i. Title: GLOBAL LAND-USE AND LAND-COVER DATA FOR ECOLOGISTS: HISTORICAL, CURRENT AND FUTURE SCENARIOS

ii. Running title: GLOBAL LULC DATA

iii. Authors: MARIANA M. VALE [1, 2, 3], MATHEUS S. LIMA-RIBEIRO [3, 4] AND TAINÁ C. ROCHA [3,5 *]

IV: Author affiliations:

[1] Ecology Department, Federal University of Rio de Janeiro, Brazil

[2] Brazilian Research Network on Global Climate Change (Rede Clima), São José dos Campos, Brazil

[3] National Institute of Science and Technology on Ecology (INCT EECBio), Evolution and Biodiversity Conservation, Federal University of Goiás, Goiânia, Brazil

[4] Biodiversity Department, Federal University of Jataí (UFJ), Goiás, Brazil

[5] Botanical Garden Research Institute of Rio de Janeiro, Rio de Janeiro, Brazil.

v. Corresponding Authors contact details: taina013@gmail.com

vi. Author contributions: MV and TR designed the study, TR carried out the analyses, MV, TR and ML-R wrote the paper.

vii. Acknowledgements:

This initiative was possible due to the high-quality data that is maintained and made publicly available by Land-Use Harmonization (https://luh.umd.edu/). We are grateful to Ritvik Sahajpa for helping in the early stages of data acquisition and transformation. This study was developed in the context of the the National Institute for Science and Technology in Ecology, Evolution and Conservation of Biodiversity (INCT EECBio, CNPq Grant n 465610|2014-5, FAPEG 201810267000023) and the Brazilian Network on Global Climate Change Research (Rede CLIMA). MMV and ML-R received support from the National Council for Scientific and Technological Development (CNPq Grants no. 304309/2018-4 and 301514/2019-4, respectively) and TR from CAPES (Coordination for the Improvement of Higher Education Personnel - Grant No. 88887.373031/2019-00).

viii. Data accessibility:

All files are freely available online in ecoClimate database: https://www.ecoclimate.org/

ix. Conflict of interest

The authors have declared that no competing interests exist.

Abstract:

Land-use land-cover (LULC) data are important predictors of species occurrence and biodiversity threat. Although there are LULC datasets available for ecologists under current conditions, there is a lack of such data under historical and future climatic conditions. This hinders, for example, projecting niche and distribution models under global change scenarios at different times. The Land Use Harmonization Project (LUH2) is a global terrestrial dataset at 0.25^{o} spatial resolution that provides LULC data from 850 to 2300 for 12 LULC state classes. The dataset, however, is compressed in a file format (NetCDF) that is incompatible with most ecological analysis and intractable for most ecologists. Here we selected and transformed the LUH2 data in order to make it more useful for ecological studies. We provide LULC for every year from 850 to 2100, with data from 2015 on provided under two Shared Socioeconomic Pathways (SSP2 and SSP5). We provide two types of file for each year: separate files with continuous values for each of the 12 LULC state classes, and a single categorical file with all state classes combined. To create the categorical layer, we assigned the state with the highest value in a given pixel among the 12 continuous data. The final dataset provides LULC data for 1251 years that will be of interest for macroecology, ecological niche modeling, global change analysis, and other applications in ecology and conservation. We also provide a description of LUH2 prediction of future LULC change through time.

keywords: Conservation biogeography, ecological niche modelling, macroecology, CMIP6, climate change, deforestation

INTRODUCTION

Changes in land cover derived from human-mediated land use changes, particularly conversion of native vegetation to agriculture, is still among the greatest threats to biodiversity (Maxwell et al. 2016). Mapping land-use land-cover (LULC) changes through time is, therefore, important and desirable to predict these threats and propose effective conservation policies (Jetz et al. 2007). LULC is also an important predictor of species' occurrence and, thus extensively used in ecological studies (Eyringet al. 2016; Ruiz-Benito et al. 2020; Sobral-Souza et al. 2021). There are several LULC datasets available for ecological studies at a global scale under current conditions, such as the Global Land Survey, the 30 Meter Global Land Cover, and the GlobeLand30 (Gutman et al. 2013; Pengra et al. 2015; Brovelli et al. 2015), as well as the near historical period, such as the ESA Climate Change Initiative (1992 to 2015), the Finer Resolution Observation, Monitoring of Global Land Cover (1984 to 2011) (Hollmann et al. 2013; Gong et al. 2013) and GCAM (2015- 2100) (Chen et al. 2020). These datasets are usually available in standard Geographic Information System (GIS) formats (e.g. TIF or KMZ), routinely used by landscape ecologists, macroecologists and biogeographers. However, there is an important gap of historical LULC data covering pre-industrial periods (i.e. older than 1700) and, perhaps more importantly, projections of LULC changes into the future. The absence of compatible dataset across past, present and future scenarios, for example, hinders the use of LULC predictors in projections of ecological niche and species distribution models throughout the time and hamper global change analyses (Escobar et al. 2018).

