https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

TimeSpec4LULC: A Global Deep Learning-driven Dataset of MODIS Terra-Aqua Multi-Spectral Time Series for LULC Mapping and Change Detection

Rohaifa Khaldi[1], Domingo Alcaraz-Segura[1,3], Emilio Guirado[4], Yassir Benhammou[5], Abdellatif El Afia[2], Francisco Herrera[5], and Siham Tabik[5]

[1]Dept. of Botany, Faculty of Science, University of Granada, 18071 Granada, Spain
[2]ENSIAS, Mohammed V University, Rabat, 10170, Morocco
[3]iEcolab, Inter-University Institute for Earth System Research, University of Granada, 18006 Granada, Spain
[4]Multidisciplinary Institute for Environment Studies "Ramón Margalef", University of Alicante, 03690, Spain
[5]Dept. of Computer Science and Artificial Intelligence, Andalusian Research Institute in Data Science and Computational Intelligence, DaSCI, University of Granada, 18071, Granada, Spain

Correspondence: Rohaifa Khaldi (rohaifa@ugr.es); Domingo Alcaraz-Segura (dalcaraz@ugr.es)

Abstract. Land Use and Land Cover (LULCs) mapping and change detection are of paramount importance to understand the distribution and effectively monitor the dynamics of the Earth's system. An unexplored way to create global LULC maps is by building good quality LULC-models based on state-of-the-art deep learning networks. Building such models requires large global good quality time series LULC datasets, which are not available yet. This paper presents TimeSpec4LULC (Khaldi et al., 2021), a smart open-source global dataset of multi-Spectral Time series for 29 LULC classes. TimeSpec4LULC was built based on the 7 spectral bands of MODIS sensor at 500 m resolution from 2002 to 2021, and was annotated using a spatial agreement across the 15 global LULC products available in Google Earth Engine. The 19-year monthly time series of the seven bands were created globally by: (1) applying different spatio-temporal quality assessment filters on MODIS Terra and Aqua satellites, (2) aggregating their original 8-day temporal granularity into monthly composites, (3) merging their data into a Terra+Aqua combined time series, and (4) extracting, at the pixel level, 11.85 million time series for the 7 bands along with a set of metadata about geographic coordinates, country and departmental divisions, spatio-temporal consistency across LULC products, temporal data availability, and the global human modification index. To assess the annotation quality of the dataset, a sample of 100 pixels, evenly distributed around the world, from each LULC class, was selected and validated by experts using very high resolution images from both Google Earth and Bing Maps imagery. This smartly, pre-processed, and annotated dataset is targeted towards scientific users interested in developing and evaluating various machine learning models, including deep learning networks, to perform global LULC mapping and change detection.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

1 Introduction

Accurate Land-Use and Land-Cover (LULC) information, including distribution, dynamics and changes, is of paramount importance for understanding and modelling the natural and human-modified behavior of the Earth's system (Tuanmu and Jetz, 2014; Verburg et al., 2009). Land cover is an essential variable that provides powerful insights for the assessment and modelling of terrestrial ecosystem processes, biogeochemical cycles, biodiversity, climate, and water resources, etc. (Luoto et al., 2007; Menke et al., 2009; Polykretis et al., 2020). Land uses incorporate many types of modifications that an increasing human population, 9 billion expected by 2050, causes to the Earth surface. LULCs are subjected to anomalies, trends and changes both from anthropogenic and natural origins (Polykretis et al., 2020), with subsequent modifications on their biophysical properties. LULC change is usually interpreted as the conversion from one LULC category to another or/and the modification of land management within LULC (Meyer et al., 1994). The assessment of LULC and LULC change has been identified as an essential planetary boundary to assess the status and trends of social-ecological systems from the local to the global scale in the pursuit of a safe operating space for humanity. For instance, urban sprawl and agriculture expansion or abandonment affect the biodiversity, soil quality, climate, food security, and human health (Lambin and Geist, 2008; Feddema et al., 2005). For this reason, continuous and accurate LULC and LULC change mapping is essential in policy and research to monitor ecological and environmental change at different temporal and spatial scales (Polykretis et al., 2020; García-Mora et al., 2012), and as a decision support system to ensure an effective and sustainable planning and management of natural resources (Kong et al., 2016; Congalton et al., 2014; Grekousis et al., 2015).

Satellite remote sensing in combination with geographic information systems (GIS) have provided convenient, inexpensive, and continuous spatio-temporal information for mapping LULCs and detecting changes on the Earth's surface from regional to global scales (Kong et al., 2016; Kerr and Ostrovsky, 2003; Pfeifer et al., 2012) thanks to their strong ability to cover, timely and repeatedly, wide and inaccessible areas, and to get high spatial and temporal resolution data (Alexakis et al., 2014; Yirsaw et al., 2017; Patel et al., 2019). Since the 1980s, multiple global LULC products (Table 1) have been derived from remotely sensed data, providing alternative characterizations of the Earth surface at varying extents of spatial and temporal resolutions (Townshend et al., 1991; Loveland et al., 2000; Bartholome and Belward, 2005).

One of the most important limitations of global LULC products is the within-product variability of accuracy (across different years, regions, and LULC types), besides the generally low agreement among products in many regions of the world (Tsendbazar et al., 2015b, 2016; Gao et al., 2020; Gong et al., 2013; Zimmer-Gembeck and Helfand, 2008). Such lack of consensus can translate into huge implications for subsequent global assessments of biodiversity status, carbon balance, or climate change (Estes et al., 2018; de la Cruz et al., 2017). In addition, accuracy at the local level can be too low, which impedes the use of global or regional LULC products in local studies (Hoskins et al., 2016; Tsendbazar et al., 2016), since it can lead to different conclusions due to the compelling amount of inconsistencies, uncertainties, and inaccuracies (Tsendbazar et al., 2015a; Estes et al., 2018). Multiple methodological reasons lay behind these discrepancies among LULC products (Congalton et al., 2014; Grekousis et al., 2015; Gómez et al., 2016):

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 1. Description of the GEE Global LULC products used in this study.

Product ID(s)	Product	Version	Provider	Sensor	Satellite or Spaceborne	Spatial resolution	Acquisition time	Data type
P1:P5	MCD12Q1 (Type 1 to 5)	v6	NASA LP DAAC at the USGS EROS Center	MODIS	Aqua, Terra	500 m	2001-2019	Image collection over years
P6	CGLS-LULC100	v3.0.1	Copernicus Global Land Service (CGLS)	-	PROBA-V	100 m	2015-2019	Image collection over years
P7	GFCC	v3	NASA LP DAAC at the USGS EROS Center	Multi-sensor	Multi-satellite	30 m	2000,2005, 2010,2015	Image collection over years
P8	GLOBCOVER	v2	ESA and by the Catholic University of Louvain	MERIS	ENVISAT	300 m	2009	Single Image
P9	GFSAD	v0.1	Global Food Security support Analysis Data at 30m Project (GFSAD30)	Multi-sensor	Multi-satellite	1000 m	2010	Single Image
P10	PALSAR2	vfnf	JAXA EORC	SAR	ALOS, ALOS 2	25 m	2007-2010 2015-2017	Single Image
P11	HANSEN	v1.7	Hansen, UMD, Google, USGS, NASA	OLI	Landsat 8	1 arc seconds	2000-2019	Single Image
P12	GFCH	v2005	NASA, JPL	-	Lidar	30 arc seconds	2005	Single Image
P13	JRC Yearly History (YH)	v1.2	EC JRC, Google	Multi-sensor	Landsat (5,7,8)	30 m	1984-2019	Single Image
P14	JRC Global Surface Water (GSW)	v1.2	EC JRC, Google	Multi-sensor	Landsat (5,7,8)	30 m	1984-2019	Image collection over years
P15	Tsinghua FROM-GLC	v10	Tsinghua University	Multi-sensor	Landsat	30 m	1985-2018	Single Image

- Satellite sensors: the spatial, temporal and spectral resolutions of the source satellite images strongly determine the precision and accuracy of derived LULCs. Native pixel size can vary from dozens of meters to kilometers, which determines the precision. Revisiting frequency can vary from daily images to several weeks, which determines the possibility of removing cloud and atmospheric noise effects. In addition, the greater the number of spectral bands in a sensor, the greater the amount of complementary information that can help to differentiate among LULC classes.
- Processing techniques: the different algorithms for atmospheric correction, cloud filtering, image composition, viewing geometry corrections, etc. can also influence LULC accuracy.
- Acquisition year(s): some LULC products just refer to a particular year while others are regularly updated.
- Classification schemes: LULC legends can greatly differ in the number of classes and typology definitions. In general, LULC products tend to agree more in broader general categories than in finer specific ones.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021


Open Access Earth System Science Data Discussions

- Classification algorithms: the approaches and rules used to identify each LULC have evolved from decision trees, to multivariate clustering, and machine learning, including now deep learning.
- Validation techniques of the final product. The amount and global distribution of ground truth samples differs across products and influences their reported accuracy.

Many efforts have been made to assess, compare, and harmonize the increasing plethora of global, regional and local LULC products, including their integration into synthetic products, which has shed light onto their strengths and weaknesses (Feng and Bai, 2019; Zhang et al., 2019; Gao et al., 2020; Liu et al., 2021). Still, the myriad of existing products with different specifications and accuracies have made their selection by the users problematic, and discouraging because it is frequently unknown whether a product meets the user needs for a particular area or LULC class (Tsendbazar et al., 2015b; Xu et al., 2020). In addition, many of these efforts are either limited to regional or national scale (e.g. Pérez-Hoyos et al. (2012); Gengler and Bogaert (2018)), coarse spatial resolution (e.g. Tuanmu and Jetz (2014); Jung et al. (2006)), or just one LULC type (e.g. Fritz et al. (2011)). The use of synergistic products takes advantage from the strengths of individual products while attenuating their respective weaknesses. However, they still face the challenge of taking into consideration the heterogeneity within pixels (spatial purity), i.e., they only provide a single discrete class per pixel (Tuanmu and Jetz, 2014), and of considering land cover change over time (temporal purity), i.e., they do not integrate the inter-annual variability while combining the related products. In general, given a target maximum error of 5–15% either per class or for the overall accuracy, most of the current global land-cover maps still do not meet the accuracy demands of many applications (Liu et al., 2021).

In parallel to these efforts, deep artificial neural networks, also known as Deep Learning (DL) models, have shown impressive performance in computer vision and promising ones in remote sensing during the last decade. Currently, two specific types of DL models, i.e., CNNs (Convolutional Neural Networks) and RNNs (Recurrent Neural Networks), constitute the state-of-art in respectively extracting spatial and temporal patterns from data-records. Indeed, DL models are showing great performance in LULC tasks such as scene classification (Zhang et al., 2018a), object detection (Zhao et al., 2015; Guirado et al., 2021) and segmentation (Zhao and Du, 2016; Guirado et al., 2017; Safonova et al., 2021) in RGB and multi-spectral satellite and aerial images. However, such good performance is only possible when DL models are trained on smart data. The concept of smart data involves all pre-processing methods that improve value and veracity of data and of associated expert annotations (Luengo et al., 2020), resulting in high quality and accurately annotated datasets. In general, remote sensing datasets contain noise, missing values, and high variability and complexity across space, time and spectral bands. Applying pre-processing methods, such as gap filling and noise reduction to data and consensus across multiple sources to annotations, contributes to creating smart remote sensing datasets.

Currently, there only exist few multi-spectral datasets annotated for training DL models to map LULC and LULC change (Table 2). However, most of these datasets provide very short time series of data, very few LULC classes, and do not have a global coverage. As far as we know, there is no dataset designed for DL models that allows for global scale analysis of many LULC classes and that are long enough to characterize change over time, such as phenology, inter-annual dynamics, trends, anomalies, and abrupt changes.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 2. A list of existing times series dataset, including the proposed TimeSpec4LULC dataset, for training machine learning models.

