,
Horizon 2020 Societal challenge 5:
Climate action, environment, resource
efficiency and raw materials

VERIFY

Observation-based system for monitoring and verification of greenhouse gases

GA number 776810, RIA

Deliverable number (relative in WP)	**D3.3**
Deliverable name:	Final state of the art database
WP / WP number:	3
Delivery due date:	Month 38 (March 2021)
Actual date of submission:	Month 50 (09/03/2022)
Dissemination level:	Public
Lead beneficiary:	University of Aberdeen
Responsible	Matthias Kuhnert
Contributor(s):	Matthias Kuhnert, Pete Smith, Matthew McGrath, Philippe Peylin, Karina Winkler, Richard Fuchs, Martin Herold, Mart-Jan Schelhaas, Gert-Jan Nabuurs, Goulven G. Laruelle, Frank Deneter, Pierre Regnier, Are Olsen, Maximilian Reuter, Meike Becker
Internal reviewer:	Philippe Peylin, Pete Smith, Matthew McGrath

Changes with respect to the DoA
None
Dissemination and uptake **(Who will/could use this deliverable, within the project or outside the project?)**
The collected data are freely available (in some case registration is/will be necessary). In the project these data build the basis for model simulations in WP3 and WP4. The web-pages for data download are listed in section 3. This data forces the models in addition to being used for calibration and validation**.**
Short Summary of results (<250 words)
The state of the art database addresses the data demand by the different modeling groups in order to provide data for forcing, calibration and evaluation of the models. We were already able to summarize most relevant datasets in deliverable D3.1 (the first version of the database deliverable) and updated serval data sets for the second deliverable D3.2. Therefore, this new version only summarizes previous datasets. All details are reported in the core of the deliverable. All data are up-loaded on the VERIFY THREDDS server (accessible through http://verify.lsce.ipsl.fr/index.php/products). Section 3 of the deliverable summarizes how to access all datasets**.**
Evidence of accomplishment **(report, manuscript, web-link, other)**
All the datasets will be accessible though the VERIFY web site and the dedicated data-products page: http://verify.lsce.ipsl.fr/index.php/products Note that some of these data are password protected during a consolidation phase and thus only accessible to the VERIFY partners. Most data have been uploaded on the portal and are accessible through a catalogue and a THREDDS server (see section 3).

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

Version	Date	Description	Author (Organisation)
V0.1	21/03/2020	Creation/Writing	Matthias Kuhnert (University of Aberdeen)
V0.2	21/03/2020	Writing/Formatting/Delivery	Matthias Kuhnert, Pete Smith, Matthew McGrath, Philippe Peylin, Karina Winkler, Richard Fuchs, Martin Herold, Emanuele Lugato, Adrian Leip, Philippe Ciais, Mart-Jan Schelhaas, Gert-Jan Nabuurs, Frank Deneter, Ronny Lauerwald, Pierre Regnier, Are Olsen, Maximilian Reuter, Heike Becker
V1	09/03/2022	Formatting/Delivery on the Participant Portal	Matthias Kuhnert, Pete Smith, Matthew McGrath, Philippe Peylin, Géraud Moulas

Contents

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

1. Glossary

Abbreviation / Acronym	Description/meaning
CAPRI	Common Agricultural Policy Regional Impact
CORINE	Coordination of Information on the Environment
CRU	Climatic Research Unit
EMEP	The co-operative programme for monitoring and evaluation of the long-range transmission of air pollutants in Europe (unofficially 'European Monitoring and Evaluation Programme' = EMEP)
ESA	European Space Agency
ESDAC	European Soil Data Center
FAO	Food and Agriculture Organization
HILDA	Historic Land Dynamics Assessment
HWSD	Harmonized World Soil Database
LULC	Land use/Land Cover
RUSLE	Revised Universal Soil Loss Equation
THREDDS	Thematic Real-time Environmental Distributed Data Services
UERRA	Uncertainties in Ensembles of Regional ReAnalysis

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

2. Executive Summary

This deliverable updates the status of the state-of-the-art database in VERIFY. The objective of task T3.1 is to provide the database for process-based and statistical models. Additionally, with recent progress, it expands this scope to also provide data for the prior estimates of the top-down approaches. Most data were already collected and described for deliverables D3.1 and D3.2. The current deliverable summarises the available data in a table and describes improvements and extensions of data sets in detail up to the year 2020 where applicable.

A series of major improvements has been made in the past year. 1) The biomass data set is now complete. This data set includes aboveground and belowground biomass and carbon content, along with uncertainty estimates. 2) Climate data and land use data are updated for an extended period to cover the period until 2020 (climate) and 2019 (land use). Additionally, the land use data are improved by the Integration of new Earth Observation datasets/Land Cover classifications for LULC fraction maps. 3) Data on the coastal fluxes are updated, but not yet completed; a unified data set of merged ocean and coastal CO_2 fluxes will be provided as soon as the data for 2020 are completely available. A unified data sets avoids large differences in the transition area.