A recent and robust LULC dataset for ecologists is the Land-Use Harmonization project (https://luh.umd.edu/data.shtml, Hurtt et al. 2006, 2011, 2020). This dataset is part of the Coupled Model Intercomparison Project (CMIP), which coordinates modeling experiments worldwide used by the Intergovernmental Panel on Climate Change (IPCC) (Eyring et al. 2016). The data is an input to Earth System Models (ESMs) to estimate the combined effects of human activities on the carbon-climate system. Currently, CMIP datasets are available in NetCDF format, a quite complex file format for most ecologists. This CMIP LULC has rarely been used in ecological studies, as opposed to CMIP's climate data already simplified for ecologists on standard GIS formats in widely used datasets such as WorldClim (https://www.worldclim.org/, Fick and Hijmans 2017) and ecoClimate (https://www.ecoclimate.org/, Lima-Ribeiro et al. 2015).

In order to make the global Land-Use Harmonization data more accessible and readily usable, here we filtered, combined and transformed it in standard GIS formats, making the dataset accessible for the ecologist with standard GIS skills. Besides providing the Land-Use Harmonization data in regular GIS format at yearly temporal resolution covering 1251 years of past, present and future (from 850 to 2100), we also derived new data based on the existing dataset.

METHODS

The Land-Use Harmonization project (LUH2) provides LULC data from 850 to 2300 at 0.25^{o} spatial resolution (ca. 30 km). The first generation of models (LUH1, Hurtt et al. 2006, 2011) made future land-use land-cover projections under CMIP5's Representative Concentration Pathways greenhouse gas scenarios (RCPs, see Vuuren et al. 2011), and the current generation of models (LUH2, Hurtt et al. 2020) makes projection under CMIP6's Shared Socioeconomic Pathways greenhouse gas scenarios (SSP, see Popp et al. 2017). Both provide data on 12 lan-use land-cover state classes: forested primary land (primf), non-forested primary land (primn), potentially forested secondary land (secdf), potentially non-forested secondary land (secdn), managed pasture (pastr), rangeland (range), urban land (urban), C3 annual crops (c3ann), C3 perennial crops (c3per), C4 annual crops (c4ann), C4 perennial crops (c4per), C3 nitrogen-fixing crops (c3nfx). The "forested" and "non- forested" land-use states are defined on the basis of the aboveground standing stock of natural cover; where "primary" are lands previously undisturbed by human activities, and "secondary" are lands previously disturbed by human activities and currently recovered or in process of recovering of their native aspects (see Hurtt et al. 2006, 2011, 2020 for more details). They were computed using an accounting-based method that tracks the fractional state of the land surface in each grid cell as a function of the land surface at the previous time step through historical data. Because it deals with a large and undetermined system, the approach was to solve the system for every grid cell at each time step, constraining with several inputs including land-use maps, crop type and rotation rates, shifting cultivation rates, agriculture management, wood harvest, forest transitions and potential biomass and biomass recovery rates (see Fig. S1 in the Supplementary Material for details).

We downloaded the 12 land-use land-cover state layers (state.nc) provided in Network Common Data Form (NetCDF) from the Land-Use Harmonization Project

(LUH2, https://luh.umd.edu/data.shtml). To manipulate the NetCDF files, we used the ncdf4 version 1.16 and rgdal packages in R environment version 4.2 (R Core Team 2020, Pierce 2019; Hijmans et al. 2020; Bivand et al. 2021). We also used the Panoply software version 4.8 for quick visualization of the original data (states.nc) (Schmunk, 2017 https://www.giss.nasa.gov/tools/panoply/).

We created two sets of files for each year, the continuous "state-files" and the categorical "LULC-files" (Fig.1, Fig.2 and Fig. S2 of supplemental material). The state-files are the same data provided in the original LUH2 dataset (states.nc), transformed into Tag Image File Format (TIFF) and standardized for ranging from 0 to 1. We built the new LULC-files, also in TIFF format, assigning the highest value among the 12 available states to each pixel. The LULC-files are categorical data ranging from 1 to 12, where each number represents each one of the 12 existing states in the dataset (Table S1 in Supplementary Material). We generated states-files and LULC-files for each single year from 850 to 2100 for two greenhouse gas scenarios: an optimistic (SSP2) and a pessimistic (SSP5) (see Fig. S2 in Supplementary Material for the workflow to create state files and LULC-files).