Dataset	Source	# images × (pixels)	Spatial Resolution (m)	Temporal Resolution	No. Bands	No. Classes	Extent	Intra/Inter time series	labeled for
CaneSat (Virnodkar et al., 2020)	Sentinel-2	1627× (10×10)	10	Monthly	6	2	India	[2018, 2019]	Sugarcane classification
SpaceNet-7 (Van Etten et al., 2021)	Dove Satellite Constellation Planet Labs'	24× (1024×1024)	4	Monthly	8	2	100 cities	[2017,2020]	Buildings tracking
Time series spectral dataset for croplands in France (Hubert-Moy et al., 2019)	MODIS and LPIS	21,129 pixels	250	8-day intervals	4	19	France	[2006, 2017]	Crop type mapping
TiSeLaC[1]	Landsat	8 × (2866× 2633)	30	Annually	10	9	The Reunion Island	2014	classification
BreizhCrops Rußwurm et al. (2019)	Sentinel-2	610,000 pixels	60	-	10	9	Brittany dept. France	[01/01/2017, 31/12/2017]	Crop type mapping
TimeSpec4LULC (Ours)	MODIS	11,856,992 pixels	500	Monthly	7	29	Global	[07/2002, 01/2021]	LULC mapping and change detection

This paper presents, TimeSpec4LULC, a new open-source, smart and global dataset of multi-spectral time series targeted towards the development and evaluation of DL models to map LULCs and monitor their change globally. TimeSpec4LULC was built using Google Earth Engine (GEE) by combining the seven 500m spectral bands of MODIS Aqua and MODIS Terra satellite sensors at a monthly time step from 2002 to 2021. It contains 11.85 million of pixels that were annotated based on a consensus across up to 15 global LULC products (Table 1) for 29 broad and globally harmonized LULC classes. In addition, it provides this metadata at pixel level: geographic coordinates, country and departmental divisions, spatio temporal consistency across LULC products, statistics on temporal data availability, and the global human modification index. The annotation quality was further assessed by experts using Google Satellite and Bing Maps very high resolution images in 100 samples per class evenly distributed around the world.

2 Methods

To build TimeSpec4LULC, we first determined the spatial and temporal agreement across 15 heterogeneous global LULC products (listed in Table 1) for 29 broad and globally harmonized LULC classes. Then, for each class, we extracted a 19-year monthly time series for the seven 500-meter spectral bands of MODIS Terra and Aqua combined. We carried out this process in GEE (Gorelick et al., 2017) since it provides access to freely available satellite imagery under a unified programming, processing and visualization environment.

2.1 Finding spatial agreement across 15 global LULC products

To find the spatial consensus across global LULC products for different LULC classes, we followed five steps: 1) selection of global LULC products, 2) standardization and harmonization of LULC legends, 3) combination of products across space and

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

time, and 4) reprojection and selection of spatial purity thresholds to get a final consistent mask across the 15 products for each one of the 29 LULC classes.

2.1.1 Selection of global LULC products

We used the 15 most updated global LULC products available in GEE (Table 1). These products widely differ in their source satellite data, spatial resolution, temporal coverage, class legend, and accuracy. Given such heterogeneity, we used the consensus across all of them in space and time as a source of reliability to support our annotation. That is, a given LULC class is assigned to a 500 m pixel only if it was consistent over time and space across all the 15 LULC products.

2.1.2 Standardization and Harmonization of LULC legends

Broadly, land cover (LC) refers to the different vegetation types (usually following biotype, plant functional type, or physiognomy schemes, such as forests, shrublands, or grasslands) or other biophysical classes (such as water bodies, snow, or bare soil) that cover the Earth's surface. Whereas, land use (LU) describes the set of human activities that significantly modify a LC (such as urban areas and croplands).

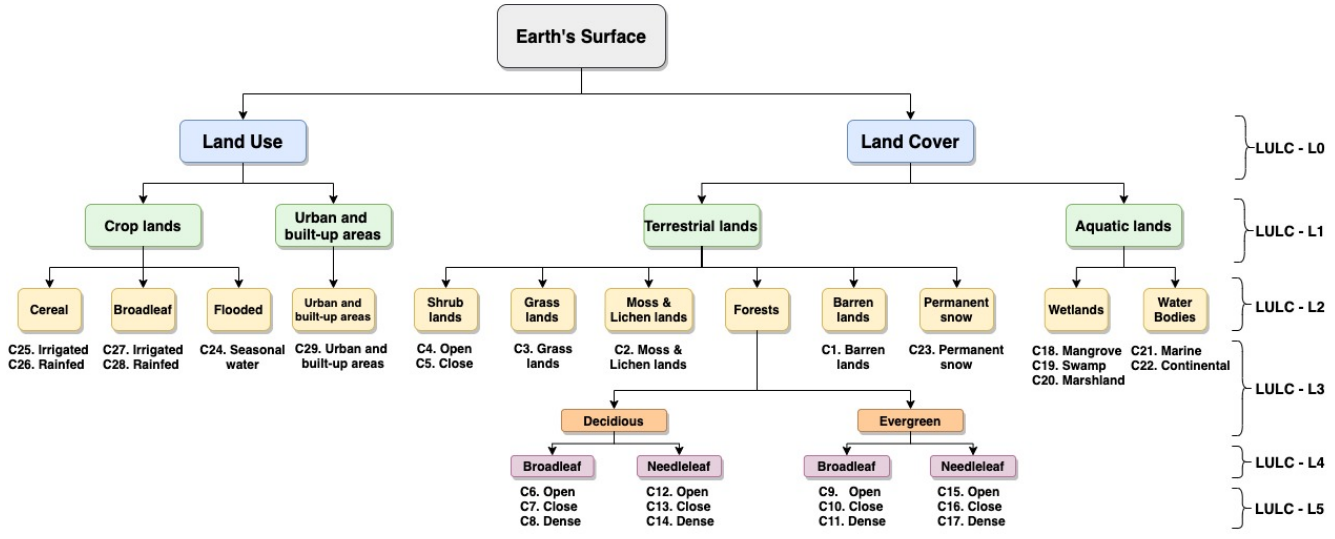


Figure 1. The hierarchical structure of the LULC classes contained in the TimeSpec4LULC dataset. C1 to C29: class 1 to class 29. LULC-L0 to LULC-L5: LULC level 0 to LULC level 5. The level LULC-L0 includes the 2 blue boxes. The level LULC-L1 includes the 4 green boxes. The level LULC-L2 includes the 12 yellow boxes. The level LULC-L3 includes all the classes of the 12 yellow boxes (from C1 to C5 and from C18 to C29) except the Forests class where it includes only the 2 orange boxes (Decidious and Evergreen). The level LULC-L4 includes the same classes but expands the Forests class into the 4 purple boxes: Decidious (Broadleaf and Needleleaf) and Evergreen (Broadleaf and Needleleaf). The last level LULC-L5 includes all the 29 LULC classes (from C1 to C29).

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

To standardize and harmonize the LULC legends across the 15 LULC products, we used expert knowledge (Vancutsem et al., 2013) to create a common nomenclature of 29 broad classes (6 LU classes and 23 LC classes) that were interoperable across all products (see the hierarchical structure of our legend in Fig. 1 and the rule set across products in Table 3). The LULC legend was structured into 6 hierarchical levels (L0 to L5).

The six anthropogenic LU classes contained Urban and built-up areas and five types of croplands. The 23 natural or semi-natural LC classes covered three aquatic systems (Marine water bodies, Continental water bodies, and Wetlands) and 18 terrestrial systems (Permanent snow, Barren lands, Moss and Lichen lands, Grasslands, Close Shrublands, Open Shrublands, and 12 types of Forests that differed in their canopy type, phenology and tree cover).

Table 3. The rule set used to build the LULC legend for TimeSpec4LULC dataset. P1 to P15: product 1 to 15. C1 to C29: class 1 to class 29. The numbers from 0 to 220 correspond to the class label in GEE. NU: Not Used, NA: Not Available, TC: Tree Cover, G: Gain, L: Loss, D: Datamask, TH: Tree Hight, TCC: Tree Canopy Cover, TCF: Tree-Cover Fraction, and SCF: Shrub-Cover Fraction)

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15
C1	16	15	NA	7	11	60	$TCC<10$	200	0	2	$(TC<10)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH<1$	$1\cup 0$	0	Not(≥ 1)
C2	16	15	NA	7	11	100	$TCC<10$	$200\cup 150$	0	2	$(TC<10)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH<1$	$1\cup 0$	0	Not(≥ 1)
C3	10	10	1	6	6	30	$TCC<10$	140	NA	2	$(TC<10)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH<2$	$1\cup 0$	0	Not(≥ 1)
C4	7	7	2	NA	5	$20\cup(10<SCF<50)$	$TCC<10$	150	0	2	$(TC<10)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH<2$	$1\cup 0$	0	Not(≥ 1)
C5	6	6	2	NA	5	$20\cup(SCF>50)$	$TCC<10$	130	0	2	$(TC<10)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH<2$	$1\cup 0$	0	Not(≥ 1)
C6	NA	NA	NA	4	4	$4+(15<TCF<30)$	$15<TCC<30$	60	NA	1	$(15<TC<30)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C7	NA	NA	NA	4	4	$4+(40<TCF<60)$	$40<TCC<60$	50	NA	1	$(40<TC<60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C8	4	4	6	4	4	$4+(TCF>60)$	$TCC>60$	50	NA	1	$(TC>60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C9	NA	NA	NA	3	3	$3+(15<TCF<30)$	$15<TCC<30$	NA	NA	1	$(15<TC<30)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C10	NA	NA	NA	3	3	$3+(40<TCF<60)$	$40<TCC<60$	NA	NA	1	$(40<TC<60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C11	3	3	8	3	3	$3+(TCF>60)$	$TCC>60$	NA	NA	1	$(TC>60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C12	NA	NA	NA	2	2	$2+(15<TCF<30)$	$15<TCC<30$	40	NA	1	$(15<TC<30)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C13	NA	NA	NA	2	2	$2+(40<TCF<60)$	$40<TCC<60$	40	NA	1	$(40<TC<60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C14	2	2	5	2	2	$2+(TCF>60)$	$TCC>60$	40	NA	1	$(TC>60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C15	9	9	NA	1	1	$1+(15<TCF<30)$	$15<TCC<30$	90	NA	1	$(15<TC<30)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	$Not(\geq 1)$
C16	8	8	4	1	1	$1+(40<TCF<60)$	$40<TCC<60$	70	NA	1	$(40<TC<60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C17	1	1	7	1	1	$1+(TCF>60)$	$TCC>60$	70	NA	1	$(TC>60)\cap(G=0)\cap(L=0)\cap(D\neq 2)$	$TH>2$	$1\cup 0$	0	Not(≥ 1)
C18	11	11	NA	NA	NA	90	$TCC>10$	170	NA	NA	$(TC>10)\cap(G=0)\cap(L=0)\cup(D=2)$	$TH>2$	$2\cup 3$	1	Not(≥ 1)
C19	11	11	NA	NA	NA	90	$TCC>10$	$a.160\cup 180$ $b.Not(170)$	NA	NA	$(TC>10)\cap(G=0)\cap(L=0)\cup(D=2)$	$TH>2$	$2\cup 3$	1	Not(≥ 1)
C20	11	11	NA	NA	NA	90	$TCC<10$	$160\cup 170$ $\cup 180$	NA	NA	$(TC<10)\cap(G=0)\cap(L=0)\cup(D=2)$	$TH<2$	$2\cup 3$	1	Not(≥ 1)
C21	17	0	0	0	0	200	NA	210	NA	3	NA	NA	3	1	Not(≥ 1)
C22	17	0	0	0	0	80	NA	210	NA	3	NA	NA	3	1	Not(≥ 1)
C23	15	NA	NA	NA	10	70	NA	220	NA	NA	NA	NA	$1\cup 0$	0	Not(≥ 1)
C24	12	12	$3\cup 1$	$5\cup 6$	$7\cup 8$	40	NA	$11\cup 14$	$1\cup 2\cup 3$ $\cup 4\cup 5$	NA	NA	NA	$2\cup 3$	$0\cup 4\cup 5$ $\cup 8\cup 10$	Not(≥ 1)
C25	12	12	1	6	7	40	NA	11	$1\cup 2$	NA	NA	NA	$1\cup 0$	0	Not(≥ 1)
C26	12	12	1	6	7	40	NA	14	$3\cup 4\cup 5$	NA	NA	NA	$1\cup 0$	0	Not(≥ 1)
C27	12	12	3	5	8	40	NA	11	$1\cup 2$	NA	NA	NA	$1\cup 0$	0	Not(≥ 1)
C28	12	12	3	5	8	40	NA	14	$3\cup 4\cup 5$	NA	NA	NA	$1\cup 0$	0	Not(≥ 1)
C29	13	13	10	8	9	50	NA	190	NA	NA	NA	NA	$1\cup 0$	0	NU