Since this is the final deliverable in the project on this issue, the data sets are complete and further changes will only improve the ocean/coastal CO_2 flux data or extend the coverage period until 2021. As soon as the data for 2020 are available the data will be aggregated and provided. As the merging of coastal and ocean CO_2 fluxes will be a new data set, there may be some challenges to overcome with the harmonization of methods.

Now, the provided land use/land cover data cover the period 1900-2019 with annual resolution. This is an exceptional data set, which addresses the demand of project partners for a longer temporal coverage in a high spatial resolution. The climate forcing data, essential for most models, have been extended for another year with a larger number of years coming from the updated ERA5-Land dataset. The data are available for modelling groups to run simulations through 2020. The new data sets of biomass and emissions at the European shelf seas addressed known gaps in the database. With these additions the database is complete and covers all relevant areas for greenhouse gas emission inventories. This database addresses the needs of partners in WP3 and WP4.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

3. Introduction

In this deliverable we report the final state of data collected for the VERIFY project. There are no plans to add more data types to this list. However, there might be still some improvements and changes to single data sets (e.g., additional forest data) or improvements on datasets which are already being worked on (e.g., combining coastal fluxes with ocean fluxes). Some of the data sets may also be extended to cover year 2020 to meet the goal of an operational system.

Almost all data sets have already been described in the former deliverables (D3.1 and D3.2). As was done in previous deliverables, Table 1 shows the complete list of datasets and contact person to get an overview about the available data. Detailed descriptions are listed on the VERIFY SharePoint platform (for VERIFY partners only). All data are provided to all project partners and will be made available to everyone (when the data is freely available) through a THREDDS server (see section 5). The datasets will mainly contribute to the simulation approaches in WP3 and WP4.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

Table 1 : Available dataset, from the internet (CORINE, HWSD and NO3/NH4 data) and provided by project participants.

Dataset	Name/ model	Inst.	Coverage	Resolution	Time frame	Contact in the project
Land use	HILDA[1]	KIT/WU	global	1 km	1900-2019	Richard Fuchs[a], Karina Winkler[b] Martin Herold[c]
Biomass	BIOMASS-CCI[1]	WU	global	0.1 °	2000, 2010, 2017	Martin Herold[c]
N-Deposition	EMEP model[2]	JRC		1 km	2010	Frank Dentener[d]
Erosion	RUSLE[3]	ESDAC	Europe	100 m	2015	Emanuele Lugato[e]
Soil data	LUCAS[4]	ESDAC	Europe	500 m/ 1 km	2009-2015	Emanuele Lugato[e]
Climate	C3S-ERA-5[5]	ECMWF	global	31 km	2008-current	Richard Engelen[f] Matt McGrath[p]
Climate	UERRA[6]	ECMWF	Europe	10 km	1961-2018	Richard Engelen[f] Matt McGrath[p]
Flux data	FLUXNET[7] network		global	sites	diverse	Dario Papale, Werner Kutsch
Fertiliser application rates	CAPRI	JRC	Europe	0.25°	2000-2012	Adrain Leip[g]
Management timing (crop)		UNIABDN	EU28	0.25°	2000-2015	Matthias Kuhnert[h]
Fresh water fluxes		ULB	Europe	0.1°	2016	Goulven G. Laruelle[i]
Ocean coastal fluxes	SOCAT[9]	UiB	Northern Europe	0.125°	1998-2018	Are Olsen[j], Meike Becker[o]
Forest management		WU	Europe	0.125°	2000-2015	M.-J. Schelhaas[k], G.-J. Nabuurs[l]
Grassland management	NO3/NH4 data[10]	CEA-LSCE	Europe	0.5°	1860-2012	Philippe Ciais[m]
atmospheric CO_2	FOCAL[8]	UBremen	global	2 km	2015-2016	Maximilian Reuter[n]
Fertilizer data		Nishina et al., 2017	global	0.5°	1961-2010	Matthias Kuhnert[h]

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

Email addresses:

a richard.fuchs@kit.edu
b karina.winkler@kit.edu
c martin.herold@wur.nl
d frank.dentener@ec.europa.eu
e Emanuele.LUGATO@ec.europa.eu
f richard.engelen@ecmwf.int
g Adrian.LEIP@ec.europa.eu
h Matthias.kuhnert@abdn.ac.uk
i Goulven.Gildas.Laruelle@ulb.ac.be
j are.olsen@uib.no
k martjan.schelhaas@wur.nl
l gert-jan.nabuurs@wur.nl
m philippe.ciais@lsce.ipsl.fr
n mreuter@iup.physik.uni-bremen.de
o meike.becker@uib.no
p matthew.mcgrath@lsce.ipsl.fr

Data for climate, soil, erosion, freshwater fluxes land use/land cover and atmospheric CO_2 have already been uploaded onto the THREDDS server (WP3-input dataset).