We performed an accuracy assessment of our classification for the LULC-files following Olofsson et al.'s (2014) good practices, for the all continents together and for Newton and Dale's (2001) zoogeographic regions separately. We compared our classified LULC-file for the year 2000 with that of the Global Land Cover SHARE (GLC-SHARE) data, used as the ground truth reference data in the accuracy assessment. The GLC-SHARE was built from a combination of "best available" high resolution national, regional and/or sub-national land cover databases (Latham et al. 2014), and has a better spatial resolution (1 km) than the LUH2 (30 km). The great advantage of GLC-SHARE is to combine a global extent with land cover information at regional level obtained by spatial and multi-temporal source data (Latham et al. 2014). Furthermore, GLC-SHARE has 11 classes that are very similar with those from the LUH2 database: artificial surfaces (01), cropland (02), grassland (03), tree covered areas (04), shrubs covered areas (05), herbaceous vegetation, aquatic or regularly flooded (06), mangroves (07), sparse vegetation (08), bare soil (09), snow and glaciers (10), and water bodies (11) . To make the two datasets comparable, we reclassified LUH2 and GLS-SHARE to the following classes: forest, crops, open areas and urban (Fig. 3, Table S1 in Supplementary Material). We also masked-out ice and water areas from GLS-SHARE, as they do not have an equivalent in the LUH2 dataset. Thus, Greenland was removed from analysis and is not present in the LULC-files. The accuracy assessment was performed in QGIS 3.10, we computed a confusion matrix, quantified the commission and omission errors for each class (ranging from 0 to 100%), and then the Kappa Index for overall accuracy (ranging from 0 to 1).

The entire resulting dataset will be made freely available for download at the ecoClimate repository (https://www.ecoclimate.org/), an open database of processed environmental data in a suitable resolution and user-friendly format for macroecological and biogeographical studies (Lima-Ribeiro et al. 2015).

RESULTS

We generated 17.394 files, 16.056 of which are the LUH2 original (continuous data) states files transformed into TIFF (Fig. 1), and the other 1.338 are new (categorical data) files created by combining the 12 states files (Fig. 2). The LULC-files had good results for some geographic regions and land-use land-cover classes, but not for all (Fig. 3, Table 1). The overall accuracy (Kappa index) at the global scale was 0.51, with good overall accuracy (> 0.55) for Indomalayan and Australasia zoogeographic regions, moderate overall accuracy (ca. 0.50) for the Neartic and Paleartic, and fair overall accuracy (> 0.50) for the Neotropic and Afrotropic (Table 1). The accuracy for land-use land-cover classes showed a pattern. There was good accuracy for crops, moderate accuracy for forest and urban areas, and fair accuracy for open areas at the global scale, with similar results for zoogeographic regions, where crops showed the best accuracy (except for Australasia where forest had the best Kappa Index), and open and urban areas showing the worst (Table 1).

DISCUSSION

This data paper is another contribution to ecology in the big data era, aiming at making the Land-Use Harmonization project data more accessible to ecologists. Our main goal here was to provide support for ecology and biodiversity studies, and disseminate the use of open datasets and open-source tools for a science of quality, transparency and inclusion. We contributed not only by transforming the data into standard GIS file format, but also by providing new categorical data on land-use land-cover through time. This is important given that several studies have highlighted the need for more integration between climate and land-use change within biodiversity science (Titeux et al. 2017, Albert et al. 2020, Hanna et al. 2020). The dataset we provide here has fine time resolution, with 1251 years of data from 850 to 2100, and considerable spatial resolution (0.25°) for a global dataset.

The state-files may be particularly useful as predictors in ecological niche modeling (Peterson et al. 2011). The forested primary land state, for example, can be used to model the distribution of forest-dependent species, as is the case, of birds in the Brazilian Atlantic Forest biodiversity hotspot (Vale et al. 2018), predicted to lose more forests in the future. This data has the advantage of being represented in continuous values, as opposed to most discrete land cover data (e.g. all datasets cited in this paper), overcoming the shortcoming of using categorical data as layers in ecological niche modeling (Peterson 2001). More importantly, it allows for the use of land cover data in projections of species distribution under future climate change scenarios.

The LULC-files may be particularly useful for ecological and biogeographical studies focused on past or future conditions, given that the datasets currently available in standard GIS format are restricted to specific periods in the last few decades. In the accuracy assessment, moderate and fair accuracy values for some classes may be due to the reclassification of the data from 12 to four classes, and much greater spatial resolution of the reference map (1 km) as compared to the our LUH2 data (30 km). In any case, we suggest that the user consult Table 1 for general class accuracy at different zoogeographic regions when performing regional analysis.

The Land-use Harmonized project shows important changes in LULC through time (Fig. 1 and 2), although with no noticeable difference between greenhouse gas scenarios within the same year (Fig. 4). It predicts a pronounced decrease in primary forest, and an equally pronounced increase in secondary forest and non-forest lands (Fig. 4). The decrease in primary forest is particularly noticeable in the Amazon, the Brazilian Atlantic Forest, the Congo Basin and the boreal forests (Fig. 1), coupled with an increase in secondary forest in these regions (Fig. 2). A predicted increase in C4 annual, C3 nitrogen-fixing and C3 perennial crops is especially pronounced in the Brazilian Atlantic Forest and sub-Saharan Africa (Fig. 2). These crops will apparently replace managed pastures in Africa's Great Lakes region. Finally, there is also a specially pronounced predicted decrease in non-forested primary land (Fig. 4), especially in northern Africa and in the Horn of Africa (Fig. 2).