2.1.3 Combining products across time and space

For each LULC class, we built a consensus image with its global distribution by agreement across the 12 LULC products (P1 to P12). For products that only contained one image (P1 to P7) referred to a particular year or period, we obtained a binary

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 4. Description of the Temporal-Spatial Combination of the 15 global LULC products (P1:P15) masks to build a consensus image for each LULC class. ([1]: Inter-annual combination used in all products except in P13, where we first calculated the inter-annual Mean and then transformed it into a water no-water binary mask)

Class Id(s)	LULC class(s)	Spatial Combination	Temporal Combination [1]
C1	Barren lands	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C2	Moss and Lichen lands	Rule 2: Mean(P1 : P5, P7 : P12) * P6 * P13 * P14 * P15	MEAN
C3	Grasslands	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C4	Open Shrublands	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C5	Close Shrublands	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	MEAN
C6:C17	Forests	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	MEAN
C18	Mangrove Wetlands	Rule 3: Mean(P1 : P7, P9 : P14) * P8 * P15	MEAN
C19	Swamp Wetlands	Rule 4: Mean(P1 : P8.a, P9 : P12) * P8.b * P15	MEAN
C20	Marshland Wetlands	Rule 5: Mean(P1 : P6, P8 : P10, P13 : P14) * P7 * P11 * P12 * P15	MEAN
C21:C22	Water Bodies	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C23	Permanent Snow	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C24	Croplands Flooded with seasonal water	Rule 6: Mean(P1 : P12) * (P13 OR P14) * P15	AND
C25:C26	Cereal Croplands	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C27:C28	Broadleaf Croplands	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND
C29	Urban and built-up areas	Rule 1: Mean(P1 : P12) * P13 * P14 * P15	AND

mask with the targeted LULC class as 1. As further explained below, for products that were a collection over years (P8 to P12), we first obtained a binary mask for each year and then produced their combination over years (Table 4). Then, we obtained the spatial agreement over these 12 masks and used two water bodies products (P13 and P14) and one impervious surface product (P15) to further refine the consensus.

The inter-annual combination of LULC masks was based on two types of rules according to the inter-annual consistency of some products for specific LULC classes. In general, the combination over time was done using the AND operator when handling classes with high temporal stability, namely Urban and built-up areas, water bodies, Permanent Snow, Croplands, Open Shrublands, Barren lands, and Grasslands. Alternatively, since some LULC products showed implausible inter-annual changes for specific classes (e.g., wetlands affected by droughts or large areas of no-forest cover in one year preceded and followed by forest in the previous and following years, respectively), to avoid unrealistic data loss, the MEAN operator was used in the following classes: Moss and Lichen lands, Forests, Close Shrublands, and Wetlands.

Afterwards, for each LULC class, a spatial combination of the 12 masks, one per product, was performed following six specific rules according to the global abundance of each class (Table 4). In general (Rule 1), the MEAN operator was used across products P1 to P12 and the result was multiplied by two water class masks (P13 and P14) to eliminate water pixels from land classes and land pixels from water classes, and by impervious surface class mask of P15 to eliminate impervious pixels from all classes but urban. More relaxed rules where applied to five LULC classes that had too few pixels (less than 1000) with the general rule (see details in Table 4). The application of the explained spatio-temporal combination resulted in a mask for

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

each LULC class produced at the resolution of the finest product (i.e., 30 m), where each pixel had a consensus level value p in $[0, 1]$. Hence, p indicates the consensus degree for a pixel to belong to the class represented by each LULC mask.

2.1.4 Re-projection and Selection of purity threshold

The 30-meter resolution LULC consensus was reprojected to MODIS resolution (approximately equal to 500 m) using the spatial MEAN reducer. This 500 m average consensus was used to explore different pixel-purity thresholds θ for each LULC class. We kept $\theta = 1$, when the number of 500-meter pixels retrieved was greater than 1000, which meant that those pixels were 100% pure for that class. When the number of these pixels was lower than 1000 at $\theta = 1$, we decreased θ to 0.75%, which meant that up to 25% of the pixel could be occupied by another LULC class. In any case, our dataset provides as metadata at pixel level, the spatial purity percentage, so the user can control the desired purity and subsequent sample size (Table 6).

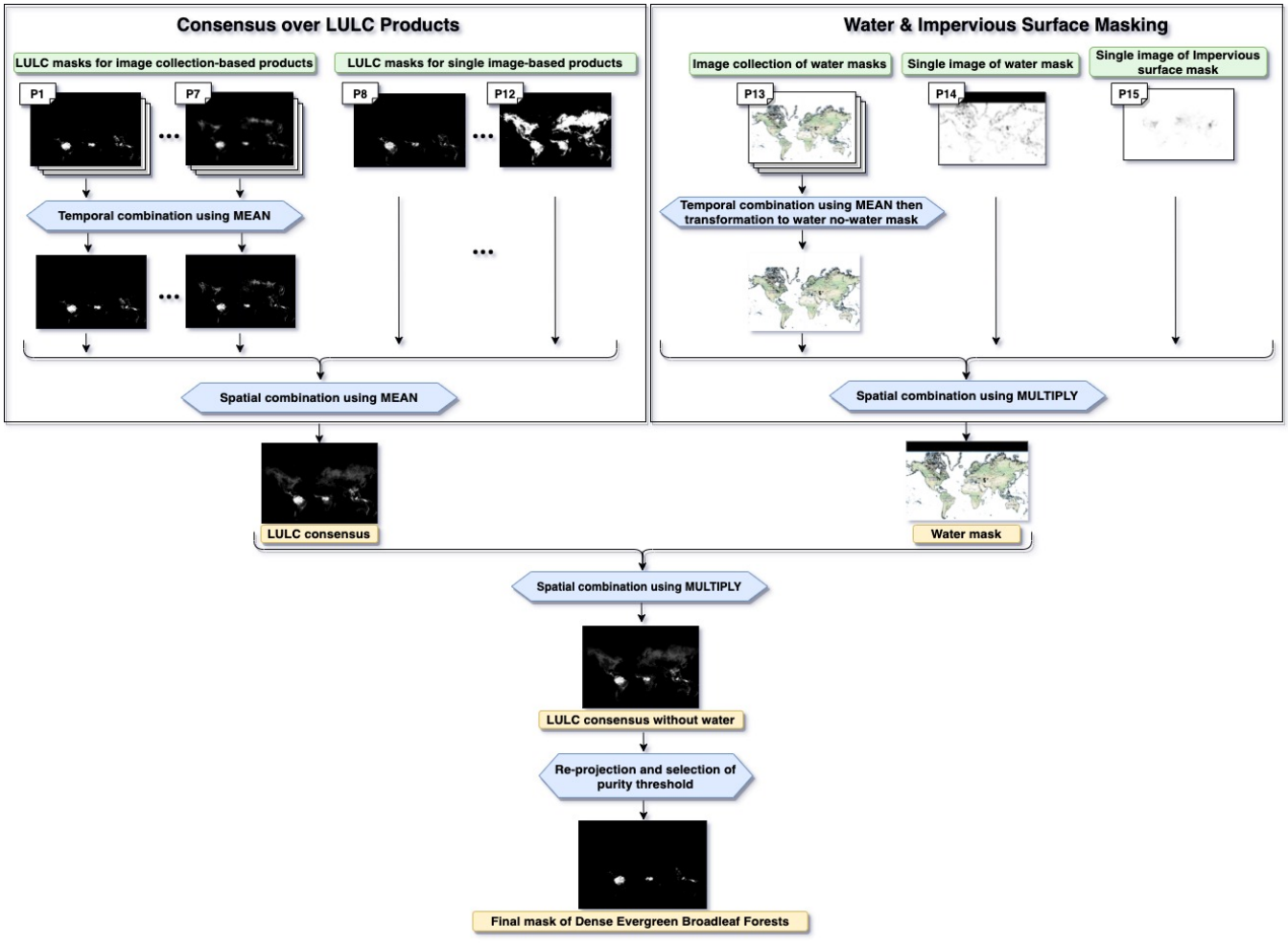
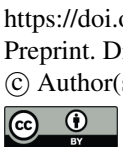


Figure 2. Example of the final mask creation process for the Dense Evergreen Broadleaf Forests LULC class produced through the spatial agreement over the 15 global LULC products available in GEE.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021



Open Access Earth System Science Data Discussions

To clarify the process of spatial agreement creation across the 15 global LULC products, we provide an example explaining the final mask creation for the class Dense Evergreen Broadleaf (Fig. 2). After performing the spatial agreement process for each LULC class, we combined the final class masks for all the 29 LULC classes to generate one global LULC map describing their distribution (Fig. 3).

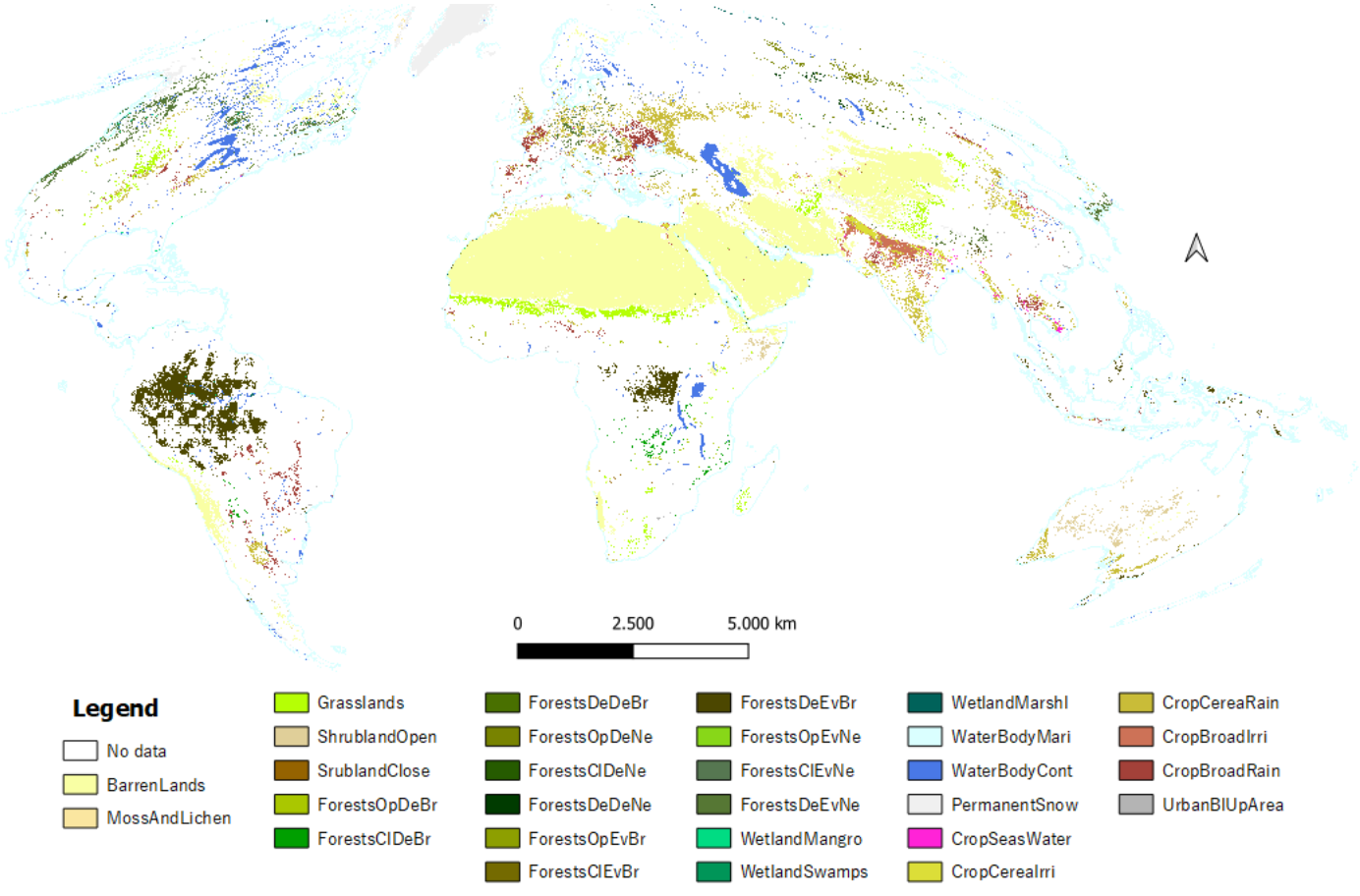


Figure 3. Distribution of the number of covered countries (FAO-L0) over the 29 LULC classes. This map combines all the LULC classes maps that were generated from the process of spatial agreement across the 15 global LULC products available in GEE.