1 https://www.wur.nl/en/Research-Results/Chair-groups/Environmental-Sciences/Laboratory-of-Geo-information-Science-and-Remote-Sensing/Models/Hilda/HILDA-data-downloads.htm
2 http://webdab.emep.int/Unified_Model_Results/
3 https://esdac.jrc.ec.europa.eu/content/soil-erosion-water-rusle2015
4 https://esdac.jrc.ec.europa.eu/resource-type/european-soil-database-soil-properties
5 https://www.ecmwf.int/en/about/media-centre/science-blog/2017/era5-new-reanalysis-weather-and-climate-data
6 https://confluence.ecmwf.int//display/UER
7 https://fluxnet.fluxdata.org/login/?redirect_to=/data/download-data/
8 http://www.iup.uni-bremen.de/~mreuter/TN_XCO2-OCO2-FOCAL_v08.pdf
9 www.socat.info
10 https://doi.pangaea.de/10.1594/PANGAEA.861203

For some sections no progress or changes are included. This may be because the final product was already provided for previous deliverables or because the ongoing work does not yet show significant changes to the last deliverable.
Based on the demand and needs of the different groups, all required data are provided. Further datasets, which will be provided later or in an improved version (e.g., the combined coastal and ocean fluxes), will enhance the model simulation and evaluation, but do not block any progress for ongoing work.

4. Description of the different input and forcing datasets for WP3 activities

4.1. State of the art climate data

In 2020, the VERIFY project, through the combined efforts of the University of East Anglia, ECMWF, and the LSCE, successfully processed high-resolution meteorological forcing data from the ERA5-Land dataset at three-hour resolution across Europe for the years 2018 and 2019. As noted in D3.2, data production of the HARMONIE dataset from the UERRA project ceased in August 2019, which means we could not continue using it for VERIFY. We made the choice to pursue ERA5-Land as it was the only available dataset which satisfied the requirements of the VERIFY project: spatial resolution around 10 km, sub-daily resolution, temporal coverage back to as close to the beginning of the observed 20th -century warming as possible, and an operational status that guaranteed that data for the previous year would be available by April of the current year.

The years 2018 and 2019 were chosen as initial test year and due to the time constraints imposed by the Covid pandemic and the increased size of the data files (which first have to be downloaded at hourly resolution and then transformed into the 3h, daily, and monthly temporal resolutions used in the project). Integration of these two years along with the CRUHAR dataset from 1901-2018 was facilitated by re-alignment with the CRU observational dataset. This procedure changes the monthly means of each 0.5° pixel to match that of CRU observations; consequently, the regional monthly-averaged climate is identical, with variations only arising on the local (sub 0.5°) and sub-monthly levels.

In 2021, we are already well-along in the process. Due to the Covid pandemic, ECMWF has delayed publication of the full ERA5-Land dataset. Originally, the years 1951-2020 were set to be available by April 2021, which would have enabled us to redo the whole 1901-2020 meteorological forcing for the last round of bottom-up simulations in the VERIFY project in the summer of 2021. However, only the years 1981-2020 are currently available. We have, therefore, recovered, transmitted, and re-aligned the years 1981-2017 and 2020. All files are now available to the VERIFY consortium.

4.2. Land use datasets and high-resolution land cover change and biomass mapping

4.2.1. High resolution land cover

We provide an updated version (EUR v201101) of the European subset of the HILDA+ Land Use Change, a global dataset of land use/land cover (LULC) change at 1 km spatial resolution. We developed the HILDA+ land use/land cover maps using a data-driven reconstruction approach as described in deliverable D3.1 and updated in D3.2.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