We hope that the data provided here can foster the use of land-use land-cover data in ecology and biogeography. It will be freely available for download at ecoClimate repository (https://www.ecoclimate.org/).

REFERENCES

Albert, C. H., M. Hervé, M. Fader, A. Bondeau, A., Leriche, A.C. Monnet, and W. Cramer. 2020. What ecologists should know before using land use/cover change projections for biodiversity and ecosystem service assessments. *Reg. Environ. Change* 20:1–12. doi:10.1007/s10113-020-01675-w.

Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Sumner, M., Hijmans, R., Rouault, E. and Bivand, M.R. 2015. rgdal: Bindings for the Geospatial Data Abstraction Library. R package version 1.5-12.

Brovelli, M. A., M. E. Molinari, E. Hussein, J. Chen, and R. Li. 2015. The first comprehensive accuracy assessment of global and 30 at a national level: Methodology and results". *Remote Sensing* 7:4191–4212. doi:10.3390/rs70404191.

Chen, M., C. R. Vernon, N. T. Graham, M. Hejazi, M. Huang, Y. Cheng, and K. Calvin. 2020. Global land use for 2015–2100 at 0.05° resolution under diverse socioeconomic and climate scenarios. *Sci. Data* 7:1–11. doi:10.1038/s41597-020-00669-x.

Escobar, L. E., H. Qiao, J. Cabello, and A. T. Peterson. 2018. Ecological niche modeling reexamined: A case study with the Darwin's fox. *Ecol. Evol.* 8:4757–4770. doi:10.1002/ece3.4014 .

Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor. 2016. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* 9:1937–1958. doi:10.5194/gmd-9-1937-2016.

Gong, P., J. Wang, L. Yu, Y. Zhao, Y. Zhao, L. Liang, Z. Niu, X. Huang, H. Fu, S. Liu, C. Li, X. Li, W. Fu, C. Liu, Y. Xu, X. Wang, Q. Cheng, L. Hu, W. Yao, H. Zhang, P. Zhu, Z. Zhao, H. Zhang, Y. Zheng, L. Ji, Y. Zhang, H. Chen, A. Yan, J. Guo, L. Yu, L. Wang, X. Liu, T. Shi, M. Zhu, Y. Chen, G. Yang, P. Tang, B. Xu, C. Giri, N. Clinton, Z. Zhu, J. Chen, and J. Chen. 2013. Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* 34:2607–2654. doi:10.1080/01431161.2012.748992.

Gutman, G., C. Huang, G. Chander, P. Noojipady, and J.G. Masek. 2013. Assessment of the NASA-USGS Global Land Survey (GLS) datasets. *Remote Sens. Environ.* 134:249–265. doi:10.1016/j.rse.2013.02.026.

Hanna, D. E. L., C. Raudsepp-Hearne, and E. M. Bennett. 2020. Effects of land use, cover, and protection on stream and riparian ecosystem services and biodiversity. *Conserv. Biol.* 34: 244–255. doi:10.1111/cobi.13348.

Fick, S. E., and Hijmans, R. J. 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* 37: 4302–4315. doi:10.1002/joc.5086.

Hijmans, R. J. 2020. raster: geographic data analysis and modeling. R package version 3.3.13.

Hollmann, R., C. J. Merchant, R. Saunders, C. Downy, M. Buchwitz, A. Cazenave, E. Chuvieco, P. Defourny, G. de Leeuw, R. Forsberg, T. Holzer-Popp, F. Paul, S. Sandven, S. Sathyendranath, M. van Roozendael, and W. Wagner. 2013. The ESA climate change initiative: Satellite data records for essential climate variables. *Bull. Am. Meteorol. Soc.* 94:1541–1552. doi:10.1175/BAMS-D-11-00254.1.

Hurtt, G. C., L. P. Chini, S. Frolking, R. A. Betts, J. Feddema, G. Fischer, J. P. Fisk, K. Hibbard, R. A. Houghton, A. Janetos, C. D. Jones, G. Kindermann, T. Kinoshita, Kees Klein Goldewijk, K. Riahi, E. Shevliakova, S. Smith, E. Stehfest, A. Thomson, P. Thornton, D. P. van Vuuren, and Y. P. Wang. 2011. Harmonization of land-use scenarios for the period 1500-2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Clim. Change* 109:117–161. doi:10.1007/s10584-011-0153-2.7.

Hurtt, G. C., L. Chini, R. Sahajpal, S. Frolking, B. L. Bodirsky, K. Calvin, J. C. Doelman, J. Fisk, S. Fujimori, K. K. Goldewijk, T. Hasegawa, P. Havlik, A. Heinimann, F. Humpenöder, J. Jungclaus, J. Kaplan, J. Kennedy, T. Krisztin, D. Lawrence, P. Lawrence, L. Ma, O. Mertz, J. Pongratz, A. Popp, B. Poulter, K. Riahi, E.

