017). This filtering process removed about 1.3 million points from our training dataset, resulting in a classification matrix with a total of ca. 5.3 million samples and 178 covariates. The classification matrix used to produce ODSE-LULC is available from http://doi.org/10.5281/zenodo.4740691.

**Table 4.** Per-class conditions applied only to CLC points during the filtering step. All the raster layers were upsampled to 30 m² by average and the points that did not meet the specified condition were omitted from the training dataset.

| Class | Condition | Copernicus HRL | OSM | HRL + OSM |
|-------|-----------|----------------|-----|-----------|
| 111   | -         |                |     | >50 and <150 |
| 122   | OR        | >30            | >30 | >30       |
| 131   | AND equals 0 | equals 0      |     |           |
| 141   | OR equals 0 | >0             | >0  |           |
| 211   | AND equals 0 |                |     |           |
| 221   | AND equals 0 |                |     |           |
| 222   | AND equals 0 |                |     |           |
| 223   | AND equals 0 |                |     |           |
| 231   | AND equals 0 |                |     |           |
| 311   | AND >0     |                |     |           |
| 312   | AND >0     |                |     |           |
| 321   | AND equals 0 |                |     |           |
| 411   | OR equals 0 | >0             | >0  |           |
| 512   | -         | equals 100     |     |           |

**Comparison with other land cover products**

We compare a number of existing land cover products to the harmonized training dataset used by our model to (see Table 1):
• GLC FCS30–2015 (Zhang et al., 2020);
• GLC FCS30–2020 (Zhang et al., 2020);
• S2GLC (Malinowski et al., 2020);
• The European land cover product for 2015 created by Pflugmacher et al. (2019);
• ELC10 (Venter and Sydenham, 2021).

For each comparison, we reclassified the training dataset to the nomenclature of the target dataset. We overlaid our LUCAS and CLC training points from years adjacent to each land cover product. We then performed a classification accuracy assessment, calculating the associated accuracy metrics. Points with classes of the target products that were completely absent in the training point subsets (due to the target nomenclature of the training points) were removed before these assessments.

The GLC FCS30 nomenclature was not suitable for direct translation because some land cover groups (such as forests) are separated in several subcategories. We therefore aggregated their thematic resolution to the higher level of abstraction described in Zhang et al. (2020). The complete translation scheme is available via the GitLab repository of the GeoHarmonizer project (https://gitlab.com/geoharmonizer_inea/spatial-layers).

### Accuracy assessment

We assess performance of our final ensemble model in two ways. Firstly through spatial 5-fold cross-validation, and secondly by validating the final model predictions on an independently collected test dataset. In all comparisons and experiments, we discriminate model performance with the Weighted F1-score metric (Van Rijsbergen, 1980):

\[
WF_1 = \sum_{c=1}^{n} S_c \cdot \frac{2 \cdot P_c \cdot R_c}{P_c + R_c}
\]

where \(n\) is the number of classes, and \(S_c\) is the support, \(P_c\) the precision, and \(R_c\) the recall of a given class \(c\). We used this metric because it distinguishes classification performance more strictly on imbalanced datasets.

### Spatial cross-validation

We performed spatial 5-fold cross-validation using the hyperparameters of the final EML model and assessed its performance based on its cross-validation predictions. During the hyperparameter optimization step, we saved the spatial 5-fold cross-validation predictions of the model with the hyperparameters that would be used to train the final model on all available data. We merged these predictions into one dataset, which we treated as an independent prediction. Because no test dataset was available that had the same thematic resolution as our model, we used this validation step to assess its performance on a class-by-class basis. This validation strategy also allowed us to assess average model performance per year. The hierarchical nature of the target nomenclature allows for accuracy assessment at three levels.
For each dataset size ranging from 1000 to 500,000 points, we created a training dataset from CLC data for 2000, 2006, and 2012. Each training dataset was accompanied by a validation dataset half its size with points from the same year. Lastly, for each dataset size, we created a validation set with points from 2018. For each dataset size, we trained a spatial model on data from each separate year, as well as a spatiotemporal model on all data from the three years.

Independent test data

After training an ensemble model with the same hyperparameters on all training data, we classified LULC in 2017. This prediction was validated with the S2GLC dataset (Malinowski et al., 2020). This dataset contains 51,926 points with human-verified land cover classifications, which were collected and used for the same purpose by Malinowski et al. (2020). We performed a spatial overlay with these points on the ODSE-LULC 2017 predictions.

As the S2GLC points follow a different nomenclature, we translated the ODSE-LULC predicted classes according to Table 5. As our model was not trained to predict peat bogs, we removed all points with this land cover class from the validation dataset. In addition, because any predicted classes outside the S2GLC nomenclature would be automatically counted as errors, we performed two validations: (1) a conservative assessment that included points with such predictions, and (2) an optimistic assessment where they were omitted.

Comparison of spatial and spatiotemporal models

We include two experiments to compare model performance depending on the temporal nature of the training data and either the data source or the size of the training dataset. In both cases we trained spatial (trained on spatial data from one year) models and spatiotemporal (trained on spatial data from multiple years) models, and compared their performance both when predicting on years that were included in their training dataset, and on data from 2018, which we omitted from all training datasets. We averaged the performance of all spatial models to obtain the performance of one ‘spatial model’ approach.

To investigate the effect of different temporal extents and data source combinations, we compiled three datasets: One with only CLC points, one with only LUCAS points, and one with points from both datasets, resembling the composition of our full dataset.

We compared the performance of multiple spatial models trained on data from one year with the performance of a spatiotemporal model trained on a separate, but equally-sized multi-year dataset. We also trained a spatiotemporal model on all spatial model training data to quantify the performance gain from having access to more training points than spatial models. Each model was validated on data from each year, a multi-year validation set, and data from a year that was completely excluded from the training datasets.

To investigate the effect of different temporal extents and training dataset size, we created 9 training datasets from the best-performing data source with sizes ranging between 1000 and 500,000 points from 2000, 2006 and 2012. We accompanied each training dataset by a validation set half its size. We also made nine validation datasets with data from 2018 to compare model performance on data from previously unseen years. For each dataset size, three spatial models were trained on data from the three separate years, and a spatiotemporal model was trained on all data for each year. The models were validated on the
Table 5. Reclassification key used to validate the predictions of our ensemble model on the test set collected by Malinowski et al. (2020).

| S2GLC | ODSE-LULC |
|-------|-----------|
|       | 111: Urban fabric |
|       | 122: Road and rail networks and associated land |
|       | 123: Port areas |
|       | 124: Airports |
|       | 132: Dump sites |
|       | 133: Construction sites |

Artificial surfaces

|       | 311: Broad-leaved forest |
|       | 312: Coniferous forest |

Broadleaf tree cover

|       | 211: Non-irrigated arable land |
|       | 212: Permanently irrigated arable land |
|       | 213: Rice fields |

Coniferous tree cover

|       | 231: Pastures |
|       | 321: Natural grasslands |

Cultivated areas

|       | 411: Inland wetlands |
|       | 421: Maritime wetlands |

Herbaceous vegetation

|       | 311: Broad-leaved forest |
|       | 312: Coniferous forest |

Marshes

|       | 311: Broad-leaved forest |
|       | 312: Coniferous forest |

Moors and heathland

|       | 311: Broad-leaved forest |
|       | 312: Coniferous forest |

Natural material surfaces

|       | 511: Water courses |
|       | 512: Water bodies |
|       | 521: Coastal lagoons |
|       | 522: Estuaries |
|       | 523: Sea and ocean |

None (not in nomenclature)

|       | 131: Mineral extraction sites |
|       | 331: Beaches, dunes, sands |
|       | 332: Bare rocks |

Peatbogs

|       | 335: Glaciers and perpetual snow |

Permanent snow

|       | 321: Natural grasslands |

Sclerophyllous vegetation

|       | 323: Sclerophyllous vegetation |

Vineyards

|       | 221: Vineyards |

Water bodies

|       | 222: Fruit trees and berry plantations |
|       | 223: Olive groves |
|       | 332: Bare rocks |

Time-series analysis

We analysed changes using NDVI and probability trend analysis, LULC change as hard classes, and finally methods to aggregate changes over larger areas. The time-series analysis has resulted in 13 data layers for NDVI slopes, 3 for probability slopes two layers for every two years for land use change classes plus two layers for changes between the years 2001–2018 and one layer for aggregated change classes.
and one for intensity of change. Here we discuss a subsection of examples.