2.2 Extracting times series of spectral data for 29 LULC classes globally

To extract the 19-year monthly time series of the seven 500-meter MODIS spectral bands for each of the 29 LULC classes throughout the entire world, we followed four steps (Fig. 4): 1) spatio-temporal filtering of Terra and Aqua data based on Quality Assessment flags, 2) aggregating the original 8-day Terra and Aqua data into monthly composites, 3) merging the two monthly time series into a Terra+Aqua Combined time series, and 4) data extraction and archiving (Fig. 5).

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

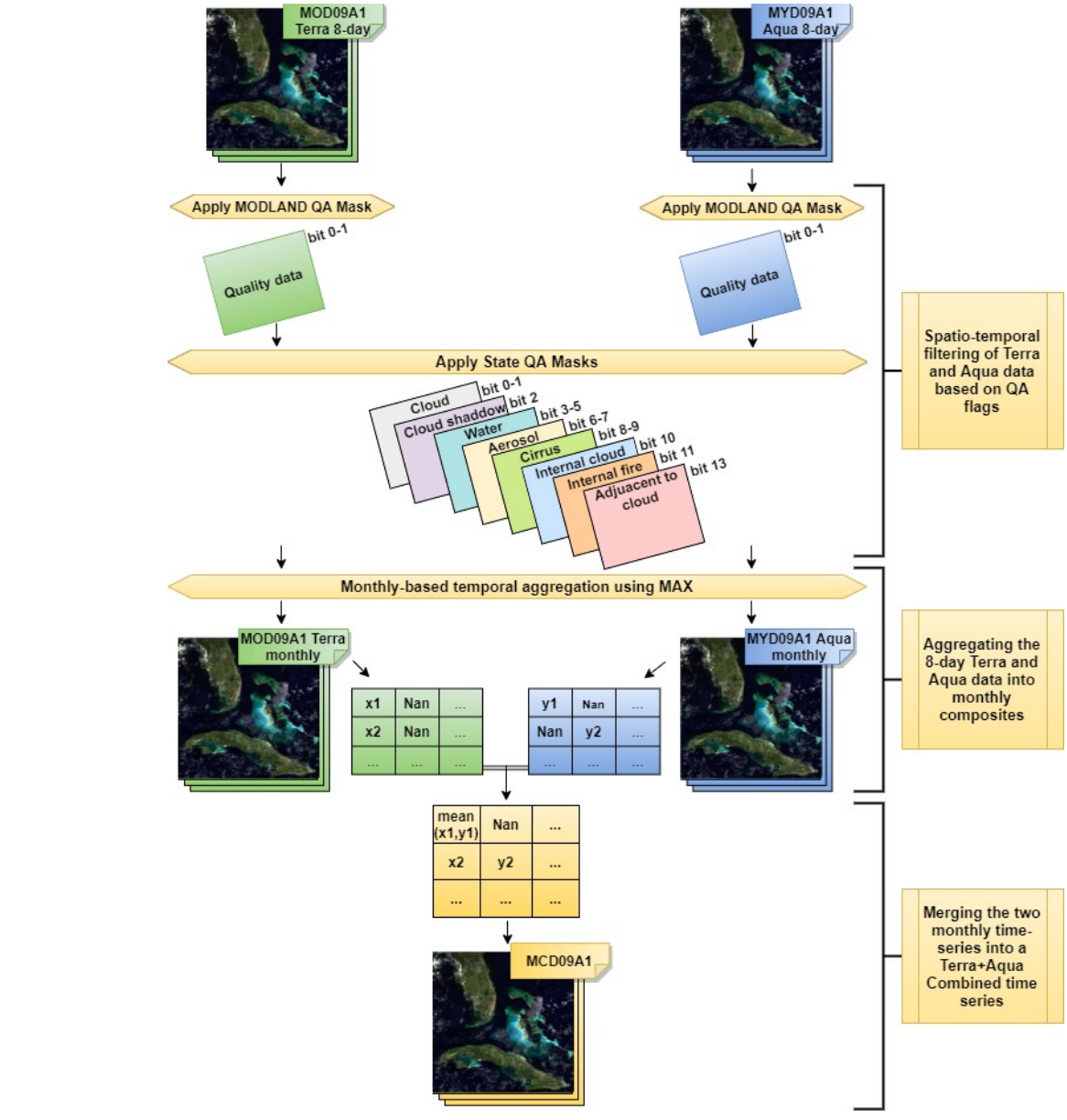


Figure 4. Description of the spatio-temporal filtering of Terra and Aqua, their aggregation into monthly composites, and their merging into Terra+Aqua Combined time series. This process aims to filter out spectral values affected by disruptive conditions, and to reduce the number of gaps in the multi-spectral time series for the 29 LULC classes.

2.2.1 Spatio-temporal filtering of Terra and Aqua data based on Quality Assessment flags

MODIS sensor is known by its high temporal coverage ensured by Terra and Aqua satellites, and by its innovative features (spectral, spatial, and temporal) that are highly suitable for LULC mapping and change detection (García-Mora et al., 2012;

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Xiong et al., 2017). Thus, we used two MODIS products MOD09A1 and MYD09A1 that estimate the 8-day surface spectral reflectance for the seven 500-meter bands from Terra and Aqua, respectively. Since Aqua was launched three years later (in 2002) after Terra had been launched, the acquisition time of our dataset is from 2002-07-04 to 2021-01-25.

The quality of any time series of satellite imagery is affected by the internal malfunction of satellite sensors, atmospheric (e.g., clouds, shadows, cirrus, etc.) or land (e.g., floods, snow, fires, etc.) conditions. In addition to the spectral bands, MODIS products provide 'Quality Assessment' (QA) flags as metadata bands to allow the user to filter out spectral values affected by disruptive conditions. Therefore, all QA flags were used to remove noise, spurious values, and outliers in the image collection. "MODLAND QA" flags (bits 0-1) were used to only select pixel values produced at 'ideal quality'. Then, "State QA" flags were used to mask out clouds (bits 0-1), internal clouds (bit 10), pixels adjacent to clouds (bit 13), cirrus (bits 8-9), cloud shadows (bit 2), high aerosol quantities (bits 6-7), and internal fires (bit 11). Regarding the water flag (bits 3-5), it was used to mask out water pixels in terrestrial systems (but not in Permanent Snow, and in Croplands flooded with seasonal water).

2.2.2 Aggregating the original 8-day Terra and Aqua data into monthly composites

Filtering the MODIS Terra and Aqua data records produced many missing-value gaps in their 8-day time series. To overcome this issue and further reduce the presence of noise in our dataset, the original 8-day time series were aggregated into monthly composites by computing the maximum over the observations of each month. Indeed, despite reducing the temporal resolution from 8-day to monthly composites shortened the time series size, it generated two datasets with less missing values and clear monthly patterns, which are more intuitive to track LULC dynamics than the 8-day patterns.

2.2.3 Merging the two monthly time series into a Terra+Aqua Combined time series

Terra satellite daily orbits above the Earth's surface from north to south in the morning at around 10:30 local time, while Aqua orbits in the opposite direction in the afternoon at around 13:30. Having two opportunities per day at each location increases the chances of capturing an image under good atmospheric conditions. To further reduce the number of missing values in our dataset, we merged the monthly time series provided by these two satellites into a Terra+Aqua Combined time series. That is, for each pixel and month, when both Terra and Aqua had values, we used the mean between them; when one satellite had a missing value, we used the available one; and when both of them had missing values, the combined value remains missing.

2.2.4 Data extraction and archiving

One of the biggest advantages of our dataset is its global scale characteristic since all the LULC classes data were extracted globally from all regions over the world. To take into account all the differences across the globe and thinking of regional interests that users could have, for each LULC class we organized the dataset into two levels. We used the Food and Agricultural Organization's (FAO's) Global Administrative Unit Layers (GAUL) product available in GEE to arrange our dataset into FAO-L0 (i.e., countries) and FAO-L1 (i.e., departments, states, provinces). For all 500-meter pixels selected at each FAO-L1 administrative unit, we retrieved the time series of the 7 spectral bands, together with the LULC class code and name,

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

205 pixel id, pixel purity percentage, percentage of non-missing data per band throughout the time series, average Global Human Modification index (gHM), and the latitude and longitude of the pixel center. Average gHM was derived as the spatial mean projected to MODIS resolution of the gHM (Kennedy et al., 2019) product available in GEE, which provides a cumulative measures of human modification of terrestrial lands. Finally, we exported the 7-band time series of each FAO-L1 unit and structured it as explained in section 3.