Shevliakova, E. Stehfest, P. Thornton, F. N. Tubiello, D. P. van Vuuren, and X. Zhang. 2020. Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6. *Geosci. Model Dev.* 13:5425–5464. doi:10.5194/gmd-13-5425-2020.

Hurtt, G. C., S. Frolking, M. G. Fearon, B. Moore, E. Shevliakova, S. Malyshev, S. W. Pacala, and R. A. Houghton. 2006. The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands. *Glob. Change Biol.* 12:1208–1229. doi:10.1111/j.1365-2486.2006.01150.x.

Jetz, W., D. S. Wilcove, and A. P. Dobson. 2007. Projected impacts of climate and land-use change on the global diversity of birds. *PLOS Biol.* 5:1211–1219. doi:10.1371/journal.pbio.0050157.

Latham, J., R. Cumani, I. Rosati, and M. Bloise. 2014. Global land cover SHARE (GLC-SHARE). FAO: Rome, Italy. Version 1.0-2014.

Lima-Ribeiro, M. S. 2015. EcoClimate: a database of climate data from multiple models for past, present, and future for macroecologists and biogeographers. *Biodivers. Inform.* 10:0–21. doi:10.17161/bi.v10i0.4955.

Maxwell, S. L., R. A. Fuller, T. M. Brooks, and J. E.M. Watson. 2016. Biodiversity: the ravages of guns, nets and bulldozers. *Nature* 536:143–145. doi:10.1038/536143a.

Newton I, and L. Dale. 2001. A comparative analysis of the avifaunas of different zoogeographical regions. *J. Zool.* 254:207–218. doi:10.1017/S0952836901000723

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

Peterson, A. T. 2001. Predicting species' geographic distributions based on ecological niche modeling. Condor 103:599–605. doi:10.1093/condor/103.3.599.

Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, Araújo, M. B. 2011. Ecological Niches and Geographic Distributions. Princeton University Press. JSTOR .

Pierce, D., and M. D. Pierce. 2019. ncdf4: Interface to Unidata netCDF. R package version 1.17.

Popp, A., K. Calvin, S. Fujimori, P. Havlik, F. Humpenöder, E. Stehfest, B. L. Bodirskyah, J. P. Dietrich, J. C. Doelmann, M. Gusti, T. Hasegaw, P. Kyl, M. Obersteiner, A. Tabeau, K. Takahashi, H. Valin, S. Waldhoff, I. Weindl, … D. P. V. Vuuren. 2017. Land-use futures in the shared socio-economic pathways. *Glob. Environ. Change* 42:331–345. doi:10.1016/j.gloenvcha.2016.10.002.

R Core Team. 2020. R: a language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Ruiz-Benito, P., G. Vacchiano, E. R. Lines, C. P. O. Reyer, S. Ratcliffe, X. Morin, F. Hartig, A. Mäkelä, R. Yousefpour, J. E. Chaves, A. Palacios-Orueta, M. Benito-Garzón, C. Morales-Molino, J. J. Camarero, A. S. Jump, J. Kattge, A. Lehtonen, A. Ibrom, and M. A. Zavala. 2020. Available and missing data to model impact of climate change on European forests. *Ecol. Model.* 416:108870. doi:10.1016/j.ecolmodel.2019.108870.

Schmunk, R. B. 2017. Panoply: netCDF, HDF and GRIB Data Viewer. NASA Goddard Institute for Space Studies. Version 4.7.

Sobral-Souza, T., J. P. Santos, M. E. Maldaner, M. S. Lima-Ribeiro, , and M. C. Ribeiro. 2021. EcoLand: A multiscale niche modelling framework to improve predictions on biodiversity and conservation. Perspectives in Ecology and Conservation (in press). doi: 10.1016/j.pecon.2021.03.008

Taylor, K. E., R. J. Stouffer, and G. A. Meehl. 2012. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* 93:485–498. doi:10.1175/BAMS-D-11-00094.1.

Titeux, N., K. Henle, J. B. Mihoub, A. Regos, I. R. Geijzen-Dorffer, W. Cramer, P. H. Verburg, and L. Brotons. 2017. Global scenarios for biodiversity need to better integrate climate and land use change. *Divers. Distrib.* 23:1231–1234. doi:10.1111/ddi.12624.

Vale, M. M., L. Tourinho, M. L. Lorini, H. Rajão, and M. S. L. Figueiredo. 2018. Endemic birds of the Atlantic Forest: traits, conservation status, and patterns of biodiversity. *J. Field Ornithol.* 89:193–206. doi:10.1111/jofo.12256.

Vuuren, D. P. van, J. A. Edmonds, M. Kainuma, K. Riahi, and J. Weyant. 2011. A special issue on the RCPs. *Clim. Change* 109:1–4. doi:10.1007/s10584-011-0157-y.