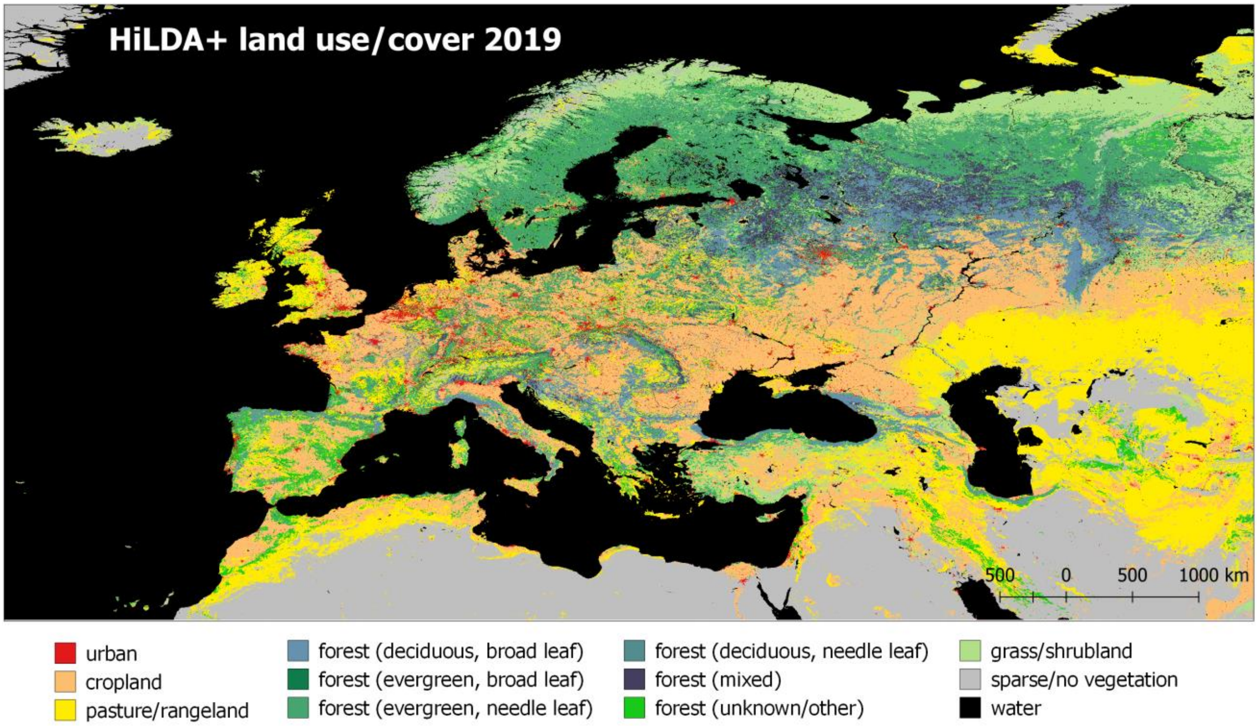

Figure 1: HILDA+ Land use/cover data of the year 2019.

Updates compared to the previous version:

- Extension of time series to 2019: HiLDA+ full time series covers now 1899*/1960-2019 (*using trend extrapolation, note that observational data were only for used for 1960-2019)
- Integration of new Earth Observation datasets/Land Cover classifications for LULC fraction maps (change allocation):
    - Copernicus LC100 2016-2019 https://land.copernicus.eu/global/products/lc
    - CORINE land cover 2012 and 2018 https://land.copernicus.eu/pan-european/corine-land-cover Forest class sub-division and extension to forest dynamics to 2018 with integration of ESA C3S Global Land Cover products C3S-LC-L4-LCCS-Map-300m-P1Y-v2.1.1 (2016-2018).
- Forest class sub-division extended to 2018 with integration of ESA C3S Global Land Cover products C3S-LC-L4-LCCS-Map-300m-P1Y-v2.1.1 (2016-2108). The forest dynamics included in HiLDA+ EUR v200111 given in chronological order:
    - 1900-1959: static forest division from ESA-CCI year 1992 and extrapolated forest area dynamics from HILDA+
    - 1960-1991: static forest division from ESA-CCI year 1992 and forest area dynamics from HILDA+
    - 1992-2018: within-forest dynamics from ESA-CCI and forest area dynamics from HILDA+

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

  - 2019: static forest division from ESA-CCI year 2018 and forest area dynamics from HILDA+

- Update in change allocation procedure: changed minimum LULC fraction threshold for change allocation procedure from 0.0 to 0.3 (used in ranking the "probability maps" for finding most probable change cell)
- Update of the FAO land use database, annual national land use trends based on FAOSTAT land use, updated version 2020-09-10
  http://www.fao.org/faostat/en/#data/RL/metadata

HILDA+ Land Use Change EUR v201101 has the following specifications:

- Two spatial datasets:
  1) hildaplus_vEUR-v201101_LULC_states.nc, which contains annual land cover/use types (11 classes: 6 LULC generic LULC classes - Urban, Cropland, Pasture/Rangeland, Forest, Grass/Shrubland, No/sparse vegetation - and 6 Forest subclasses), and
  2) hildaplus_vEUR-v201101_LULC_transitions.nc, which contains annual LULC transition types between two years, are provided with respective documentation.
- Format: NetCDF
- Projection: EPSG:4326 - WGS 84 – Geographic
- Spatial coverage/extent: -48.0000000000000000,27.0040397419643625 : 88.0016336493667097,82.0000000000000000
  Temporal coverage: 1899-2019
- Spatial resolution: ~0.00998837° (~1 km)
- Temporal resolution: 1 year/ annual
- Dimensions(sizes): time(116), latitude(5506), longitude(13616)
- Variables(dimensions):: uint16 time(time), float32 latitude(latitude), float32 longitude(longitude),
  1) uint8 LULC_states(time,latitude,longitude)
  2) uint16 LULC_transitions(time,latitude,longitude)