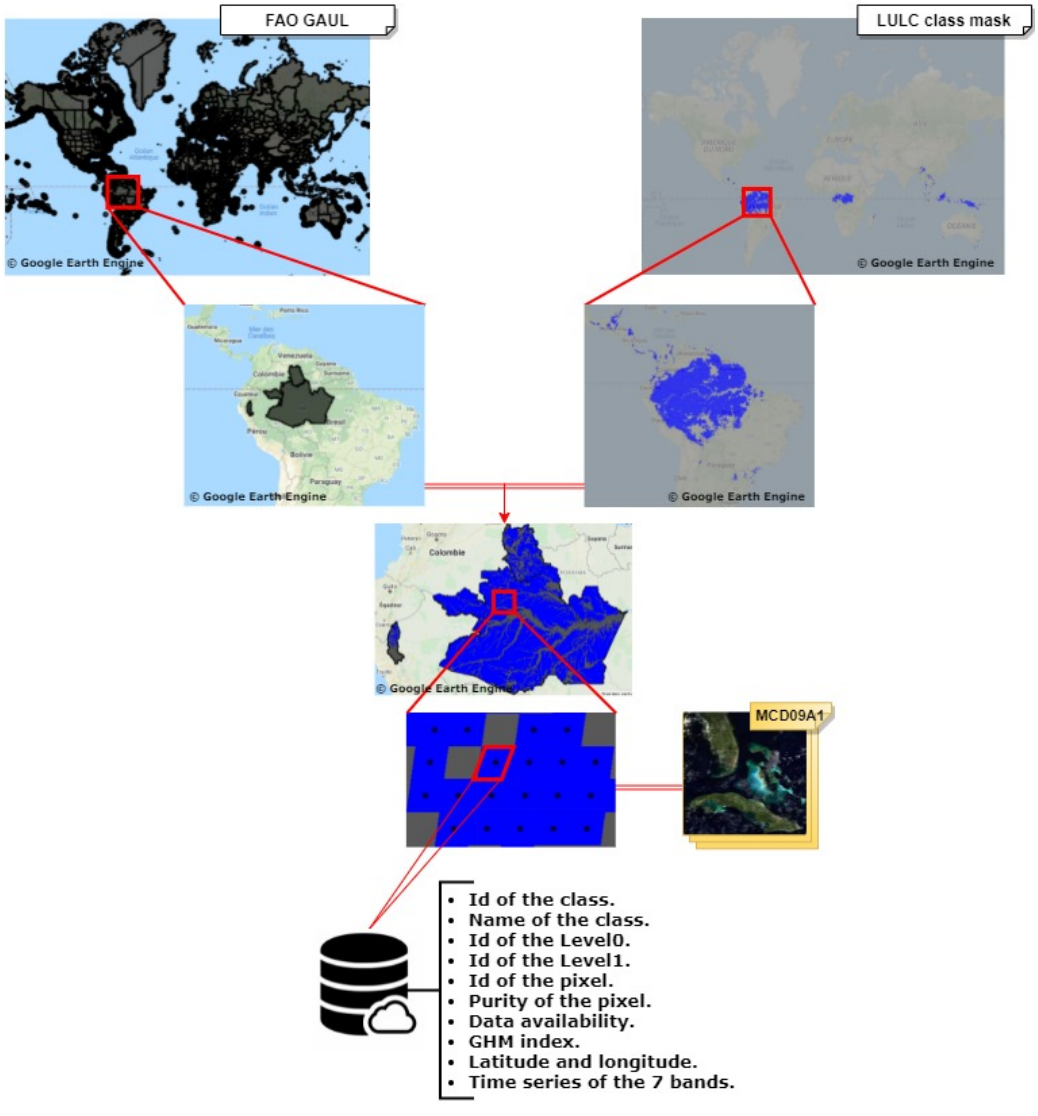


Figure 5. Description of the data extraction process for the LULC class Dense Evergreen Broadleaf Forests. It consists in looping over all the world's countries (FAO-L0) and departments (FAO-L1) that intersect with this LULC class mask. Then, it extracts the time series of the seven spectral bands of MODIS (MCD09A1-monthly-500m) along with a set of metadata at the pixel level.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

3 Data

To organize and assess the quality of the extracted global data for all the 29 LULC classes, we first present the description of the dataset structure stored in the repository[2] (Khaldi et al., 2021), then we evaluate the quality of its annotation process.

3.1 Data structure description

TimeSpec4LULC is provided in 30 different ZIP files owning the name of the 29 LULC classes (The Barren Lands class is divided into two files since it is too large). Within each ZIP file, there exists a set of seven CSV files, each one corresponding to one of the seven spectral bands. The naming of each file follows this structure: IdOfTheClass_NameOfTheClass_ModisBand.csv For example, for band 1 of the Barren Lands class, the filename is: 01_BarrenLands_MCD09A1b01.csv

Inside each CSV file, rows represent the collected pixels for that class. The first 11 columns contain the following metadata:

- "IdOfTheClass": Id of the class.
- "NameOfTheClass": Name of the class.
- "IdOfTheLevel0": Id of the FAO-L0 (i.e., countries).
- "IdOfTheLevel1": Id of the FAO-L1 (i.e., departments, states, or provinces depending on the country).
- "IdOfThePixel": Id of the pixel.
- "PurityOfThePixel": Spatial and inter-annual consensus for this class across multiple land-cover products, i.e., purity of the pixel.
- "DataAvailability": percentage of non-missing data per band throughout the time series.
- "Index_GHM": average of Global Human Modification index (gHM).
- "Lat": Latitude of the pixel center.
- "Lon": Longitude of the pixel center.
- ".geo": (Longitude, Latitude) of the pixel center with more precision.

And, the last 223 columns contain the 223 monthly observations of the time series for one spectral band from 2002-07 to 2021-01. Along with the dataset, an Excel file named 'Countries_Departments_FAO-GAUL' containing the FAO-L0 and the FAO-L1 Ids and names (following the FAO-GAUL standards) is provided.

[2]https://zenodo.org/record/5020024.YWARotpBxaQ

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

3.2 Data quality control

The quality of the dataset annotation was assessed and validated visually by two co-author experts using two very high resolution imagery (< 1m/pixel) sources, namely Google Earth and Bing Maps imagery[3]. The assessment process includes three stages.

- First, a set of 100 samples is carefully selected from each class following the minimum distance criteria. That is, depending on the overall size of LULC class, the pixels within a minimum distance with respect to each pixel are eliminated until getting 100 pixels evenly distributed over the globe. Fig. 6 shows the distribution of the 2,900 selected pixels.
- Second, the class of each pixel of the 29×100 samples is identified visually by the expert eye following the next rule. We consider as ground truth the dominant LULC class, such LULC class occupies at least 70% of the pixel. The presence of up to 30% of features of other different LULC classes within the dominate class are ignored.
- Once the validated LULC classification matrix was obtained (Table 5), the F1 score was calculated at all the LULC levels (from L0 to L5), labeled as LULC-L5 in Fig. 1. As it can be noticed, as we go up from level L0 to L5 the obtained dataset accuracy increases from 87% to 96% due mainly to the forests classification.

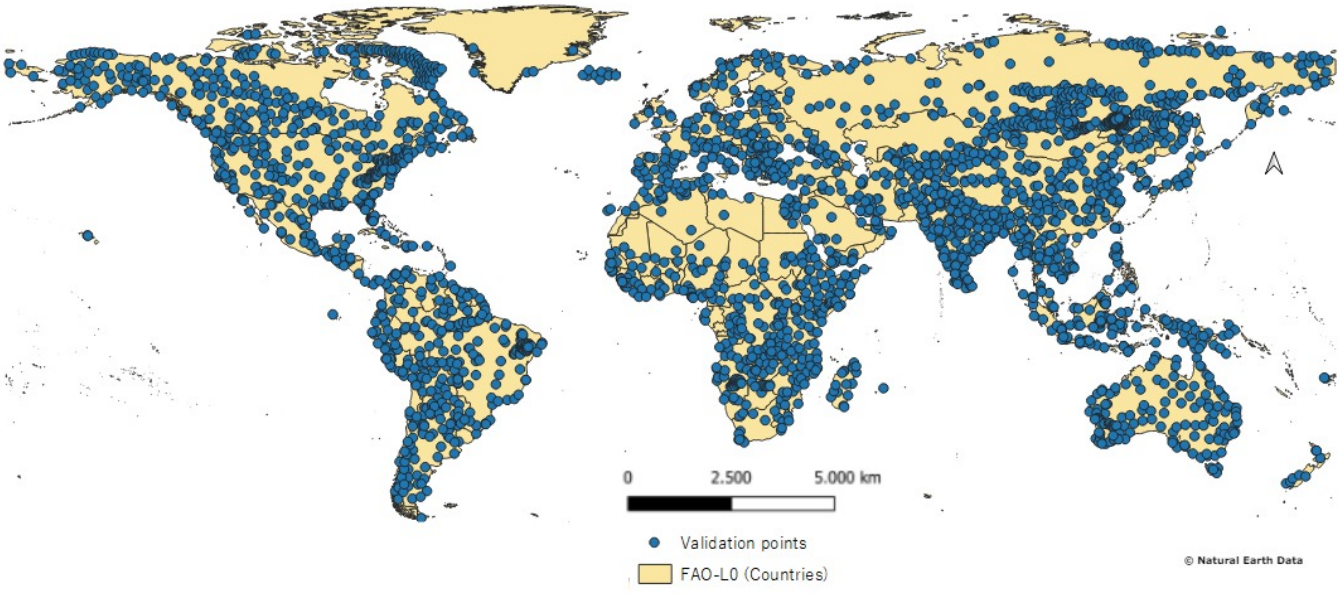


Figure 6. Global distribution of the selected 2,900 pixels to perform the quality control of the TimeSpec4LULC dataset for all the 29 LULC classes.

[3]https://www.bing.com/maps/

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 5. Description of the data quality control results of the 2900 pixels for each LULC level (L0 to L5) using the F1 score. The correspondence between long and short names of the LULC-classes is provided in Table 6

L0	F1	L1	F1	L2	F1	L3	F1	L4	F1	L5	F1
Land Cover	0.98	Terrestrial Lands	0.98	BarrenLands	0.88	BarrenLands	0.88	BarrenLands	0.88	BarrenLands	0.88
				MossAndLichen	NA	MossAndLichen	NA	MossAndLichen	NA	MossAndLichen	NA
				Grasslands	0.68	Grasslands	0.68	Grasslands	0.68	Grasslands	0.68
				Shrubland	0.87	ShrublandOpen	0.74	ShrublandOpen	0.74	ShrublandOpen	0.74
						SrublandClose	0.96	SrublandClose	0.96	SrublandClose	0.96
				Forests	0.98	ForestsDe	0.99	ForestsDeBr	0.98	ForestsOpDeBr	0.81
										ForestsCIDeBr	0.79
										ForestsDeDeBr	0.93
								ForestsDeNe	1	ForestsOpDeNe	0.9
										ForestsCIDeNe	0.86
										ForestsDeDeNe	0.91
						ForestsEv	0.97	ForestsEvBr	0.94	ForestsOpEvBr	0.67
										ForestsCIEvBr	0.78
										ForestsDeEvBr	0.93
								ForestsEvNe	1	ForestsOpEvNe	0.85
										ForestsCIEvNe	0.84
										ForestsDeEvNe	0.95
				PermanentSnow	0.98	PermanentSnow	0.98	PermanentSnow	0.98	PermanentSnow	0.98
		Aquatic Lands	0.94	Wetland	0.92	WetlandMangro	0.88	WetlandMangro	0.88	WetlandMangro	0.88
						WetlandSwamps	0.91	WetlandSwamps	0.91	WetlandSwamps	0.91
						WetlandMarshl	0.95	WetlandMarshl	0.95	WetlandMarshl	0.95
				WaterBody	0.97	WaterBodyMari	0.88	WaterBodyMari	0.88	WaterBodyMari	0.88
						WaterBodyCont	0.86	WaterBodyCont	0.86	WaterBodyCont	0.86
Land Use	0.93	Crop Lands	0.95	CropSeasWater	0.85	CropSeasWater	0.85	CropSeasWater	0.85	CropSeasWater	0.85
				CropCerea	0.92	CropCereaIrri	0.92	CropCereaIrri	0.92	CropCereaIrri	0.92
						CropCereaRain	0.91	CropCereaRain	0.91	CropCereaRain	0.91
				CropBroad	0.98	CropBroadIrri	0.96	CropBroadIrri	0.96	CropBroadIrri	0.96
						CropBroadRain	0.96	CropBroadRain	0.96	CropBroadRain	0.96
		UrbanBlUpArea	0.93	UrbanBlUpArea	0.93	UrbanBlUpArea	0.93	UrbanBlUpArea	0.93	UrbanBlUpArea	0.93
Mean	0.96		0.95		0.91		0.90		0.91		0.87

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

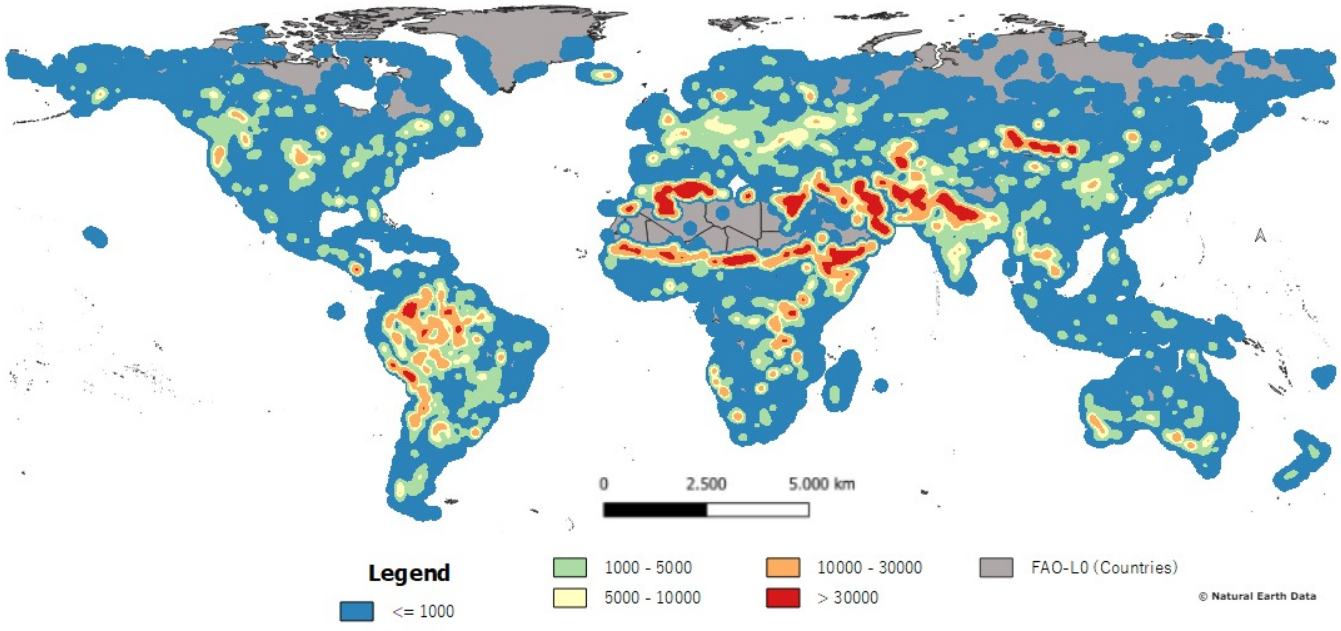


Figure 7. Global density of all the 11,856,992 collected pixels in the TimeSpec4LULC dataset for all the 29 LULC classes.