**NDVI and LULC probability trend analysis**

For both the NDVI trend analysis and the probability trend analysis we analysed the trend over the years between 2000 and 2019. NDVI is widely used as a proxy for chlorophyll concentration and hence as an indicator for vegetation health.

We fitted a regression for the time-series of every pixel (e.g. 20×4 of NDVI images) to map the slope of change through time. Before applying the OLS estimate of the beta coefficients, we applied a logit transformation to the input data because probabilities only have meaningful values between 0 and 1 and NDVI are only meaningful for values between -1 and 1.

We applied OLS in parallel and then saved the slope, intercept, and R-squared for each pixel in Europe. For NDVI values we also deseasonalize the data before deriving the slope by applying a loess regression for seasonal decomposition (Cleveland et al., 1990) as implemented in the python statsmodels library (Seabold and Perktold, 2010). An example in Python code can be found below:

```python
from statsmodels.tsa.seasonal import STL
import statsmodels.formula.api as smf
import pandas as pd
import numpy as np
import math

def logit(p):
    return math.log(p/(1-p))

def slope_analysis(ndvi_timeseries):
    # apply deseasonalization
    ndvi = pd.Series(np.array(ndvi_timeseries))
    stl = STL(ndvi, period = 4, robust = True)
    res = stl.fit()

    # apply logit transformation
    data_norm = [logit(x) for x in res.trend]
    df_ols = pd.DataFrame(data_norm).reset_index()
    df_ols.columns = ['t', 's']

    # apply OLS
    model = smf.ols('s~t', df_ols)
    results = model.fit()
    a, b = results.params
```
For the analysis of probability slopes a similar function is applied with only the deseasonalization removed. Fig. 4A, and B show an example of the application and result of this method on a single pixel time-series in Björnrike, Sweden. Note computation of slopes is computationally very intensive often taking couple of days of continuous computing.

We applied a trend analysis on probability slopes for the four most prevalent LU classes: (1) coniferous forest, (2) non-irrigated arable land, (3) broad leaved forest, and (4) pastures. Fig. 4C illustrates this process for a single pixel time-series in Björnrike, Sweden.

**Figure 4.** Example of deseasonalization (Seabold and Perktold, 2010) and subsequent Logit Ordinary Least Squares (OLS) applied on a single pixel in Sweden (coordinates: 62°24'43.7"N 13°56'00.3"E). a) red dots represent pixel values, blue line represents a local weighted regression smoothed line based on the pixel values plus a light blue area indicating the confidence interval, the red line represents the trend after removing the seasonal signal. b) red line and crosses represent the trend after removing the seasonal signal, the blue line visualizes the regression model based NDVI values in the logit space. c) Trend analysis on probability values for non-irrigated arable land. In the case above the gradient value is 0.09 with the model R-square = 0.88

**LULC change classes**

The simplest way to detect changes is to analyze difference land cover data from two years. However, this method propagates classification errors to consequent change maps and metrics (Carmel et al., 2001), which can lead to significant overestimation of change (Olofsson et al., 2014). To avoid this we follow these steps:
• Perform change detection on the input data before classifying land cover (Zhu and Woodcock, 2014);

• implement post-processing rules that prohibit specific types of transformations (Song et al., 2016);

• perform change detection and smooth over any undetected changes in predicted land cover (Li et al., 2018).

For the hard class change analysis we applied a comparison between two subsequent years. We applied one temporal post-processing step in which we consider the classification of two neighboring years for every year between 2001 and 2018. In order to maximize the re-usability of our results, we only performed limited post-processing: If both neighboring years are the same class and the class under consideration is another class we assume this classification is an error and match the middle year with its neighboring years. We call this a “T-3 temporal filter”. Both the filtered and the noise layers are saved for further analysis and interpretation.

We mirror the change classes seen in the Copernicus land cover map (Buchhorn et al., 2020) and described in Table 6. The change classes applied by the Copernicus land cover map, however, use classes of a higher abstraction level. Therefore we translate the CLC classes to the land use classes used by the Copernicus land cover map. Some examples of changes include: changing from Dump sites into Urban fabric is classified as “No change”, changing from Non-irrigated arable land into Urban fabric to “Urbanization”, changing from Airports to Mineral extraction sites to “Other” etc.

Prevalent change

We also mapped prevalent change on a 5×5 km grid and the change intensity on 20×20 km grid. We divide the entire area of Europe into 5×5 km grids and count the number of pixels for each change class within these blocks. The change class that covers the biggest amount of pixels is then assigned to the corresponding pixel / grid node. We also save the number of pixels that the most prevalent change class covers, as part of the total of pixels in the area for 20×20 km areas. For example: we used 30×30 m resolution data, in each 20×20 km block we have \((20,000/30) \cdot (20,000/30) = 444,444\) pixels. If the prevalent change class covers >94,000 pixels this means that it covers >20% of the total area.
Table 6. Harmonization scheme used to convert ODSE-LULC nomenclature to Copernicus Global Land Cover classes. On the left side, ODSE-LULC classes are converted to Forest, Other Vegetation, Wetland, Bare, Cropland, Urban, and Water classes. Each transition from one Copernicus class to another is then categorized into a change class in the cross-table.

| ODSE-LULC class | Copernicus change class | Forest | Other Vegetation | Wetland | Bare | Cropland | Urban | Water |
|-----------------|-------------------------|--------|-----------------|---------|------|----------|-------|-------|
| 311: Broad-leaved forest | Forest | Deforestation | | | | | | |
| 312: Coniferous forest | Forest | Deforestation | | | | | | |
| 311: Broad-leaved forest | Other Vegetation | Other | Desertification | Crop expansion | Urbanization | | | |
| 311: Broad-leaved forest | Wetland | Wetland degradation | Wetland degradation | Crop expansion | Wetland degradation | | | |
| 311: Broad-leaved forest | Bare | | Desertification | Crop expansion | Urbanization | | | |
| 311: Broad-leaved forest | Cropland | Reforestation | | | | | | |
| 311: Broad-leaved forest | Urban | | Land abandonment | Land abandonment and desertification | Urbanization | | | |
| 311: Broad-leaved forest | Water | | | | Water reduction | | | |
RESULTS

Comparison of training data to other land cover products

Table 7 provides an overview of each compared land cover product’s accuracy when validated on subsets of ODSE-LULC training data. The S2GLC product scored the highest in both 2016 and 2018 subsets of our dataset, while the land cover product made by Pflugmacher et al. (2019) fits more closely to the 2015 subset. The 2019 point subset was considered too small to perform any meaningful comparison between ELC10 and GLC FCS30.