4.2.2. Biomass maps

Over the past years, several global maps of above-ground biomass (AGB) have been produced and are updated, but they exhibit significant differences that reduce their value for climate and carbon cycle modelling. As part of the current project year, we developed and provided European-scale data packages as data input to climate models multi-epoch and multi-resolution aboveground biomass (AGB), aboveground carbon (AGC), below-ground biomass (BGB), and below-ground carbon (BGC) data layers with uncertainty estimates in the form of standard deviations (SD). The data layers include adjusted AGB and SD layers from the original data sources (see AGB map data sources) after an uncertainty assessment and comparison with globally distributed biomass plot data. The original biomass map data sources included those from Baccini 2000 (Global Forest Watch, 2002), GEOCARBON 2008 (Avitabile et al. 2016), GlobBiomass 2010

(Santoro et al. 2020), and CCIBiomass 2017 (Santoro et al. 2019). The data package delivered to the WP3 model users include:

- AGB map: The original AGB map from a data source was averaged at coarser resolution and compared with the weighted average of plot biomass (weighing is based on plot uncertainty). Differences of the map and plot AGB or map "bias" were modelled using spatially exhaustive covariate and random forest regression. The spatial bias was subtracted to provide an adjusted "best-available" aboveground biomass map.
- SD layer: the original SD layers from the data source were adjusted based on the length of autocorrelated AGB residuals from variogram models, and after assessing the map error conformity (either as pessimistic or optimistic about map precision)
- BGB and uncertainty layer: root-to-shoot ratios from Spawn et al., (2020) was used. The ratio layer used a regression model for woody vegetation using phylogeny, temperature, forest management and empirical equations for non-woody vegetation.
- Carbon stocks: AGC, BGC, AGB, BGB, and SD layers were multiplied by 0.475 (average of global AGB-AGC ratio)

Each VERIFY modelling group received a biomass data package targeted at their specific needs recorded as part of a dedicated survey that was conducted as part of the project.

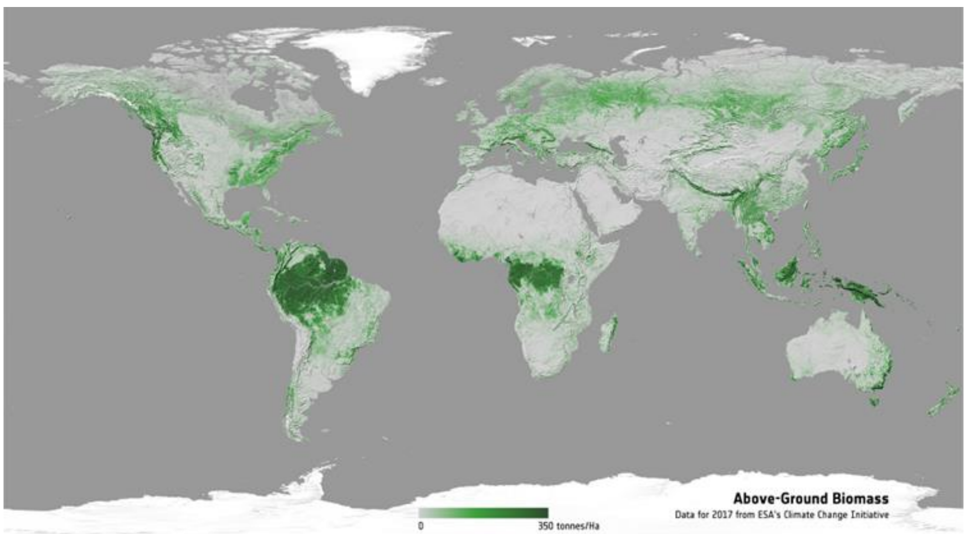

Figure 2: Global aboveground biomass at 100m resolution derived from space-based data source as one input to the VERIFY biomass data package (Source: ESA: http://cci.esa.int/biomass)

The framework for estimating biomass map combining plot and map data provides a general and objective method to estimate the accuracy and uncertainty of AGB maps that is applicable to both local and global-scale analysis. We also provide a tool for users of the framework to pre-process plot data, estimate their uncertainty and compare plot and map AGB estimates (https://github.com/arnanaraza/PlotToMap). The study therefore constitutes a major step towards improved AGB map validation and improvement.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

4.3. Soil property and soil erosion datasets

There are no changes in the provided datasets since the last deliverable (see D3.1).

4.4. Flux datasets for model testing

There are no changes in the provided datasets since the last deliverable (see D3.1).

4.5. Cropland management data

4.5.1. CAPRI fertilizer application data

There are no changes in the provided datasets since the last deliverables (see D3.1).