4 Results and Discussion

The total number of collected time series (pixels) in all the 29 LULC classes is 11,856,992, which is large enough to build high quality DL models. The density of the collected pixels over the world is depicted by Fig. 7.

The number of collected time series, at the lowest purity, differs from one class to another according to the global abundance of each class (Table 6). But, in general, we have at least 1,158 time series in each LULC class which is sufficient for DL modeling. In 15 LULC classes, the number of collected time series, at purity 100%, is at least 2,225 per class. This means that these classes have enough pure spectral information describing their behavior over time to train DL models with very high accuracy. However, in the remaining 14 LULC classes, the number of time series, collected at purity 100%, is either small (with Close shrublands, Dense Evergreen Needleleaf Forests, and Marshland wetlands) or null in the remaining Wetlands classes and Forests classes (except Dense Broadleaf classes). This implies that, within 500 m pixels, the LULC products are less consistent within these classes and that it may exist remaining noise in one LULC class from other LULC classes. Since our dataset provides the level of spatial purity at pixel level, the user can always select the desired purity threshold.

Certain number of time series in each LULC class still contain some missing data that could be handled neither with the monthly aggregation process nor with the Terra-Aqua merging process (Table 7). The LULC classes that have less than 1000 time series (i.e., MODIS pixels) without missing data in at least one spectral band are Moss and Lichen lands, Close Shrublands, Open and Dense Deciduous Broadleaf Forests, all Deciduous Needleleaf Forests, Open Evergreen Broadleaf Forests, Mangrove wetlands, Swamp wetlands, Permanent snow, and Croplands flooded with seasonal water. This means that the multi-spectral

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

time series information of these classes is hugely affected by atmospheric and/or land conditions. For almost all LULC classes (except Barren Lands, Continental Water Bodies, and Rainfed Cereal Croplands), the number of time series without gaps is exactly the same over all the spectral bands excluding Band 6 which is the most contaminated by gaps. This severe drop in the number of time series without gaps in Band 6 compared to the other bands is due to the known "dead lines" in Aqua band 6 caused by the already reported malfunctioning or noise in some of its detectors (Zhang et al., 2018b).

Our dataset is distributed all over the world's FAO-L0 and FAO-L1 partitions (Fig. 8, Fig. 9, and Table 8). For almost all the 29 LULC classes, the collected time series covers more than 9 countries, except with Moss and Lichen lands, which were collected from 2 countries, and Deciduous Needleleaf Forests, which were collected from 3 countries. This small number of countries covered by these classes is due to their natural scarce distribution over the world. Contrary, some of the LULC classes, namely Water Bodies (Marine and Continental), Rainfed Croplands (Cereal and Broadleaf), and Urban and built-up areas have a broad world coverage, i.e., more than 80 countries and more than 570 departments (Fig. 8 and Fig. 9).

According to Fig. 10, the gHM index of the five cropland classes, and the Urban and built-up areas class is widely higher (more than 59% of human change) compared to the other land cover classes, which proves their accurate annotation as human Land Uses.

This smartly, pre-processed, and annotated dataset is targeted towards scientific users interested in developing and evaluating various DL models to detect LULCs and monitor their change. For example, TimeSpec4LULC can be used i) to characterize the seasonal and inter-annual dynamics and changes of vegetation types and LULC classes, ii) to perform environmental monitoring, management and planning, ii) to study the intra-class behaviors of LULCs, i.e., assess the behavior of one specific LULC in different areas of the world and see whether it maintains the same pattern or it reveals different patterns, and iii) to study the inter-class differences and similarities of LULCs, i.e., recognise and compare the patterns and dynamics of all LULCs (e.g., time series segmentation).

5 Conclusions

Accurate LULC mapping and change detection is highly relevant for many applications, including Earth system modeling, environmental monitoring, management and planning, or natural hazards assessment, among many others. However, there still exists a high level of disagreement across current global LULC products, particularly for some LULC classes. To address the challenge of improving LULC products and change detection, we have created a smart open-source global dataset of multi-spectral time series for 29 LULC classes containing almost 11.85 million pixels annotated by using the spatio-temporal agreement across the 15 global LULC products available in GEE. The 29 LULC classes were hierarchically grouped into a legend with five levels. The monthly 7-band time series dataset was made by merging the two MODIS sensor data records, Terra and Aqua, at 500 m resolution and expands 19 years from 2002 to 2021. Each pixel is provided with a set of meta-data about geographic coordinates, country and departmental divisions, spatio-temporal consistency across LULC products, temporal data availability, and the global human modification index. Finally, to assess the annotation quality of the dataset, a sample of 100 pixels per class, evenly distributed around the world, was selected by maximizing the distance among sampled

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

pixels, and validated with photo-interpretation by experts using very high resolution images from both Google Earth and Bing Maps. The overall accuracy (F1 value) of the annotation varied from 96% at the coarser classification level to 87% at the finest level. This smartly, pre-processed, and annotated dataset is targeted towards scientific users interested in developing and
300 evaluating various machine learning models, including deep learning networks, to perform global LULC mapping and change detection.

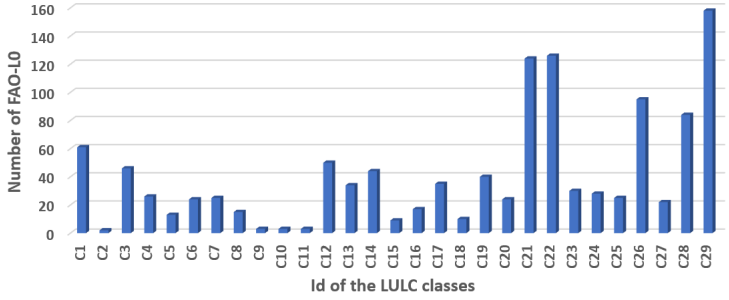


Figure 8. Distribution of the number of covered countries (FAO-L0) over the 29 LULC classes.

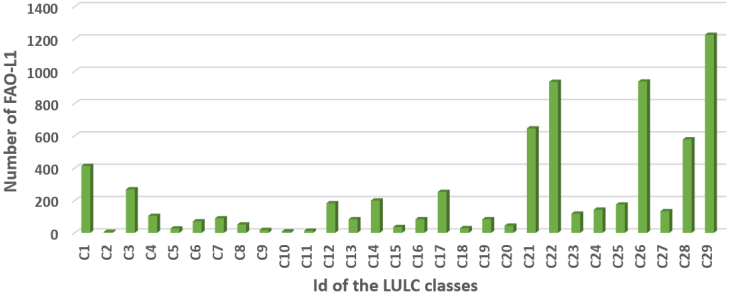


Figure 9. Distribution of the number of covered departments (FAO-L1) over the 29 LULC classes.

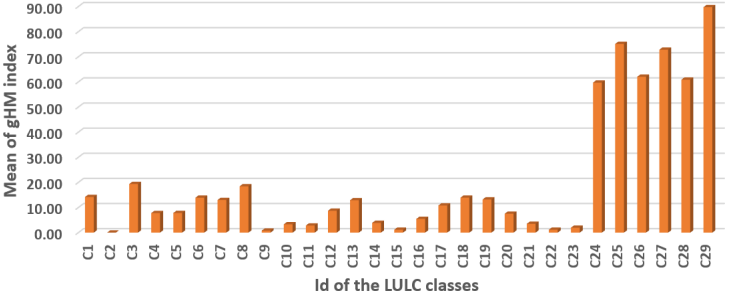


Figure 10. Distribution of the mean of the Global Human Modification (gHM) index over the 29 LULC classes.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 6. Description of the full and short name of each LULC class along with the number of pixels collected at purity 100% and at the lowest purity retrieved.

Class Id	LULC class full name	LULC class short name	# Pixels retrieved at purity 100%	Lowest purity in the dataset	# Collected pixels
C1	Barren lands	BarrenLands	5,866,886	1	5,866,886
C2	Moss and lichen lands	MossAndLichen	2,804	1	2,804
C3	Grasslands	Grasslands	890,842	1	890,842
C4	Open shrublands	ShrublandOpen	121,273	1	121,273
C5	Close shrublands	SrublandClose	4	0.95	2,910
C6	Open Deciduous Broadleaf Forests	ForestsOpDeBr	0	0.85	6,956
C7	Close Deciduous Broadleaf Forests	ForestsClDeBr	0	0.85	35,445
C8	Dense Deciduous Broadleaf Forests	ForestsDeDeBr	2,225	1	2,225
C9	Open Deciduous Needleleaf Forests	ForestsOpDeNe	0	0.75	5,266
C10	Close Deciduous Needleleaf Forests	ForestsClDeNe	0	0.85	1,348
C11	Dense Deciduous Needleleaf Forests	ForestsDeDeNe	0	0.95	10,716
C12	Open Evergreen Broadleaf Forests	ForestsOpEvBr	0	0.75	22,315
C13	Close Evergreen Broadleaf Forests	ForestsClEvBr	0	0.85	3,245
C14	Dense Evergreen Broadleaf Forests	ForestsDeEvBr	1,198,448	1	1,198,448
C15	Open Evergreen Needleleaf Forests	ForestsOpEvNe	0	0.75	20,314
C16	Close Evergreen Needleleaf Forests	ForestsClEvNe	0	0.85	4,232
C17	Dense Evergreen Needleleaf Forests	ForestsDeEvNe	199	0.95	323,892
C18	Mangrove wetlands	WetlandMangro	0	0.85	4,707
C19	Swamp wetlands	WetlandSwamps	0	0.85	1,158
C20	Marshland wetlands	WetlandMarshl	80	0.85	9,303
C21	Marine water bodies	WaterBodyMari	107,467	1	107,467
C22	Continental water bodies	WaterBodyCont	977,759	1	977,759
C23	Permanent snow	PermanentSnow	99,798	1	99,798
C24	Croplands flooded with seasonal water	CropSeasWater	38,636	1	38,636
C25	Irrigated cereal croplands	CropCerealIrri	402,049	1	402,049
C26	Rainfed cereal croplands	CropCerealRain	786,837	1	786,837
C27	Irrigated broadleaf croplands	CropBroadIrri	392,091	1	392,091
C28	Rainfed broadleaf croplands	CropBroadRain	359,141	1	359,141
C29	Urban and built-up areas	UrbanBlUpArea	158,929	1	158,929

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 7. Description of the number of time series without missing data per spectral bands.