TABLES AND FIGURES

Table 1. Classification accuracy. The overall accuracy (Kappa Index) for each geographic region presented in parenthesis (in bold) by the side of the region's name. Land-use land-cover classes' Kappa Index (k) are presented in bold, as well as commission error (EC) and omission error (EO) percentage rates (%). Geographic regions: All continents together (Global), Indomalayan (Indo.), Neartic, Paleartic, Australasian (Austral.), Neotropical (Neotrop.) and Afrotropical (Afrotrop.). Data in the table is arranged in decreasing order of Kappa Index for regions and land-use land-cover classes.

	Crops			Forest			Urban Areas			Open Areas		
region	**k**	EC	EO	**k**	EC	EO	**k**	EC	EO	**k**	EC	EO
Global (**.51**)	**.76**	20.3	52.6	**.54**	29.5	33.1	**.55**	44.4	86.8	**.42**	28.7	17.4
Indomalayan (**.57**)	**.86**	8.4	37.8	**.50**	31.0	17.0	**.46**	53.3	87.9	**.45**	42.2	28.1
Australasian (**.57**)	**.79**	19.0	43.3	**.86**	10.4	53.7	**.66**	33.3	81.0	**.41**	18.5	2.1
Neartic (**.52**)	**.78**	19.4	41.5	**.40**	38.7	15.9	**.68**	18.6	79.5	**.60**	18.6	32.8
Paleartic (**.51**)	**.71**	24.3	43.4	**.52**	31.6	35.1	**.40**	59.1	94.1	**.44**	26.5	18.4
Neotropical (**.42**)	**.90**	7.9	84.7	**.72**	12.4	32.8	**.25**	74.0	66.7	**.25**	52.2	11.3
Afrotropical (**.35**)	**.71**	25.6	84.5	**.61**	28.2	58.0	**1.0**	0	95.2	**.23**	29.3	5.9