4.5.2. Management timing

There are no changes in the provided datasets since the last deliverables (see D3.1).

4.5.3. Nitrate and ammonium fertilizer data for croplands

In addition to the dataset provided by the JRC, nitrate and ammonium fertilizer data from literature were added to the data base. This is a global data set with fertilizer data as published by Nishina et al. (2017). The data provide global data in a 0.5°X0.5° spatial resolution for nitrate and ammonium fertilizer for the period 1961-2010. The data set is accessible online (doi:10.1594/PANGAEA.861203 link last used 12th of April, 2021).

4.6. Grassland management data

There are no changes in the provided datasets since the last deliverables (see D3.1).

4.7. Forest management data

An approach was developed to map forest management strategies (see Nabuurs et al. 2020) as an example of what can be achieved with innovative applications where experts and the latest digital data sources are combined. This is an interactive tool where various preferences and drivers regarding forest functions can be tested and effects quantified. With any new update of harmonized data or higher-resolution information, such a tool can help steer policy discussions and visualize how the main drivers of forest change might affect the forest management strategies from very intensive (short rotation, highly productive species addressing mainly raw material provision) to strict nature

management (focused on biodiversity conservation) and all variations in between (Fig. 4.3). We used a Bayesian belief network (BBN) as an experimental mapping procedure. BBNs facilitate modelling when data availability is insufficient to use a deterministic approach and can mix statistical and encapsulated expert knowledge. During several sessions, forest experts brainstormed on drivers, pressures and impacts resulting from forest management. Then the relationship from a single driver to the management intensity was described independently by using probabilistic relationships in a BBN. Since no statistical response functions were available, expert knowledge was used instead.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

The beta versions of probability tables of drivers were then discussed in subgroups and exchanged between groups. Finally, they were discussed in a plenary session, and resulting maps were immediately displayed, discussed again, changed and agreed on. Expertise represented by these 34 people included a BBN technical expert, 2 GIS information experts, a forester, an environmental services expert, Common Agricultural Policy specialists, agronomists, nongovernmental organizations, JRC/European Commission, forest policy experts and nature conservation experts. Besides the usual data on management and environment such as tree species, elevation and soil type, indirect indicators were used such as accessibility from cities and gross domestic product, which were assumed to partly reflect management intensity and demands in the future. More detailed information on the method and indicators is presented in Nabuurs et al. 2019, Supplementary information.

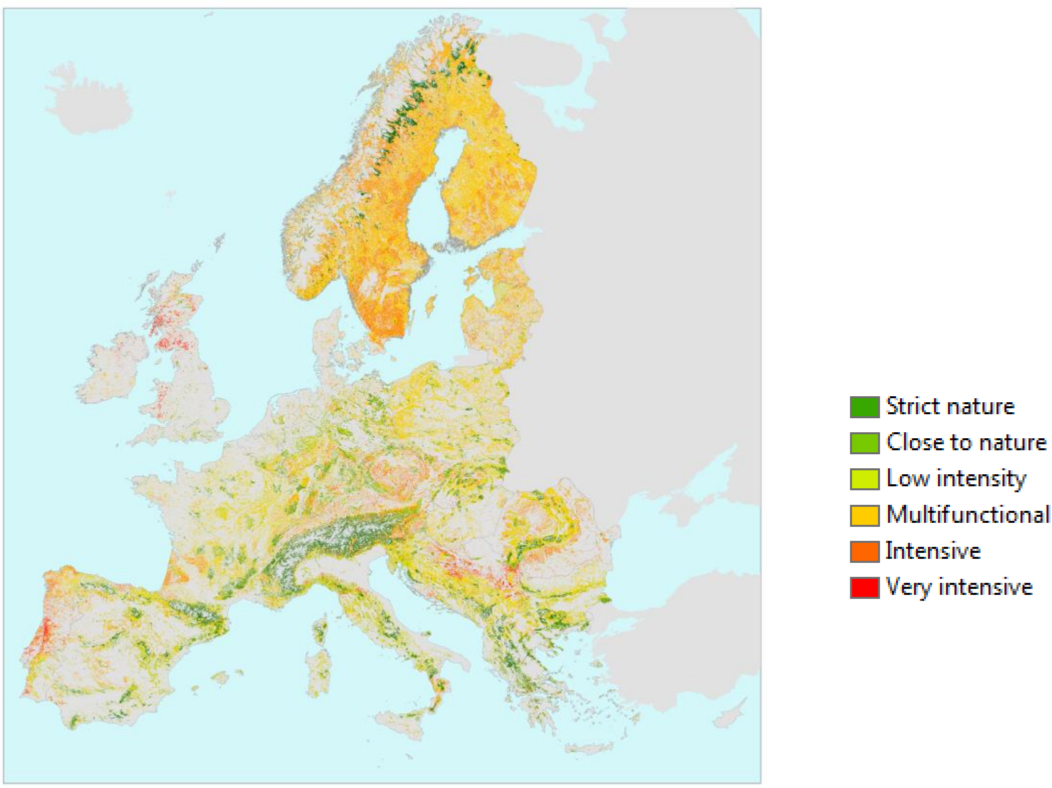

Figure 3: European (European Economic Area countries) forest management strategies map showing a possible route of strategizing forest management (Nabuurs et al. 2019).