Class Id	Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
C1	4,378,501	4,378,501	4,378,501	4,378,473	4,378,501	3,947,731	4,378,501
C2	0	0	0	0	0	0	0
C3	414,963	414,963	414,963	414,963	414,963	310,326	414,963
C4	104,633	104,633	104,633	104,633	104,633	90,892	104,633
C5	1,099	1,099	1,099	1,099	1,099	886	1,099
C6	1,035	1,035	1,035	1,035	1,035	559	1,035
C7	7,675	7,675	7,675	7,675	7,675	5,463	7,675
C8	110	110	110	110	110	25	110
C9	11	11	11	11	11	1	11
C10	0	0	0	0	0	0	0
C11	0	0	0	0	0	0	0
C12	1,024	1,024	1,024	1,024	1,024	552	1,024
C13	1,957	1,957	1,957	1,957	1,957	1,498	1,957
C14	8,488	8,488	8,488	8,488	8,488	3,610	8,488
C15	1,740	1,740	1,740	1,740	1,740	1,574	1,740
C16	1,386	1,386	1,386	1,386	1,386	1,068	1,386
C17	22,488	22,488	22,488	22,488	22,488	8,219	22,488
C18	690	690	690	690	690	201	690
C19	287	287	287	287	287	257	287
C20	1,409	1,409	1,409	1,409	1,409	1,134	1,409
C21	11,640	11,640	11,640	11,640	11,640	1,151	11,640
C22	140,856	140,854	140,856	140,856	140,856	44,302	140,856
C23	138	138	138	138	138	11	138
C24	269	269	269	269	269	96	269
C25	25,142	25,142	25,142	25,142	25,142	17,607	25,142
C26	262,632	262,630	262,630	262,630	262,630	181,382	262,630
C27	29,154	29,154	29,154	29,154	29,154	14,636	29,154
C28	79,686	79,686	79,686	79,686	79,686	43,080	79,686
C29	20,289	20,289	20,289	20,289	20,289	10,697	20,289

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Open Access Earth System Science Data Discussions

Table 8: Description of the number of FAO-L0 and FAO-L1 covered by the collected pixels along with the mean and (the standard deviation) of the pixels gHM index, purity, and temporal availability percentage over the seven spectral bands.

Class Id	# FAO L0	# FAO L1	gHM index	Purity (%)	Temporal Availability Percentage (%)						
					Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
C1	61	414	14.19	100	98.51	98.50	98.50	98.50	98.50	98.07	98.50
			(10.20)	(0.00)	(8.10)	(8.14)	(8.14)	(8.14)	(8.14)	(8.74)	(8.14)
C2	2	6	0.01	100	51.24	51.24	51.24	51.24	51.24	46.69	51.24
			(0.23)	(0.00)	(17.85)	(17.85)	(17.85)	(17.85)	(17.85)	(17.54)	(17.85)
C3	46	269	19.35	100	98.26	98.26	98.26	98.26	98.26	97.72	98.26
			(14.09)	(0.00)	(6.77)	(6.77)	(6.77)	(6.77)	(6.77)	(7.38)	(6.77)
C4	26	105	7.79	100	99.64	99.64	99.64	99.64	99.64	99.44	99.64
			(8.82)	(0.00)	(4.31)	(4.31)	(4.31)	(4.31)	(4.31)	(4.61)	(4.31)
C5	13	27	7.83	96.78	98.85	98.85	98.85	98.85	98.85	97.85	98.85
			(5.82)	(1.32)	(2.85)	(2.85)	(2.85)	(2.85)	(2.85)	(3.65)	(2.85)
C6	24	71	13.91	85.78	98.21	98.21	98.21	98.21	98.21	97.11	98.21
			(6.41)	(0.68)	(2.83)	(2.83)	(2.83)	(2.83)	(2.83)	(3.40)	(2.83)
C7	25	90	12.99	86.09	97.53	97.53	97.53	97.53	97.53	95.96	97.53
			(5.58)	(1.07)	(3.52)	(3.52)	(3.52)	(3.52)	(3.52)	(4.62)	(3.52)
C8	15	52	18.44	100	90.18	90.18	90.18	90.18	90.18	88.12	90.18
			(9.86)	(0.00)	(11.65)	(11.65)	(11.65)	(11.65)	(11.65)	(12.69)	(11.65)
C9	3	18	0.79	76.37	89.29	89.29	89.29	89.29	89.29	88.05	89.29
			(2.39)	(1.29)	(8.47)	(8.47)	(8.47)	(8.47)	(8.47)	(8.83)	(8.47)
C10	3	9	3.31	85.51	78.51	78.51	78.51	78.51	78.51	76.07	78.51
			(3.73)	(0.43)	(7.87)	(7.87)	(7.87)	(7.87)	(7.87)	(8.00)	(7.87)
C11	3	13	2.83	96.10	85.79	85.79	85.79	85.79	85.79	82.69	85.79
			(3.77)	(0.92)	(9.20)	(9.20)	(9.20)	(9.20)	(9.20)	(9.64)	(9.20)
C12	50	183	8.71	77.27	84.65	84.65	84.65	84.65	84.65	79.97	84.65
			(14.26)	(2.07)	(18.30)	(18.30)	(18.30)	(18.30)	(18.30)	(19.28)	(18.30)
C13	34	84	12.89	86.46	90.67	90.67	90.67	90.67	88.50	85.85	90.67
			(12.06)	(1.44)	(20.30)	(20.30)	(20.30)	(20.30)	(20.30)	(21.68)	(20.30)
C14	44	200	3.88	100	88.86	88.87	88.87	88.87	88.87	82.95	88.87

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Table 8: *– Continued from previous page*

Class Id	# FAO L0	# FAO L1	gHM index	Purity (%)	Temporal Availability (%)						
					Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
			(4.04)	(0.00)	(9.17)	(9.10)	(9.10)	(9.10)	(9.10)	(9.93)	(9.10)
C15	9	36	1.15	76.54	59.21	59.21	59.21	59.21	59.21	56.90	59.21
			(3.30)	(1.31)	(37.87)	(37.87)	(37.87)	(37.87)	(37.87)	(37.74)	(37.87)
C16	17	84	5.46	86.13	88.61	88.61	88.61	88.61	88.61	86.55	88.61
			(6.47)	(1.03)	(24.75)	(24.75)	(24.75)	(24.75)	(24.75)	(25.84)	(24.75)
C17	35	253	10.81	85.40	85.40	85.40	85.40	85.40	85.40	82.27	85.40
			(13.92)	(1.41)	(26.61)	(26.61)	(26.61)	(26.61)	(26.62)	(27.26)	(26.61)
C18	10	29	13.92	87.58	97.93	97.93	97.93	97.93	97.93	95.82	97.93
			(9.70)	(2.38)	(4.14)	(4.14)	(4.14)	(4.14)	(4.14)	(5.98)	(4.14)
C19	40	84	13.20	87.08	95.58	95.58	95.58	95.58	95.58	92.65	95.58
			(9.92)	(2.16)	(7.01)	(7.01)	(7.01)	(7.01)	(7.01)	(9.57)	(7.01)
C20	24	44	7.52	90.43	74.97	74.97	74.97	74.97	74.97	71.57	74.97
			(11.96)	(3.86)	(17.53)	(17.53)	(17.53)	(17.53)	(17.53)	(18.45)	(17.53)
C21	124	646	3.50	100	90.98	90.98	90.96	90.98	90.98	84.29	90.98
			(11.10)	(0.00)	(11.43)	(11.44)	(11.52)	(11.43)	(11.43)	(14.40)	(11.43)
C22	126	934	1.15	100	87.98	87.98	87.96	87.98	87.98	81.46	87.98
			(6.02)	(0.00)	(13.29)	(13.30)	(12.67)	(13.29)	(13.29)	(16.54)	(13.29)
C23	30	119	1.96	100	78.50	78.50	78.50	78.50	78.50	72.06	78.50
			(5.63)	(0.00)	(11.52)	(11.52)	(11.52)	(11.52)	(11.52)	(12.85)	(11.52)
C24	28	143	59.72	100	79.60	79.60	79.60	79.60	79.60	74.86	79.60
			(18.78)	(0.00)	(9.45)	(9.45)	(9.45)	(9.45)	(9.45)	(9.84)	(9.45)
C25	25	175	75.11	100	92.13	92.13	92.13	92.13	92.13	89.93	92.13
			(6.07)	(0.00)	(9.80)	(9.80)	(9.80)	(9.80)	(9.80)	(10.67)	(9.80)
C26	95	936	62.05	100	95.43	95.43	95.43	95.43	95.43	94.02	95.43
			(13.42)	(0.00)	(10.47)	(10.47)	(10.47)	(10.47)	(10.47)	(11.50)	(10.47)
C27	22	134	72.82	100	88.47	88.47	88.47	88.47	88.47	85.97	88.47
			(7.21)	(0.00)	(9.63)	(9.63)	(9.63)	(9.63)	(9.63)	(10.73)	(9.63)
C28	84	578	60.89	100	95.00	95.00	95.00	95.00	95.00	93.50	95.00

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Table 8: *– Continued from previous page*

Class Id	# FAO L0	# FAO L1	gHM index	Purity (%)	Temporal Availability (%)						
					Band 1	Band 2	Band 3	Band 4	Band 5	Band 6	Band 7
			(12.60)	(0.00)	(9.46)	(9.46)	(9.46)	(9.46)	(9.46)	(10.55)	(9.46)
C29	158	1225	89.75	100	80.99	80.99	80.99	80.99	80.99	77.26	80.99
			(5.30)	(0.00)	(28.39)	(28.39)	(28.39)	(28.39)	(28.39)	(29.25)	(28.39)

Code and data availability. This dataset (Khaldi et al., 2021) is available to the public through an unrestricted data repository hosted by Zenodo at https://zenodo.org/record/5020024#.YWARotpBxaQ.

Author contributions. RK contributed to the conception of the dataset, implemented the code, performed all the data extraction and wrote the paper. DA-S contributed to the conception of the dataset, assessed its quality, provided guidance, and wrote the paper. EG assessed the quality of the dataset. YB contributed to the conception of the dataset. AE and FH provided edits and suggestions. ST contributed to the conception of the dataset, provided guidance, and wrote the paper.

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

Acknowledgements. This work was partially supported by DETECTOR (A-RNM-256-UGR18 Universidad de Granada/FEDER), LifeWatch SmartEcomountains (LifeWatch-2019-10-UGR-01 Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), BBVA DeepSCOP (Ayudas Fundación BBVA a Equipos de Investigación Científica 2018), DeepL-ISCO (A-TIC-458-UGR18 Ministerio de Ciencia e Innovación/FEDER), SMART-DASCI (TIN2017-89517-P Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), BigDDL-CET (P18-FR-4961 Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), RESISTE (P18-RT-1927 Consejería de Economía, Conocimiento, y Universidad from the Junta de Andalucía/FEDER), and Ecopotential (641762 European Commission). S. Tabik was supported by the Ramón y Cajal Programme (RYC-2015-18136).

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

References

Alexakis, D., Grillakis, M., Koutroulis, A. G., Agapiou, A., Themistocleous, K., Tsanis, I., Michaelides, S., Pashiardis, S., Demetriou, C., Aristeidou, K., et al.: GIS and remote sensing techniques for the assessment of land use change impact on flood hydrology: the case study of Yialias basin in Cyprus, Natural Hazards and Earth System Sciences, 14, 413–426, 2014.

Bartholome, E. and Belward, A. S.: GLC2000: a new approach to global land cover mapping from Earth observation data, International Journal of Remote Sensing, 26, 1959–1977, 2005.

Congalton, R. G., Gu, J., Yadav, K., Thenkabail, P., and Ozdogan, M.: Global land cover mapping: A review and uncertainty analysis, Remote Sensing, 6, 12 070–12 093, 2014.

de la Cruz, M., Quintana-Ascencio, P. F., Cayuela, L., Espinosa, C. I., and Escudero, A.: Comment on "The extent of forest in dryland biomes", Science, 358, 2017.