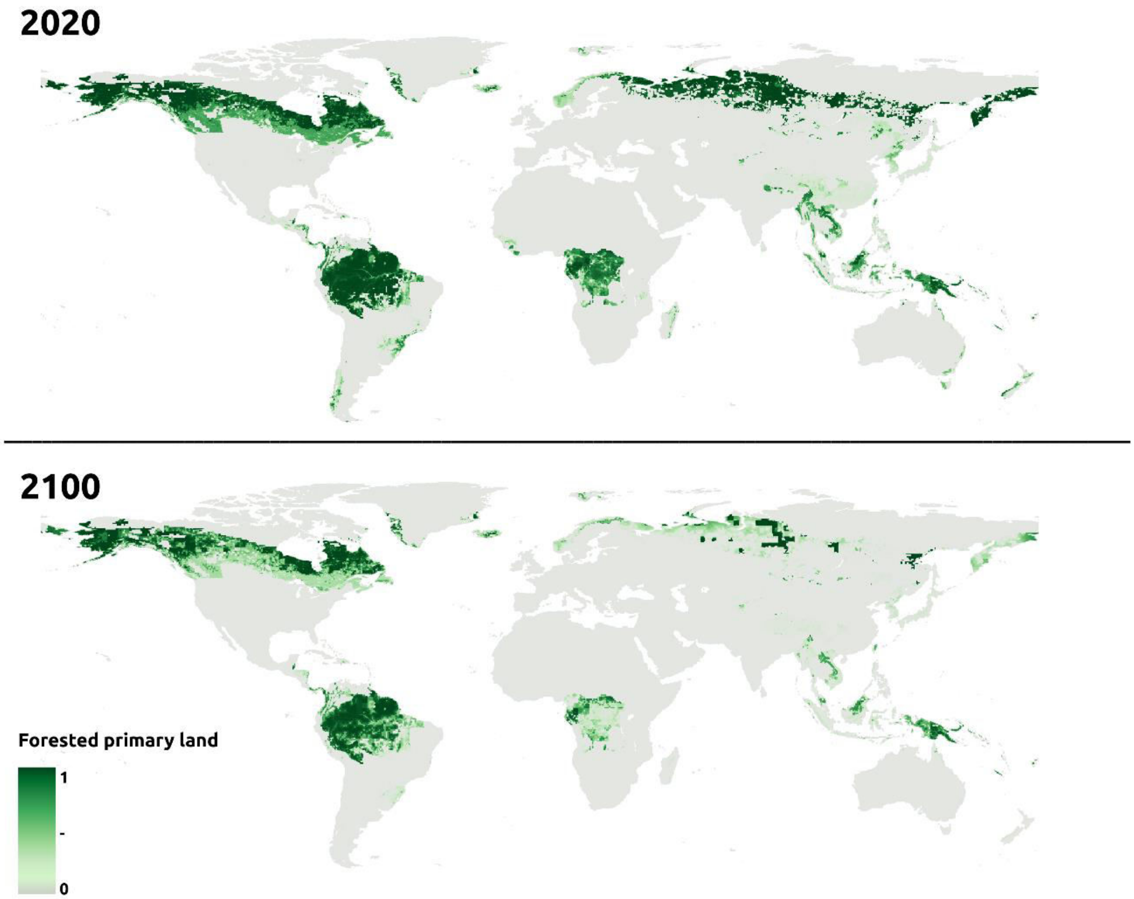

Figure 1. Example of state-files data. Continuous forested primary land state for 2020 (top) and 2100 (bottom) under SSP5 greenhouse gas scenario are shown, as originally provided by the Land-Use Harmonization (LUH2) project. State values range from 0 to 1, roughly representing the likelihood that the pixel is occupied by the land-use land-cover class depicted in the file. All other state-files have the same structure.

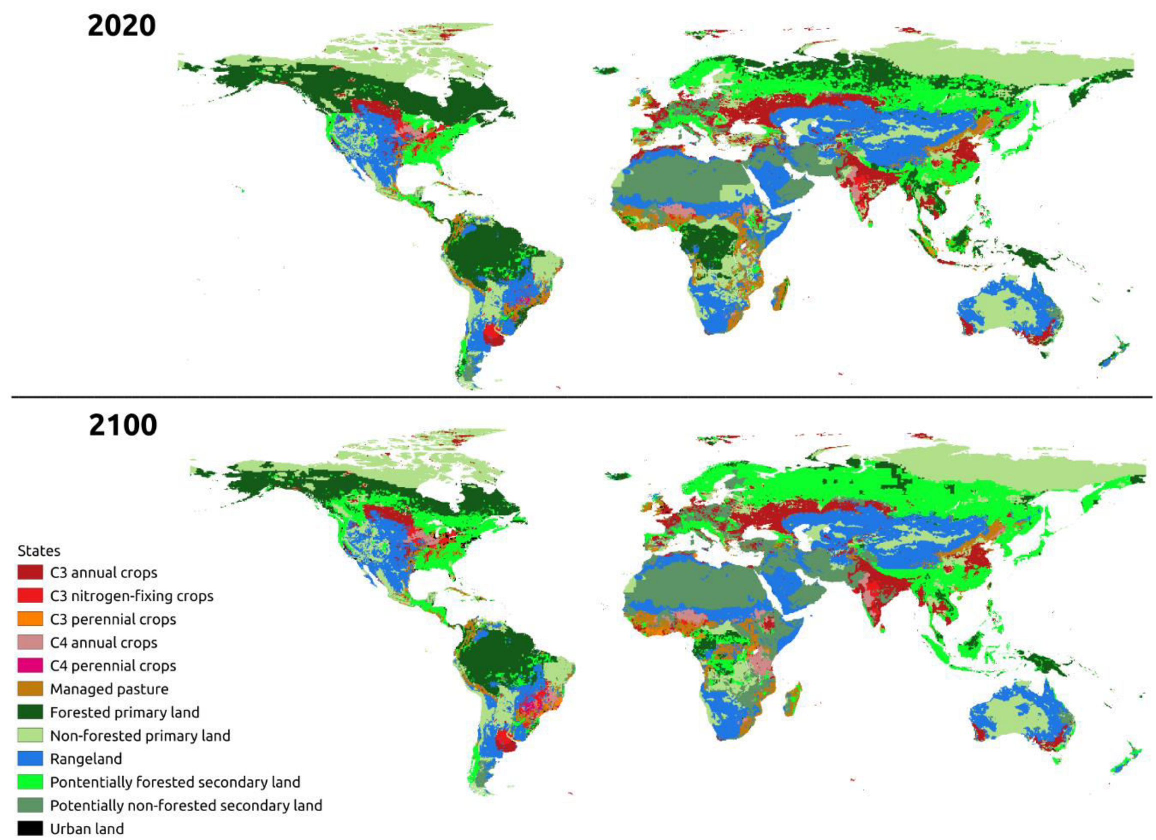

Figure 2. Example of LULC-files data. Categorical LULC for 2020 (top) and 2100 under SSP5 greenhouse gas scenarios (bottom) are shown, as a result of the combination of the 12 LULC original state-files into a single file.