Table 7. Weighted F1-score of other land cover products when validated with the ODSE-LULC training dataset.

| Land cover product | Validation year | Data source | Samples | Weighted F1-Score | Number of classes | Res. (m) |
|--------------------|-----------------|-------------|---------|-------------------|-------------------|---------|
| S2GLC              | 2016            | LUCAS       | 756     | 0.724             | 8                 | 10      |
| Pflugmacher et al. (2019) | 2016 | LUCAS | 719 | 0.719 | 10 | 30 |
| GLC FCS30–2015     | 2016            | LUCAS       | 724     | 0.677             | 10                | 30      |
| Pflugmacher et al. (2019) | 2015 | LUCAS | 144,027 | 0.657 | 11 | 30 |
| S2GLC              | 2018            | CLC         | 1,000,063 | 0.604 | 12 | 10 |
| ELC10              | 2018            | LUCAS       | 42,629  | 0.596             | 8                 | 10      |
| GLC FCS30–2015     | 2015            | LUCAS       | 138,342  | 0.503             | 12                | 30      |
| ELC10              | 2018            | CLC         | 172,382  | 0.456             | 8                 | 10      |
| GLC FCS30–2020     | 2018            | LUCAS       | 308,838  | 0.424             | 12                | 30      |
| GLC FCS30–2020     | 2018            | CLC         | 1,026,914 | 0.420 | 12 | 30 |

Spatiotemporal model

The result of the EML model optimization resulted in the following hyperparameters:

- Random forest: Number of trees equal to 85, maximum depth per tree equal to 25, number of covariates to find the best split equal to 89, and 20 as minimum number of samples per leaf.

- Gradient boosted trees: Number of boosting rounds equal to 28, maximum depth per tree equal to 7, minimum loss reduction necessary to split a leaf node equal to 1, L1 regularization term on weights equal to 0.483, learning rate equal to 0.281, greedy histogram algorithm to construct the trees, and softmax as objective function.

- Artificial Neural Network: Four fully connected hidden layers with 64 artificial neurons each; ReLU as activation function, dropout rate equal to 0.15 and batch normalization in all the layers; softmax as activation function for output layer; batch size and number of epochs equal to 64 and 50, respectively; and Adam with Nesterov momentum as optimizer considering 5e-4 as learning rate.

- Logistic Regression: SAGA solver and multinomial function to minimize the loss.

The variable importance, generated by the tree based algorithms and presented in Fig. 5, shows that the 50th quantile of Landsat green band is the most important covariate, derived on summer and fall. In addition to spectral bands, several Landsat spectral indices (NDVI, SAVI, MSAVI, NBR, NB2 and
NDWI) appeared among the 50 most important covariates, while the global surface water frequency was considered the second most important for Random Forest and the 7th for GBT. This proves that variations of NDVI data have significant impact on the predictive power of the models ranking as the 8th, 9th and 10th most important variables for RF models and 3rd most important variable for the GBT models (Fig. 5).

The geometric temperatures and terrain covariates were ranked more important for Random Forest, with the slope and elevation in 6th and 7th ranking position, respectively, while for the gradient boosted trees only the slope appears in the result of the analysis, in the 26th position of the ranking. These differences in the variable importance indicate that a wider usage of the feature space could lead to a better prediction power for the EML model, compared to using single learners.

![Figure 5. Variable importance for the top-50 covariates according to the (a) Random Forest and (b) Gradient Boosted Trees. The covariate names are composed by theme: land cover (lcv), digital terrain model (dtm), climate (clm) and hydrology (hyd); variable code: spectral bands (green, red, nir, swir1, and, swir2), spectral indices (ndvi, savi, msavi, nbr, nbr2, and ndwi), land surface temperature (lst), and openness positive negative derived (openp and openn). For the Landsat covariates the 25th, 50th and 75th quantiles are represented by p25, p50 and p75, respectively.](image-url)
Accuracy assessment of spatiotemporal models

Spatial 5-fold cross-validation

The results of accuracy assessment using 5-fold cross-validation show that the ensemble achieved 62% accuracy for 33 CLC level 3 classes, 70% accuracy for 14 level 2 classes, and 87% accuracy for the 5 level 1 classes (see also Table 9). Overall these results are consistent with similar studies by Calderón-Loor et al. (2021) and Tsendbazar et al. (2018).

Most false positive errors were recorded for classes 211: Non-irrigated arable land (16%), 324: Transitional Woodland-shrub (14.8%), 311: Broad-leaved forest (14.2%) and 231: Pastures (13.7%). Most false negative errors were for the same classes in a different order, and: 324 (17.9%), 311 (9.5%), 231 (8.7%), and 211 (8.6%). Class 1: Artificial surfaces had the lowest precision (0.78) and recall (0.65) scores of all level 1 classes. The relatively low recall value is mainly due to false positives for class 211: Non-irrigated arable land. Its most accurately predicted subclass was 111: Urban fabric (0.68). This class had a precision of 59% and a positive rate of 1.38; the main cause are false positives on other classes inside its level 1 class. Class 2: Agricultural areas had a precision of 0.81 and a recall of 0.9. The main causes of this relatively low precision score are false positives for class 211: Non-irrigated arable land in non-forested, non-water body classes, and for class 231: Pastures in sub-classes of class 3: Forests and seminatural areas. Class 3 had the highest number of classes and the largest support in the training data. The largest cause for error within the class are confusions between 311: Broad-leaved forest, 312: Coniferous forest, and 324: Transitional woodland-shrub. The highest accuracy (0.93) was achieved for class 512: Water bodies. This class and 421: Maritime Wetlands were the main false positive class for other open water classes (511, 521, 522, 523). Table 8 shows the spatial cross-validation performance separated per year when assessed on full thematic resolution. The average accuracy per year was 0.61, with a standard deviation of 0.035.

Table 8. Spatial cross-validation performance (Weighted F1-score) of our ensemble model per year in the training data.

| Year | Number of points | Weighted F1-score |
|------|------------------|-------------------|
| 2000 | 1.034.449        | 0.652             |
| 2006 | 1.228.471        | 0.645             |
| 2009 | 225.515          | 0.585             |
| 2012 | 1.333.239        | 0.637             |
| 2015 | 147.554          | 0.589             |
| 2016 | 807              | 0.562             |
| 2018 | 1.392.044        | 0.628             |
| 2019 | 149              | 0.562             |
| Average |                     | 0.607             |
| Standard deviation |                  | 0.035             |

Independent test data

Overlaying our 2017 LULC predictions with the S2GLC dataset, reclassing them to the S2GLC nomenclature, and removing any points with class 412: Peatbogs resulted in 51,279 data points. 2914 points had a predicted class that was not in the S2GLC nomenclature (see Table 5). The ‘conservative’ assessment
Table 9. Classification report for three levels of thematic resolution: 33 classes (level 3), 14 classes (level 2), and 5 classes (level 1)