Estes, L., Chen, P., Debats, S., Evans, T., Ferreira, S., Kuemmerle, T., Ragazzo, G., Sheffield, J., Wolf, A., Wood, E., et al.: A large-area, spatially continuous assessment of land cover map error and its impact on downstream analyses, Global change biology, 24, 322–337, 2018.

Feddema, J. J., Oleson, K. W., Bonan, G. B., Mearns, L. O., Buja, L. E., Meehl, G. A., and Washington, W. M.: The importance of land-cover change in simulating future climates, Science, 310, 1674–1678, 2005.

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

Fritz, S., You, L., Bun, A., See, L., McCallum, I., Schill, C., Perger, C., Liu, J., Hansen, M., and Obersteiner, M.: Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets, Geophysical Research Letters, 38, 2011.

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

García-Mora, T. J., Mas, J.-F., and Hinkley, E. A.: Land cover mapping applications with MODIS: a literature review, International Journal of Digital Earth, 5, 63–87, 2012.

Gengler, S. and Bogaert, P.: Combining land cover products using a minimum divergence and a Bayesian data fusion approach, International Journal of Geographical Information Science, 32, 806–826, 2018.

Gómez, C., White, J. C., and Wulder, M. A.: Optical remotely sensed time series data for land cover classification: A review, ISPRS Journal of Photogrammetry and Remote Sensing, 116, 55–72, 2016.

Gong, P., Wang, J., Yu, L., Zhao, Y., Zhao, Y., Liang, L., Niu, Z., Huang, X., Fu, H., Liu, S., et al.: Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data, International Journal of Remote Sensing, 34, 2607–2654, 2013.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sensing of Environment, https://doi.org/10.1016/j.rse.2017.06.031, 2017.

Grekousis, G., Mountrakis, G., and Kavouras, M.: An overview of 21 global and 43 regional land-cover mapping products, International Journal of Remote Sensing, 36, 5309–5335, 2015.

Guirado, E., Tabik, S., Alcaraz-Segura, D., Cabello, J., and Herrera, F.: Deep-learning versus OBIA for scattered shrub detection with Google earth imagery: Ziziphus Lotus as case study, Remote Sensing, 9, 1220, 2017.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Guirado, E., Blanco-Sacristán, J., Rodríguez-Caballero, E., Tabik, S., Alcaraz-Segura, D., Martínez-Valderrama, J., and Cabello, J.: Mask R-CNN and OBIA Fusion Improves the Segmentation of Scattered Vegetation in Very High-Resolution Optical Sensors, Sensors, 21, 320, 2021.

Hoskins, A. J., Bush, A., Gilmore, J., Harwood, T., Hudson, L. N., Ware, C., Williams, K. J., and Ferrier, S.: Downscaling land-use data to provide global 30 ″estimates of five land-use classes, Ecology and Evolution, 6, 3040–3055, 2016.

Hubert-Moy, L., Thibault, J., Fabre, E., Rozo, C., Arvor, D., Corpetti, T., and Rapinel, S.: Time-series spectral dataset for croplands in France (2006–2017), Data in Brief, 27, 104 810, https://doi.org/https://doi.org/10.1016/j.dib.2019.104810, 2019.

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

Kennedy, C. M., Oakleaf, J. R., Theobald, D. M., Baruch-Mordo, S., and Kiesecker, J.: Managing the middle: A shift in conservation priorities based on the global human modification gradient, Global Change Biology, 25, 811–826, 2019.

Kerr, J. T. and Ostrovsky, M.: From space to species: ecological applications for remote sensing, Trends in ecology & evolution, 18, 299–305, 2003.

Khaldi, R., Alcaraz-Segura, D., Guirado, E., Benhammou, Y., and Tabik, S.: TimeSpec4LULC: A deep learning-oriented global dataset of MODIS Terra-Aqua multi-spectral time series measured from 2002 to 2021 for LULC mapping and change detection., https://doi.org/10.5281/zenodo.5020024, this research has been supported by DETECTOR (A-RNM-256-UGR18 Universidad de Granada/FEDER), LifeWatch SmartEcomountains (LifeWatch-2019-10-UGR-01 Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), BBVA DeepSCOP (Ayudas Fundación BBVA a Equipos de Investigación Científica 2018), Ramón y Cajal Programme (RYC-2015-18136), DeepL-ISCO (A-TIC-458-UGR18 Ministerio de Ciencia e Innovación/FEDER), SMART-DASCI (TIN2017-89517-P Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), BigDDL-CET (P18-FR-4961 Ministerio de Ciencia e Innovación/Universidad de Granada/FEDER), RESISTE (P18-RT-1927 Consejería de Economía, Conocimiento, y Universidad from the Junta de Andalucía/FEDER), and Ecopotential (641762 European Commission)., 2021.

Kong, F., Li, X., Wang, H., Xie, D., Li, X., and Bai, Y.: Land cover classification based on fused data from GF-1 and MODIS NDVI time series, Remote Sensing, 8, 741, 2016.

Lambin, E. F. and Geist, H. J.: Land-use and land-cover change: local processes and global impacts, Springer Science & Business Media, 2008.

Liu, L., Zhang, X., Gao, Y., Chen, X., Shuai, X., and Mi, J.: Finer-Resolution Mapping of Global Land Cover: Recent Developments, Consistency Analysis, and Prospects, Journal of Remote Sensing, 2021, 2021.

Loveland, T. R., Reed, B. C., Brown, J. F., Ohlen, D. O., Zhu, Z., Yang, L., and Merchant, J. W.: Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data, International Journal of Remote Sensing, 21, 1303–1330, 2000.

Luengo, J., García-Gil, D., Ramírez-Gallego, S., García, S., and Herrera, F.: Big data preprocessing - Enabling Smart Data, Cham: Springer, 2020.

Luoto, M., Virkkala, R., and Heikkinen, R. K.: The role of land cover in bioclimatic models depends on spatial resolution, Global ecology and biogeography, 16, 34–42, 2007.

Menke, S., Holway, D., Fisher, R., and Jetz, W.: Characterizing and predicting species distributions across environments and scales: Argentine ant occurrences in the eye of the beholder, Global Ecology and Biogeography, 18, 50–63, 2009.

Meyer, W. B., Meyer, W. B., BL Turner, I., et al.: Changes in land use and land cover: a global perspective, vol. 4, Cambridge University Press, 1994.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Patel, S. K., Verma, P., and Singh, G. S.: Agricultural growth and land use land cover change in peri-urban India, Environmental monitoring and assessment, 191, 1–17, 2019.

Pérez-Hoyos, A., García-Haro, F. J., and San-Miguel-Ayanz, J.: A methodology to generate a synergetic land-cover map by fusion of different land-cover products, International Journal of Applied Earth Observation and Geoinformation, 19, 72–87, 2012.

Pfeifer, M., Disney, M., Quaife, T., and Marchant, R.: Terrestrial ecosystems from space: a review of earth observation products for macroecology applications, Global Ecology and Biogeography, 21, 603–624, 2012.

Polykretis, C., Grillakis, M. G., and Alexakis, D. D.: Exploring the impact of various spectral indices on land cover change detection using change vector analysis: A case study of Crete Island, Greece, Remote Sensing, 12, 319, 2020.

Rußwurm, M., Pelletier, C., Zollner, M., Lefèvre, S., and Körner, M.: BreizhCrops: A Time Series Dataset for Crop Type Mapping, arXiv preprint arXiv:1905.11893, 2019.

Safonova, A., Guirado, E., Maglinets, Y., Alcaraz-Segura, D., and Tabik, S.: Olive Tree Biovolume from UAV Multi-Resolution Image Segmentation with Mask R-CNN, Sensors, 21, 1617, 2021.

Townshend, J., Justice, C., Li, W., Gurney, C., and McManus, J.: Global land cover classification by remote sensing: present capabilities and future possibilities, Remote Sensing of Environment, 35, 243–255, 1991.

Tsendbazar, N., De Bruin, S., and Herold, M.: Assessing global land cover reference datasets for different user communities, ISPRS Journal of Photogrammetry and Remote Sensing, 103, 93–114, 2015a.

Tsendbazar, N.-E., De Bruin, S., Fritz, S., and Herold, M.: Spatial accuracy assessment and integration of global land cover datasets, Remote Sensing, 7, 15 804–15 821, 2015b.

Tsendbazar, N.-E., de Bruin, S., Mora, B., Schouten, L., and Herold, M.: Comparative assessment of thematic accuracy of GLC maps for specific applications using existing reference data, International Journal of Applied Earth Observation and Geoinformation, 44, 124–135, 2016.

Tuanmu, M.-N. and Jetz, W.: A global 1-km consensus land-cover product for biodiversity and ecosystem modelling, Global Ecology and Biogeography, 23, 1031–1045, 2014.

Van Etten, A., Hogan, D., Martinez-Manso, J., Shermeyer, J., Weir, N., and Lewis, R.: The Multi-Temporal Urban Development SpaceNet Dataset, arXiv preprint arXiv:2102.04420, 2021.

Vancutsem, C., Marinho, E., Kayitakire, F., See, L., and Fritz, S.: Harmonizing and combining existing land cover/land use datasets for cropland area monitoring at the African continental scale, Remote Sensing, 5, 19–41, 2013.

Verburg, P. H., Van De Steeg, J., Veldkamp, A., and Willemen, L.: From land cover change to land function dynamics: a major challenge to improve land characterization, Journal of environmental management, 90, 1327–1335, 2009.

Virnodkar, S. S., Pachghare, V. K., Patil, V., and Jha, S. K.: CaneSat dataset to leverage convolutional neural networks for sugarcane classification from Sentinel-2, Journal of King Saud University-Computer and Information Sciences, 2020.

Xiong, J., Thenkabail, P. S., Gumma, M. K., Teluguntla, P., Poehnelt, J., Congalton, R. G., Yadav, K., and Thau, D.: Automated cropland mapping of continental Africa using Google Earth Engine cloud computing, ISPRS Journal of Photogrammetry and Remote Sensing, 126, 225–244, 2017.

Xu, P., Herold, M., Tsendbazar, N.-E., and Clevers, J. G.: Towards a comprehensive and consistent global aquatic land cover characterization framework addressing multiple user needs, Remote Sensing of Environment, 250, 112 034, 2020.

Yirsaw, E., Wu, W., Shi, X., Temesgen, H., and Bekele, B.: Land use/land cover change modeling and the prediction of subsequent changes in ecosystem service values in a coastal area of China, the Su-Xi-Chang Region, Sustainability, 9, 1204, 2017.

https://doi.org/10.5194/essd-2021-253
Preprint. Discussion started: 12 October 2021

Zhang, C., Pan, X., Li, H., Gardiner, A., Sargent, I., Hare, J., and Atkinson, P. M.: A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification, ISPRS Journal of Photogrammetry and Remote Sensing, 140, 133–144, 2018a.

Zhang, C., Ye, Y., Fang, X., Li, H., and Wei, X.: Synergistic Modern Global 1 Km Cropland Dataset Derived from Multi-Sets of Land Cover Products, Remote Sensing, 11, 2250, 2019.

Zhang, Q., Yuan, Q., Zeng, C., Li, X., and Wei, Y.: Missing data reconstruction in remote sensing image with a unified spatial–temporal–spectral deep convolutional neural network, IEEE Transactions on Geoscience and Remote Sensing, 56, 4274–4288, 2018b.

Zhao, W. and Du, S.: Learning multiscale and deep representations for classifying remotely sensed imagery, ISPRS Journal of Photogrammetry and Remote Sensing, 113, 155–165, 2016.

Zhao, W., Guo, Z., Yue, J., Zhang, X., and Luo, L.: On combining multiscale deep learning features for the classification of hyperspectral remote sensing imagery, International Journal of Remote Sensing, 36, 3368–3379, 2015.

Zimmer-Gembeck, M. J. and Helfand, M.: Ten years of longitudinal research on US adolescent sexual behavior: Developmental correlates of sexual intercourse, and the importance of age, gender and ethnic background, Developmental review, 28, 153–224, 2008.