4.8. Nitrogen deposition data

The collected nitrogen data of deliverable D3.1 were discussed with the modeler groups and their needs assessed. The provided nitrogen deposition maps will be used by the models LSM, ORCHIDEE, ECOSSE and EPIC. Based on the discussion, JRC prepared specific nitrogen deposition maps (wet and dry) for Europe that are more accurate than the global product used in the CMIP6 exercise.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

The EMEP MSC-W model (hereafter referred to as the 'EMEP model') is a 3-D Eulerian chemistry transport model (CTM) developed at the EMEP Centre MSC-W (hosted by the Norwegian Meteorological Institute, Oslo, Norway) under the Framework of the LRTAP Convention (UN Convention on Long-Range Transboundary Air Pollution). The EMEP model has traditionally been aimed at simulations of acidification, eutrophication and air quality over Europe, to underpin air quality policy decisions (e.g., the Gothenburg Protocol), and has undergone continuous development for several decades in response to evolving scientific knowledge and increasing computer power. The EMEP model integrates comprehensive atmospheric chemistry in several hundred reactions involving hundreds of chemical species in the gas phase and particle phase. The model was described in detail by Simpson et al. (2012). Model updates since then, leading to version rv4.33 which was used in all calculations provided for the VERIFY project, have been described in later EMEP Status reports (see Simpson et al., 2019 and references therein). The model is driven by meteorological data from the ECWMF IFS (European Centre for Medium-Range Weather Forecasts – Integrated Forecast System). Land-use data are taken from the CORINE land-cover maps (de Smet and Hettelingh, 2001), the Stockholm Environment Institute at York (SEIY, which has more detail on agricultural land-cover), the Global Land Cover (GLC2000) database (JRC, 2003), and the Community Land Model (Oleson et al. 2010, Lawrence et al. 2011). More details about this can be found in Simpson et al. (2017).

This will improve the data set and addresses common issues that that LSMs are facing when using atmospheric deposition data at relatively low resolution with land-use / land-cover in the atmospheric model inconsistent with that of the LSM.

4.9. Freshwater fluxes and river exports

This dataset consists of a monthly climatology-based average CO_2 emissions from lakes (including reservoirs) and rivers at the spatial resolution of 0.1°. The outgassing of CO_2 from the inland water network has been calculated using different methods for rivers and lakes. As consequence, emissions from river include all months over the 1979-2012 period, while the emissions from lakes only represent a climatological seasonal signal without inter-annual variability. In rivers, the seasonality in CO_2 evasion has been simulated by the model ORCHILEAK and has been superimposed on the previously reported annual-mean climatology derived from observations. In lakes, a data driven approach using multiple linear regressions allows predicting for each season pCO_2 and CO_2 emission for each of the >70000 European lakes. The multiple linear regressions are performed using individual lakes as calculation units for four different seasons. Lake size, temperature and altitude are used as predictors for lake pCO_2. Both lake and river data products assume no CO_2 exchange through ice and were combined to produce a monthly climatology at the spatial resolution of 0.1°.

The model ORCHILEAK has also been used to simulate the lateral redistribution of carbon at the scale of the EU27 + UK along the riverine network, from terrestrial ecosystems to the coast. The calculated fluxes include Dissolved Organic Carbon, Particulate Organic Carbon and soil-derived

CO_2 and are calculated daily at the spatial resolution of 0.5° over the period 1901-2014. These results allow calculating carbon budgets and retentions along the European hydrological river network.