| Class Level 1 | Level 2 | Level 3 | Precision | Recall | F1-score | Support |
|---------------|---------|---------|-----------|--------|----------|---------|
| 1: Artificial surfaces | 11: Urban fabric | 111: Urban fabric | 0.59 | 0.59 | 0.78 | 146459 |
| | 12: Industrial, commercial and transport units | 122: Road and rail networks and associated land | 0.49 | 0.45 | 0.65 | 146459 |
| | | 124: Airports | 0.29 | 0.78 | 0.68 | 308799 |
| | 13: Mine, dump and construction sites | 131: Mineral extraction sites | 0.51 | 0.58 | 0.57 | 146459 |
| | | 132: Dump sites | 0.35 | 0.57 | 0.12 | 6745 |
| | | 133: Construction sites | 0.20 | 0.02 | 0.03 | 15776 |
| | 14: Artif. non-agri. veg. areas | 141: Green urban areas | 0.25 | 0.25 | 0.16 | 33196 |
| 2: Agricultural areas | 21: Arable land | 211: Non-irrigated arable land | 0.66 | 0.78 | 0.72 | 830748 |
| | | 212: Permanently irrigated arable land | 0.42 | 0.79 | 0.26 | 32637 |
| | | 213: Rice fields | 0.70 | 0.80 | 0.65 | 14840 |
| | 22: Permanent crops | 221: Vineyards | 0.52 | 0.81 | 0.43 | 51027 |
| | | 222: Fruit trees and berry plantations | 0.38 | 0.14 | 0.21 | 49432 |
| | | 223: Olive groves | 0.45 | 0.41 | 0.44 | 49628 |
| | 23: Pastures | 231: Pastures | 0.62 | 0.62 | 0.66 | 624412 |
| 3: Forests and seminatural areas | 31: Forest | 311: Broad-leaved forest | 0.67 | 0.78 | 0.71 | 773914 |
| | | 312: Coniferous forest | 0.74 | 0.74 | 0.74 | 664949 |
| | | 321: Natural grasslands | 0.38 | 0.30 | 0.34 | 209897 |
| | | 322: Moors and heathland | 0.49 | 0.33 | 0.35 | 263842 |
| | | 323: Sclerophyllous vegetation | 0.39 | 0.57 | 0.40 | 1265286 |
| | | 324: Transitional woodland-shrub | 0.49 | 0.88 | 0.47 | 659396 |
| | | 331: Beaches, dunes, sands | 0.51 | 0.23 | 0.31 | 23292 |
| | | 332: Bare rocks | 0.65 | 0.39 | 0.49 | 64101 |
| | | 333: Sparsely vegetated areas | 0.51 | 0.52 | 0.47 | 162694 |
| | | 334: Burnt areas | 0.24 | 0.00 | 0.00 | 12036 |
| | | 335: Glaciers and perpetual snow | 0.85 | 0.82 | 0.84 | 6883 |
| 4: Wetlands | 41: Inland wetlands | 411: Inland wetlands | 0.70 | 0.70 | 0.71 | 208087 |
| | | 421: Maritime wetlands | 0.62 | 0.62 | 0.64 | 15403 |
| 5: Water bodies | 51: Inland waters | 511: Water courses | 0.29 | 0.08 | 0.13 | 12830 |
| | | 512: Water bodies | 0.91 | 0.93 | 0.93 | 185788 |
| | | 521: Coastal lagoons | 0.51 | 0.93 | 0.41 | 2597 |
| | | 522: Estuaries | 0.37 | 0.11 | 0.17 | 1720 |
| | | 523: Sea and ocean | 0.52 | 0.31 | 0.31 | 481 |

Accuracy: 0.62, Macro average: 0.62, Weighted average: 0.60.
(on 50,939 points) including the non-S2GLC classes resulted in a weighted F1-score of 0.877, a macro F1-score of 0.741 and a kappa score of 0.828. The 'optimistic assessment excluding non-S2GLC predictions resulted in a weighted F1-score of 0.903, a macro F1-score of 0.772 and a kappa score of 0.882 (see Table 10).

Taking into account possible noise from the translation process, these results are similar to those reported by Malinowski et al. (2020). Weighted precision, recall and F1-scores are also higher than our cross-validation scores at all thematic resolution levels (see Table 9).

### Table 10. Optimistic classification report of our 2017 LULC prediction on 48,365 S2GLC points after removing 2914 points with predicted classes without an equivalent S2GLC class (141: Green urban areas, 222: Fruit trees and berry plantations, 223: Olive groves, 324: Transitional woodland-shrub, 333: Sparsely vegetated areas, and 334: Burnt areas).

| S2GLC Class             | Precision | Recall | F1-Score | Support |
|-------------------------|-----------|--------|----------|---------|
| Artificial surfaces     | 0.923     | 0.891  | 0.906    | 1821    |
| Cultivated areas        | 0.914     | 0.946  | 0.930    | 13404   |
| Vineyards               | 0.771     | 0.754  | 0.762    | 479     |
| Herbaceous vegetation  | 0.798     | 0.811  | 0.805    | 6486    |
| Broadleaf tree cover    | 0.959     | 0.964  | 0.962    | 9956    |
| Coniferous tree cover   | 0.965     | 0.967  | 0.966    | 8523    |
| Moors and heathland     | 0.747     | 0.618  | 0.676    | 1484    |
| Sclerophyllous vegetation | 0.686    | 0.346  | 0.460    | 619     |
| Natural material surfaces | 0.884   | 0.864  | 0.874    | 1821    |
| Permanent snow          | 0.544     | 0.840  | 0.660    | 81      |
| Marshes                 | 0.234     | 0.548  | 0.328    | 305     |
| Water bodies            | 0.996     | 0.876  | 0.932    | 3386    |
|                         |           |        |          | 48365   |
| Macro average           | 0.785     | 0.785  | 0.772    |         |
| Weighted average        | 0.907     | 0.903  | 0.903    |         |

### Comparison of spatial and spatiotemporal models

Table 11 shows the weighted F1-scores of spatial models trained on 100,000 points sampled from one year, spatiotemporal models trained on 100,000 points sampled from across multiple years, and spatiotemporal models trained on 100,000 points per year. Each model was validated on 33,333 points from the same year(s) as its training data, and on 33,000 points from the year 2018, which was left out of all training datasets.

Results show that all models performed better when classifying in known years, regardless of data source. The spatiotemporal model trained on only CLC points achieved the highest F1-scores for both known-year and unknown-year classification. This model outperformed spatial models on known-year classification by 2.7% and unknown-year classification by 3.5% as seen in Table 11.

Fig. 6 shows that spatial models achieved a higher F1 score than spatiotemporal models when classifying LULC on the validation set from the year they were trained on, and that spatiotemporal models outperformed spatial models when generalizing to the 2018 validation datasets.
Table 11. Trained year and untrained year weighted F1-scores of spatial and spatiotemporal models trained on CLC points, LUCAS points, and a combination of both.

| Model                        | CLC Trained year(s)   | CLC Trained year(s) | LUCAS Trained year(s) | LUCAS Trained year(s) | CLC+LUCAS Trained year(s) | CLC+LUCAS Trained year(s) |
|------------------------------|-----------------------|---------------------|-----------------------|-----------------------|---------------------------|---------------------------|
| Spatial - 2000               | 0.610 0.542           | 0.610 0.542         | 0.611 0.515           | 0.611 0.515           |
| Spatial - 2006               | 0.595 0.437           | 0.604 0.563         | 0.587 0.534           | 0.587 0.534           |
| Spatial - 2009               | 0.595 0.482           | 0.611 0.574         | 0.602 0.415           | 0.602 0.415           |
| Spatial - 2012               | 0.583 0.415           | 0.608 0.560         | 0.565 0.529           | 0.565 0.529           |
| Spatial - Average            | 0.583 0.465           | 0.608 0.560         | 0.591 0.498           | 0.591 0.498           |
| Spatiotemporal - 100.000 pts.| 0.612 0.576           | 0.568 0.478         | 0.574 0.532           | 0.574 0.532           |
| Spatiotemporal - 100.000 pts. per year | **0.625 0.579** | **0.608 0.491** | **0.595 0.543** | **0.595 0.543** |

Figure 6. Performance comparison of spatial and spatiotemporal models when classifying LULC with varying training dataset sizes on (A): data from years that were included in their training dataset, and (B): data from 2018, which was excluded from all training data for this experiment.

Results Time-series analysis

NDVI and land use class probability slopes

Our NDVI slope maps show which areas have an increase or decrease in NDVI over time. When reviewing the maps big features as well as details can be appreciated as seen in Fig. 4.