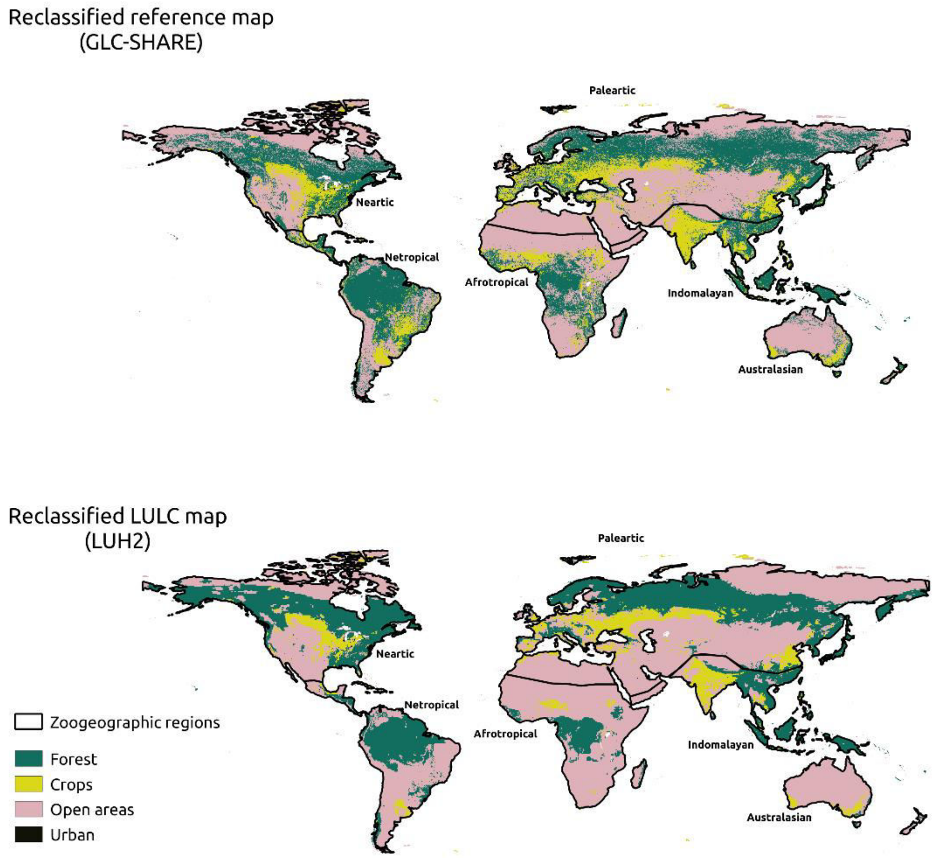

Figure 3. Data used in the accuracy assessment of LULC-files. The accuracy of the classification of the LULC-file (bottom) assessed using the GLC SHARE as reference data (top). To make the two datasets comparable, both were reclassified to the five land-use land-cover states for the year 2000 (see Table 1 for reclassification scheme). We run the analysis for the whole world, as well as for six zoogeographic regions separately: Nearctic, Neotropical, Palearctic, Afrotropical, Indomalayan and Australasian.

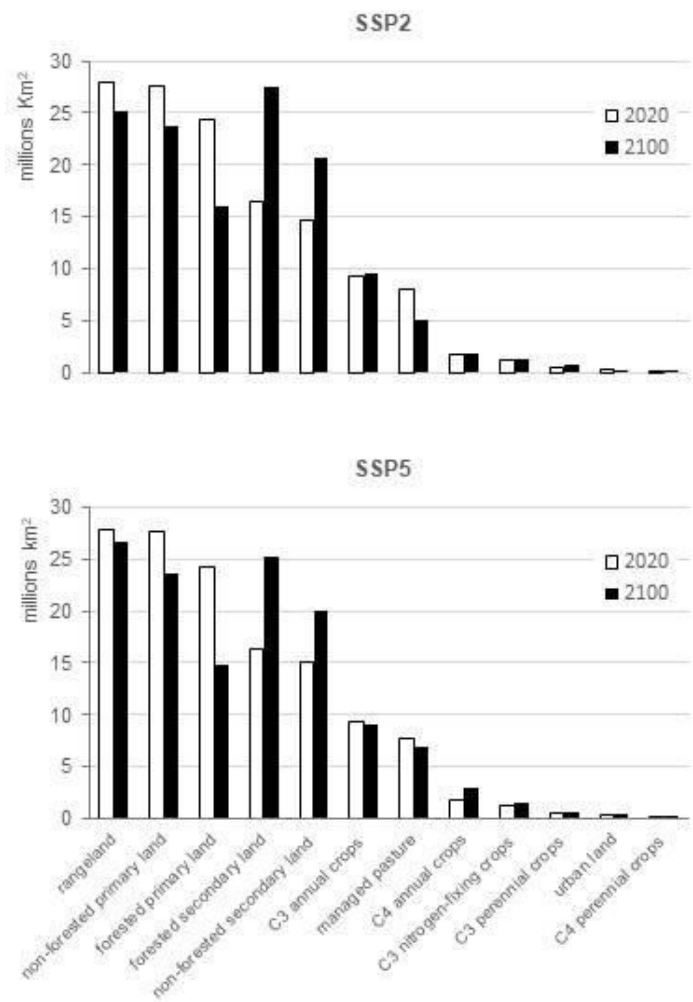

Figure 4. Land-use land cover comparison among years and scenarios. Data for the LULC-files for year 2020 and 2100 for the optimistic (SSP2, top) and pessimistic (SSP5, bottom) greenhouse gas scenarios, arranged in decreasing order of class area in 2020.