Fig. 4.1 and 4.2 show areas of negative and positive slope occur adjacent to each other without gradual transitions. Fig. 4.3 and 4.4 show examples of relatively large areas with homogeneous NDVI slope values. Overall, NDVI slopes in Europe tend to be positive, the largest exceptions being negative slope regions in Northern Scandinavia, Scotland, the Alps, South West France, Spain, Italy and Greece. The NDVI slope patterns in Sweden mostly match probability slopes for coniferous forests.
Figure 7. Dominant LULC classes, predicted probability and uncertainty for Non-irrigated arable land, Coniferous forest and Urban Fabric, RGB Landsat temporal composite (Spring season) for the years 2000 and 2019.
Table 12.Confusion matrix of 5-fold spatial cross-validation predictions for 33 classes
Land use change classes

We generated yearly maps for hard change classes. Filtered data as well as the removed noise can be viewed from the ODS-Europe viewer. The right-most subplots of Fig. 4 show examples of where sudden land cover change classes at 30×30 m tend to match relatively large negative slopes, especially for change classes such as deforestation and urbanization.

![Figure 8. Trends in NDVI values between 2000 and 2019 compared to trends in probabilities from machine learning and hard change classes between 2001 and 2018.](image)

We visualized the dominant type of change in a 5×5 km grid in Fig. 9. We also present the intensity of change as part of the total area on a separate map using 20×20 km areas (Fig. 9). Large parts of mainland Europe are characterized with reforestation as the main change with patches of urbanization scattered in between. Norway, Sweden and Finland are characterized with deforestation as the main land use change class. Large areas in Spain have land abandonment and crop expansion as the main land use class. When taking into account the intensity of the changes the central European countries seem to be fairly stable with the Iberian peninsula, Scandinavia and parts of eastern Europe exhibiting more intense changes.

DISCUSSION

Summary findings

We presented a framework for automated prediction of land cover / land use classes and change analysis based on spatiotemporal Ensemble Machine Learning and per-pixel trend analysis. In this framework we focused not only on predicting the most probable class, but also on mapping each probability and associated uncertainty. Such detailed information allows any future users to limit the decisions based
on uncertainty per pixel and/or incorporate it in further spatial modeling. Producing such detailed data, however, comes at a significant cost with the output dataset about 100–times larger (currently about 10 TiB of data representing land cover and land use for Europe at relatively fine spatial resolution of 30 m) than produced by similar European land cover mapping initiatives (Pflugmacher et al., 2019; Malinowski et al., 2020; Venter and Sydenham, 2021).

We further explained the time-series analysis framework for processing partial probabilities and NDVI values aiming at detection of significant spatiotemporal trends. We provide pixel-wise uncertainty measures (standard deviation of the slope / beta coefficient and R-square), which can also be used in any further spatial modeling. The whole framework, from hyper-parameter optimisation, fine-tuning, prediction and time-series analysis, is fully automated and generates consistent results over time with quantified uncertainty, making it more cost-effective for future updates and additions.

To enable easy access to the data and collaboration on similar research projects, we have made all the data, including training points and classification matrices reported in this paper, and code used to fine-tune the models and produce predictions, available via the project repository https://gitlab.com/geoharmonizer_inea; and all maps available via the Open Data Science Europe (ODS-Europe) viewer at https://maps.opendatascience.eu (Fig. 10). Most of the layers available via the ODS-Europe viewer are also available seamlessly through our S3 Cloud Object Service as Cloud-Optimized GeoTIFFs under the Open Data Commons Open Database License (ODbL) and/or Creative Commons Attribution-ShareAlike 4.0 and/or Creative Commons Attribution 4.0 International license (CC BY). This means that you can (a) visualize the data and run processing directly using QGIS or similar, (b) import,
Figure 10. Visualization of the ODSE-LULC predictions and training points used to produce time-series of land cover maps: the Open Data Science Europe viewer and its main functionality. The viewer supports visualization of data in 2D (OpenLayers) and 3D (Cesium).
subset, crop and overlay parts of the data for a local area. We do not however recommend downloading whole data sets by using Wasabi.com. To download the complete datasets (the whole of the European continent) we recommend using zenodo.org or similar.

Our accuracy assessment results indicate limited mapping accuracy (62%) at highest classification level (33 classes) with several classes such as “airports”, “burned areas” performing poorly, effectively being on the edge of usable. Noise in predictions is visible also by scrolling through time-lapse animations via the data portal (https://maps.opendatascience.eu), with some pixels being classified into transitional class sequences that are highly unlikely or physically impossible — for example a pixel changing into coniferous forest class, then to urban area, then next year again into forest class.

Our comparison of spatial versus spatiotemporal ML indicates that there is an added value in using spatiotemporal ML, especially for predicting land cover for years before, beyond or between different training point campaigns: the spatiotemporal model outperforms spatial models on known-year classification by 2.7% and unknown-year classification by 3.5% (Fig. 6). Also, we have demonstrated that spatiotemporal ML allows us to remove any bias in predictions through time (Table 8), which justifies further implementing time-series analysis on the produced time-series of predictions of hard classes, probabilities and remote sensing indices (NDVI).

Comparison with other land cover products for Europe

Validating our 2017 classification with the S2GLC test dataset resulted in similar performance values as reported by Malinowski et al. (2020), and higher than our cross-validation performance: results of the accuracy assessment using 48,365 independent test samples shows 87% match with the validation points. This indicates the nomenclature used by Malinowski et al. (2020) is possibly more suitable for remote sensing-based classification. It also shows that accuracy of the ODSE-LULC is comparable even with 10 m resolution products, and most importantly — with some improvements in accuracy — it could potentially in near future match the 85% accuracy threshold required by the CLC project.

Advantages and limitations of using spatiotemporal EML

The spatiotemporal ML described in this paper comes with multiple advantages:

• A single spacetime model can be used to model phenomena such as land cover — making it a holistic approach to data harmonization and prediction that performs consistently across multiple years;

• By including LUCAS points we base modeling and predictions on a consistent and quality-controlled data set that allows for unbiased assessment of land cover dynamics;

• In principle, we can also use the existing model to predict land cover for years 2020 and 2021 without collecting new training data, as preparing Landsat images for these periods would be likely enough;

• As all processes shown above are fully automated we can continue improving the models and re-running predictions with relative ease.
Possibly the biggest advantage of using spacetime ML is that it allows for using LUCAS and CLC data to make consistent predictions for periods prior to year 2000 for which very little training data is available. In the next phase of the project we will focus on producing predictions for years 1995, 1990 and to 1985. Currently we are not in position to make predictions for these years due to the Landsat ARD product (Potapov et al., 2020) being available only for periods >1997. We would need to compute ourselves and recalibrate the Landsat products, which added a higher level complexity also because of the differences between Landsat 5, 6, 7 and 8 sensors.

Some obvious limitations of our approach include:

- High computational intensity / still relatively high costs of preparing, producing and storing large amounts of data.
- Limited spatial resolution of 30 m, where currently most of land cover products aim at 10 m resolution.
- High dependence on quality training data spread through time. Many continents do not have even close to similar datasets as LUCAS and hence transferring this methodology to other continents (apart from USA and Australia) would likely be a cumbersome.
- Inconsistent accuracies per class result in a product with a variable quality. This can put many potential users off as consistency is often crucial for usability of geo-data.

The spatial cross-validation results show that the ensemble did not perform consistently across all classes at the highest level of thematic resolution. For instance, classes 111 and 323 had positive rates of 1.38 and 1.22. Fig. 11 shows that many classes in the same level 1 category (1: Artificial surfaces) were frequently incorrectly classified as class 111, while 211: Non-irrigated arable land — the level 3 class with the highest support was the leading false positive class for the largest number of classes. Poor accuracy of some classes and high dependence of spatiotemporal ML on quality training data, could be considered overall highest limitations of this work.

**Time-series analysis, interpretations and challenges**

The results of NDVI and probability trend analysis show some interesting patterns. We have focused on four geographic areas: (1) Sweden, as its forest dynamics have already garnered academic attention and it is an exemplary area where remote sensing techniques and on the ground measurements might come to different conclusions (see e.g. Ceccherini et al. (2020)). (2) South West France, as it is similar to the Sweden both in our data and compared and is also compared by other authors (Senf and Seidl, 2021). (3) Northern Romania because it shows a large region with positive trends for both NDVI and broad-leaved forest land cover, suggesting it is reforesting at high rates. Finally, we found large regions in the Alps (4) show a strong negative trend for NDVI values that does not seem to correspond to a unbalanced harvesting of forests or other obvious processes. This signal in our data is not yet understood and will need additional research.

Intense forest loss is a well studied process in Europe (Senf et al., 2018; Ceccherini et al., 2020; Senf and Seidl, 2021). Discrepancies between national forest inventories and remote sensing technique has led
to disagreements in Sweden (Paulsson et al., 2020), Finland (Korhonen, 2020), and Norway (Rossi et al., 2019). For instance, it was found that existing remote sensing products are deemed not fit for these types of analysis (Palahí et al., 2021). For these reasons, and because we do not validate our trend results, we neither attribute specific causes, nor do we analyze differences between specific time periods.

We do compare the most prominent change between 2001–2018 and our results suggest that forest is disappearing more than it is re-appearing in multiple locations. This is corroborated by Global Forest Watch; for example, the Jämtland region in Sweden lost 287k ha of tree cover and gained 164k ha (Hansen et al., 2013). We present the case of the Landes region in France here as well as it shows a similar pattern to large parts of Sweden and is a known area for large scale forest harvesting (Senf and Seidl, 2021). These cases exemplify the usefulness of our maps for finding similar processes all over Europe by using a combination of the maps that are presented here.

Our data suggests that NDVI increase, accompanied by reforestation, is the mayor component of change on a European scale. This is corroborated by the FAO’s State of Europe’s Forests report 2020 which states that European forest cover has increased by 9% between 1990 and 2020 (Rasi, 2020) and with global estimates that forest cover has increased by 7% between 1982 and 2016 (Song et al., 2018). This increase is consistent with expectations that increased CO2 will enhance plant growth in general but increase in NDVI could also be explained by other factors such as increase in cropland cover (Huang et al., 2018), having a higher NDVI value than other land use types, nitrogen depositions. Another concern that is raised is that most of the increase in forest gain is by planted forests (Payn et al., 2015) that are less valuable in terms of biodiversity and carbon sequestration (Liu et al., 2018) and less adaptable to climate change. One exemplary area for reforestation is found in Northern Romania for all three parts of our time-series analysis: Both NDVI and broad-leaved forest probability slopes show positive trends in a large area marked with reforestation as the dominant change class.

Finally, our data for the Alps shows unexpected negative NDVI trends for large parts of the Alps. This may be related to changes in snow cover as found by Wang et al. (2018) in the Tibetan Plateau and by Buus-Hinkler et al. (2006) in the Arctic regions. However, this is not corroborated by the probability slope for class “Glaciers and perpetual snow” in our data. It is also possible that this is an artifact from our gap-filling step; further study is necessary before any conclusions can be drawn.

Future work

Even though our framework is comprehensive and has produced predictions of comparable accuracy to the current state-of-the-art, after almost 14 months of processing the data and modeling land cover, we have found that that many aspects of our system could be improved:

- **Combining Landsat and Sentinel data:** We have not yet tested, even though we are aware of the Harmonized Sentinel-Landsat product, using a combination of Sentinel 1/2 and Landsat imagery for land cover mapping. This has proven to be rather complex, especially considering that our interest is in mapping land cover for large time-spans — if possible going back to year 1984 — while Sentinel data is only available for years >2016.

- **Cross-validation of land cover trends:** We have not yet put enough effort to independently identify
and quantify sudden changes in land cover due to natural hazards (fires, floods etc) but also due to slow inherent processes such as urbanisation, vegetation succession and similar. How well does our trend analysis matches the processes on the ground? This was beyond the scope of our project and we find it most important that the data is now available to any research groups interested in testing their usability for land monitoring projects.

- **Combining classification with Object-Based Image Analysis (OBIA) and pattern recognition:**

  Incorporating spatial context to our workflow could potentially improve performance for several classes that are defined by land use. For instance, class 124: “Airports” was frequently misclassified as either urban fabric, non-irrigated arable land, or pastures. These predictions likely matched the land cover of the pixel, but missed the spatial patterns that make airports easily recognizable by humans (elongated landing paths). The same issue applies to most other artificial surface LULC classes. It is also clear that the water-related classes, which are mostly distinguishable by location and spatial patterns, were often incorrectly classified as inland water bodies and maritime wetlands, which were simply more numerous in the training data.

![Figure 11](image)

**Figure 11.** Percentage of predicted classes per expected class during spatial 5-fold cross-validation

One interesting result we noticed in this work is that training spatiotemporal models on LUCAS points...
lead to lower classification accuracy estimates than when only using CLC points (see Table 11). This was unexpected, as LUCAS land cover information stems from actual ground observations, while the CLC points are pseudo-ground truth points from a dataset with a large minimum mapping unit. This suggests that either the LUCAS points are harder to reproduce with remote sensing techniques, or that the harmonization process needs to be improved. Further testing is needed to clarify this.

The field of land cover mapping is rapidly evolving. With exciting new global 10 m resolution products such as ESA WorldCover and Google’s Dynamic World Map expected in 2021, we expect the LULC mapping bar to be raised quickly to higher resolution and higher accuracy. Venter and Sydenham (2021) used low-cost infrastructure to produce land cover map of Europe at 10 m — thanks to ESA and NASA making the majority of multispectral products publicly available, today everyone could potentially map the world’s land cover from their laptop. Our primary interest, however, will remain focusing on producing better land cover maps of the past, and building land cover products that allow for unbiased estimate of long-term trends. We hope that this type of data would help eventually better understand the key drivers of land degradation and restoration, so that we can help stake-holders on the ground make better decisions and hopefully receive financial support for the ecosystem services our environment provides to us all.

ACKNOWLEDGEMENTS

OpenDataScience.eu is an Open Data project drawing inspiration from the OpenLandMap.org and OpenStreetMap.org projects. This work has received funding from the European Union’s the Innovation and Networks Executive Agency (INEA) under Grant Agreement Connecting Europe Facility (CEF) Telecom project 2018-EU-IA-0095. The authors are grateful to Radek Malinovski (Centrum Badań Kosmicznych Polskiej Akademii Nauk — CBK PAN) for providing access to independent validation points, and Martin Herold and Sytze de Bruin (Wageningen University) for providing suggestions on methodology.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Batista e Silva, F., Lavalle, C., and Koomen, E. (2013). A procedure to obtain a refined european land use/cover map. Journal of Land Use Science, 8(3):255–283.

Bossard, M., Feranec, J., and Ořáhel, J. (2000). Corine land cover—technical guide.

Breiman, L. (2001). Random forests. Machine learning, 45(1):5–32.

Buchhorn, M., Lesiv, M., Tsendbazar, N.-E., Herold, M., Bertels, L., and Smets, B. (2020). Copernicus global land cover layers—collection 2. Remote Sensing, 12(6):1044.

Buck, O., Haub, C., Woditsch, S., Lindemann, M., Kleinwillinghöfer, L., Hazeu, G., Kosztra, B., Kleeschulte, S., Arnold, S., and Hölzli”, M. (2015). Analysis of the LUCAS nomenclature and proposal for adaptation of the nomenclature in view of its use by the Copernicus land monitoring services.
Buus-Hinkler, J., Hansen, B. U., Tamstorf, M. P., and Pedersen, S. B. (2006). Snow-vegetation relations in a high arctic ecosystem: Inter-annual variability inferred from new monitoring and modeling concepts. *Remote Sensing of Environment, 105*(3):237–247.

Buyantuyev, A. and Wu, J. (2007). Effects of thematic resolution on landscape pattern analysis. *Landscape Ecology, 22*(1):7–13.

Calderón-Loor, M., Hadjikakou, M., and Bryan, B. A. (2021). High-resolution wall-to-wall land-cover mapping and land change assessment for Australia from 1985 to 2015. *Remote Sensing of Environment, 252*:112148.

Carmel, Y., Dean, D. J., and Flather, C. H. (2001). Combining location and classification error sources for estimating multi-temporal database accuracy. *Photogrammetric Engineering and Remote Sensing, 67*(7):865–872.

Castilla, G., Larkin, K., Linke, J., and Hay, G. J. (2009). The impact of thematic resolution on the patch-mosaic model of natural landscapes. *Landscape Ecology, 24*(1):15–23.

Ceccherini, G., Duveiller, G., Grassi, G., Lemoine, G., Avitabile, V., Pilli, R., and Cescatti, A. (2020). Abrupt increase in harvested forest area over Europe after 2015. *Nature, 583*(7814):72–77.

Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., He, C., Han, G., Peng, S., Lu, M., et al. (2015). Global land cover mapping at 30 m resolution: A pok-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing, 103*:7–27.

Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., and Terpenning, I. (1990). Stl: A seasonal-trend decomposition. *Journal of official statistics, 6*(1):3–73.

Conway, T. (2009). The impact of class resolution in land use change models. *Computers, Environment and Urban Systems, 33*(4):269–277.

d’Andrimont, R., Yordanov, M., Martínez-Sanchez, L., Eiselt, B., Palmieri, A., Dominici, P., Gallego, J., Reuter, H. I., Jobebs, 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.

Defazio, A., Bach, F., and Lacoste-Julien, S. (2014). Saga: A fast incremental gradient method with support for non-strongly convex composite objectives. *arXiv preprint arXiv:1407.0202*.

Duveiller, G., Caporaso, L., Abad-Viñas, R., Perugini, L., Grassi, G., Arneth, A., and Cescatti, A. (2020). Local biophysical effects of land use and land cover change: towards an assessment tool for policy makers. *Land Use Policy, 91*:104382.

Feng, M. and Bai, Y. (2019). A global land cover map produced through integrating multi-source datasets. *Big Earth Data, 3*(3):191–219.

Feranec, J., Soukup, T., Hazeu, G., and Jaffrain, G. (2016). European landscape dynamics: *CORINE land cover data*. CRC Press.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., et al. (2005). Global consequences of land use. *Science, 309*(5734):570–574.
Gao, B.-C. (1996). Ndwi—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote sensing of environment*, 58(3):257–266.

Gao, Y., Liu, L., Zhang, X., Chen, X., Mi, J., and Xie, S. (2020). Consistency Analysis and Accuracy Assessment of Three Global 30-m Land-Cover Products over the European Union using the LUCAS Dataset. *Remote Sensing*, 12(21):3479.

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O’Reilly Media.

Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. A., Tyukavina, A., Thau, D., Stehman, S., Goetz, S. J., Loveland, T. R., et al. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, 342(6160):850–853.

Hengl, T., Leal Parente, L., Križan, J., and Bonannella, C. (2021). *Continental Europe Digital Terrain Model at 30 m resolution based on GEDI, ICESat-2, AW3D, GLO-30, EUDEM, MERIT DEM and background layers*. Zenodo.

Herold, M., Mayaux, P., Woodcock, C., Baccini, A., and 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.

Hillger, D., Kopp, T., Lee, T., Lindsey, D., Seaman, C., Miller, S., Solbrig, J., Kidder, S., Bachmeier, S., Jasmin, T., et al. (2013). First-light imagery from suomi npp viirs. *Bulletin of the American Meteorological Society*, 94(7):1019–1029.

Homer, C., Dewitz, J., Fry, J., Coan, M., Hossain, N., Larson, C., Herold, N., McKerrow, A., VanDriel, J. N., Wickham, J., et al. (2007). Completion of the 2001 national land cover database for the conterminous united states. *Photogrammetric engineering and remote sensing*, 73(4):337.

Homer, C., Dewitz, J., Jin, S., Xian, G., Costello, C., Danielson, P., Gass, L., Funk, M., Wickham, J., Stehman, S., et al. (2020). Conterminous united states land cover change patterns 2001–2016 from the 2016 national land cover database. *ISPRS Journal of Photogrammetry and Remote Sensing*, 162:184–199.

Hong, C., Burney, J. A., Pongratz, J., Nabel, J. E., Mueller, N. D., Jackson, R. B., and Davis, S. J. (2021). Global and regional drivers of land-use emissions in 1961–2017. *Nature*, 589(7843):554–561.

Houghton, R. A., House, J. I., Pongratz, J., Van Der Werf, G. R., DeFries, R. S., Hansen, M. C., Quéré, C. L., and Ramankutty, N. (2012). Carbon emissions from land use and land-cover change. *Biogeosciences*, 9(12):5125–5142.

Huang, K., Xia, J., Wang, Y., Ahlström, A., Chen, J., Cook, R. B., Cui, E., Fang, Y., Fisher, J. B., Huntzinger, D. N., et al. (2018). Enhanced peak growth of global vegetation and its key mechanisms. *Nature ecology & evolution*, 2(12):1897–1905.

Huete, A. R. (1988). A soil-adjusted vegetation index (savi). *Remote sensing of environment*, 25(3):295–309.

Inglada, J., Vincent, A., Arias, M., Tardy, B., Morin, D., and Rodes, I. (2017). Operational high resolution land cover map production at the country scale using satellite image time series. *Remote Sensing*, 9(1):95.

Jin, S. and Sader, S. A. (2005). Comparison of time series tasseled cap wetness and the normalized
difference moisture index in detecting forest disturbances. Remote sensing of Environment, 94(3):364–372.

Kaplan, J. O., Krumhardt, K. M., Ellis, E. C., Ruddiman, W. F., Lemmen, C., and Goldewijk, K. K. (2011). Holocene carbon emissions as a result of anthropogenic land cover change. The Holocene, 21(5):775–791.

Key, C. H. and Benson, N. C. (1999). Measuring and remote sensing of burn severity. In Proceedings joint fire science conference and workshop, volume 2, page 284. University of Idaho and International Association of Wildland Fire Moscow, ID.

Key, C. H. and Benson, N. C. (2006). Landscape assessment (la). In: Lutes, Duncan C.; Keane, Robert E.; Caratti, John F.; Key, Carl H.; Benson, Nathan C.; Sutherland, Steve; Gangi, Larry J. 2006. FIREMON: Fire effects monitoring and inventory system. Gen. Tech. Rep. RMRS-GTR-164-CD. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Research Station. p. LA-1-55, 164.

Kilibarda, M., Hengl, T., Heuvelink, G. B., Gräler, B., Pebesma, E., Perćec Tadić, M., and Bajat, B. (2014). Spatio-temporal interpolation of daily temperatures for global land areas at 1 km resolution. Journal of Geophysical Research: Atmospheres, 119(5):2294–2313.

Korhonen, K. T. (2020). A new article in the journal Nature overestimates the increase of forest harvesting in Europe. Natural Resources Institute Finland (Luke). Accessed: 2021-07-04.

Li, X., Zhou, Y., Zhu, Z., Liang, L., Yu, B., and Cao, W. (2018). Mapping annual urban dynamics (1985–2015) using time series of landsat data. Remote Sensing of Environment, 216:674–683.

Liu, H., Gong, P., Wang, J., Clinton, N., Bai, Y., and Liang, S. (2020a). Annual dynamics of global land cover and its long-term changes from 1982 to 2015. Earth System Science Data, 12(2):1217–1243.

Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J. (2021). Finer-resolution mapping of global land cover: Recent developments, consistency analysis, and prospects. Journal of Remote Sensing, 2021.

Liu, X., Trogisch, S., He, J.-S., Niklaus, P. A., Brueelheide, H., Tang, Z., Erfmeier, A., Scherer-Lorenzen, M., Pietsch, K. A., Yang, B., et al. (2018). Tree species richness increases ecosystem carbon storage in subtropical forests. Proceedings of the Royal Society B, 285(1885):20181240.

Liu, Y., Hou, X., Li, X., Song, B., and Wang, C. (2020b). Assessing and predicting changes in ecosystem service values based on land use/cover change in the bohai rim coastal zone. Ecological Indicators, 111:106004.

Lovelace, R., Nowosad, J., and Muenchow, J. (2019). Geocomputation with R. Chapman & Hall/CRC The R Series. CRC Press.

Malinowski, R., Lewiński, S., Rybicki, M., Gromny, E., Jenerowicz, M., Krupiński, M., Nowakowski, A., Wojtkowski, C., Krupiński, M., Krätzschmar, E., et al. (2020). Automated Production of a Land Cover/Use Map of Europe Based on Sentinel-2 Imagery. Remote Sensing, 12(21):3523.

McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115–133.

Ololfsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., and Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. Remote Sensing of Environment, 148:42–57.
Palahí, M., Valbuena, R., Senf, C., Acil, N., Pugh, T. A., Sadler, J., Seidl, R., Potapov, P., Gardiner, B., Hetemäki, L., et al. (2021). Concerns about reported harvests in european forests. *Nature*, 592(7856):E15–E17.

Paulsson, J., Claesson, S., Fridman, J., and Olsson, H. (2020). Incorrect figures on harvested forests in nature article. *SLU news*. Accessed: 2021-07-04.

Payn, T., Carnus, J.-M., Freer-Smith, P., Kimberley, M., Kollert, W., Liu, S., Orazio, C., Rodriguez, L., Silva, L. N., and Wingfield, M. J. (2015). Changes in planted forests and future global implications. *Forest Ecology and Management*, 352:57–67.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540:418–.

Pflugmacher, D., Rabe, A., Peters, M., and 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.

Pielke Sr, R. A., Marland, G., Betts, R. A., Chase, T. N., Eastman, J. L., Niles, J. O., Niyogi, D. D. S., and Running, S. W. (2002). The influence of land-use change and landscape dynamics on the climate system: relevance to climate-change policy beyond the radiative effect of greenhouse gases. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 360(1797):1705–1719.

Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., and Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote sensing of environment*, 48(2):119–126.

Raši, R. (2020). *State of Europe’s Forests 2020*. Ministerial conference on the protection of forests in Europe. Accessed: 2021-10-05.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929.

Román, M. O., Wang, Z., Sun, Q., Kalb, V., Miller, S. D., Molthan, A., Schultz, L., Bell, J., Stokes, E. C., Pandey, B., et al. (2018). Nasa’s black marble nighttime lights product suite. *Remote Sensing of Environment*, 210:113–143.

Rossi, F., Breidenbach, J., Puliti, S., Astrup, R., and Talbot, B. (2019). Assessing harvested sites in a forested boreal mountain catchment through global forest watch. *Remote Sensing*, 11(5).

Sala, O. E., Chapin, F. S., Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., et al. (2000). Global biodiversity scenarios for the year 2100. *Science*, 287(5459):1770–1774.

Seabold, S. and Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. In
9th Python in Science Conference.

Senf, C., Pflugmacher, D., Zhiqiang, Y., Sebald, J., Knorn, J., Neumann, M., Hostert, P., and Seidl, R. (2018). Canopy mortality has doubled in Europe’s temperate forests over the last three decades. *Nature Communications*, 9(1):1–8.

Senf, C. and Seidl, R. (2021). Mapping the forest disturbance regimes of Europe. *Nature Sustainability*, 4(1):63–70.

Seni, G. and Elder, J. (2010). *Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions*. Synthesis lectures on data mining and knowledge discovery. Morgan & Claypool Publishers.

Shumba, T., De Vos, A., Biggs, R., Esler, K. J., Ament, J. M., and Clements, H. S. (2020). Effectiveness of private land conservation areas in maintaining natural land cover and biodiversity intactness. *Global Ecology and Conservation*, 22:e00935.

Song, X.-P., Hanssen, M. C., Stehman, S. V., Potapov, P. V., Tyukavina, A., Vermote, E. F., and Townshend, J. R. (2018). Global land change from 1982 to 2016. *Nature*, 560(7720):639–643.

Song, X.-P., Sexton, J. O., Huang, C., Channan, S., and Townshend, J. R. (2016). Characterizing the magnitude, timing and duration of urban growth from time series of Landsat-based estimates of impervious cover. *Remote Sensing of Environment*, 175:1–13.

Sy, S. and Quesada, B. (2020). Anthropogenic land cover change impact on climate extremes during the 21st century. *Environmental Research Letters*, 15(3):034002.

Szantoi, Z., Geller, G. N., Tsendbazar, N.-E., See, L., Griffiths, P., Fritz, S., Gong, P., Herold, M., Mora, B., and Obregón, A. (2020). Addressing the need for improved land cover map products for policy support. *Environmental Science & Policy*, 112:28–35.

Townshend, J. R., Masek, J. G., Huang, C., Vermote, E. F., Gao, F., Channan, S., Sexton, J. O., Feng, M., Narasimhan, R., Kim, D., et al. (2012). Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. *International Journal of Digital Earth*, 5(5):373–397.

Trisurat, Y., Shirakawa, H., and Johnston, J. M. (2019). Land-use/land-cover change from socio-economic drivers and their impact on biodiversity in Nan province, Thailand. *Sustainability*, 11(3):649.

Venter, Z. S. and Sydenham, M. A. K. (2021). Continental-scale land cover mapping at 10 m resolution over Europe (elc10).

Vilar, L., Garrido, J., Echavarría, P., Martínez-Vega, J., and 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. *International Journal of Applied Earth Observation and Geoinformation*, 78:102–117.

Wang, X., Wu, C., Peng, D., Gonsamo, A., and Liu, Z. (2018). Snow cover phenology affects alpine vegetation growth dynamics on the tibetan plateau: Satellite observed evidence, impacts of different biomes, and climate drivers. *Agricultural and Forest Meteorology*, 256:61–74.

Zhang, C. and Ma, Y. (2012). *Ensemble Machine Learning: Methods and Applications*. Springer New York.

Zhang, X., Liu, L., Chen, X., Gao, Y., Xie, S., and Mi, J. (2020). GLC_FCS30: Global land-cover product with fine classification system at 30 m using time-series Landsat imagery. *Earth System Science Data Discussions*, pages 1–31.

Zhou, W., Qian, Y., Li, X., Li, W., and Han, L. (2014). Relationships between land cover and the surface urban heat island: seasonal variability and effects of spatial and thematic resolution of land cover data on predicting land surface temperatures. *Landscape ecology*, 29(1):153–167.

Zhu, Z. and Woodcock, C. E. (2014). Continuous change detection and classification of land cover using all available landsat data. *Remote sensing of Environment*, 144:152–171.