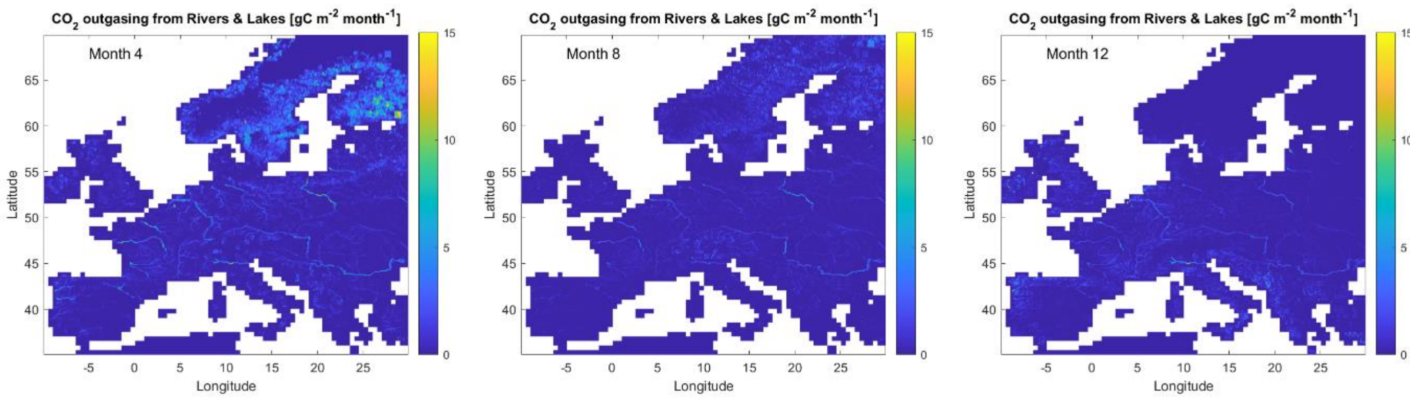

Figure 4: Averaged CO_2 outgassing from inland waters (lakes and rivers) in April, August and December.

4.10. Coastal ocean CO_2 fluxes

For estimating the air-sea CO_2 gas exchange in European shelf seas, we generate maps of sea surface pCO_2 covering the area from the western Mediterranean to the Barents Sea and then calculated the fluxes based on these maps combined with the atmospheric xCO_2 in the marine boundary layer and 6 hourly wind speed data. The pCO_2 maps are based on fitting a set of driver data (for example sea surface temperature, mixed layer depth and chlorophyll concentration) against gridded fCO_2 observations from SOCAT (Bakker et al., 2016). We use two different methods for this: Random Forest regressions and Multi-Linear Regressions.
During the last year, the focus of our work has been towards optimizing the workflow and integrating of coastal flux maps and open ocean fluxes for the inversions.

(1) *Optimizing the workflow for the 2021 version*
The coastal flux maps are using fCO_2 observations collected in SOCAT (www.SOCAT.de). In the annually-updated SOCAT database, surface ocean fCO_2 observations are collected and quality controlled. The database is typically published in June and contains observations until the end of the prior year. This means that a flux product produced in March 2021 only can be based on quality-controlled observations until December 2019. In order to provide the best possible estimate of coastal CO_2 fluxes also for 2020, we therefore decided to move the production of the coastal flux maps from March to June/July. This will allow us to also include quality-controlled observations from 2020 when producing the maps. Additionally, the open ocean maps covering 2020 which will be needed for producing the Multi-Linear Regression product and for the merging step below won't be released before late July.

(2) *Integration of coastal flux maps and open ocean maps*

For using the maps of coastal CO_2 flux in the atmospheric inversions, a joined product of open ocean and coastal fluxes has to be developed. We have decided to follow two different merging routines:

- easy merging: choses the coastal map where available and otherwise the open ocean map
- elaborate merging: regression between open ocean and coastal maps along the edge of the coastal maps, weighted by the standard error in the two products (as described in Landschützer et al. (2020))

The easy merging has the benefit that it is faster and easier to implement, while the more elaborate merging routine will help to avoid step changes between neighboring grid boxes at the border between the coastal and the open ocean maps.

The 2021 version of the coastal flux maps (including fluxes until December 2020) was produced in the summer of 2021, merged with the open ocean map of Rödenbeck et al (2014), and is now available to groups in VERIFY.

4.11. XCO_2 from OCO-2 via FOCAL algorithm

There are no changes in the provided datasets since the last deliverables (see D3.1). There are plans for a new OCO-2 data set, which includes 2019. Additionally, ground-based measurements in Russia will be considered for the new data set.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

5. Organization of the database

There are no changes in the organisation of the data. The data server is up-dated regularly and changes and new up-dates were communicated among the target groups in WP3 and WP4. A list of available data and the data itself are available from the VERIFY THREDDS data server (TDS, https://verifydb.lsce.ipsl.fr/thredds/verify/catalog.html), with some limited metadata available from the VERIFY data catalogue (available from the VERIFY web site: http://verify.lsce.ipsl.fr/index.php/products). Note that for the VERIFY partners additional information on the different datasets is available under the password protected SharePoint platform (https://projectsworkspace.eu/sites/VERIFY/Lists/WP3inputdataset/AllItems.aspx).

The filenames will be assigned to contain various information about the file itself, including the method, species, institute, region, spatial coverage, temporal resolution, and the person who uploaded the file. This information is used to automatically generate a catalogue of available data (http://webportals.ipsl.jussieu.fr/VERIFY/CountryTot2.html). The TDS is developed by Unidata, a member of the UCAR Community Programs, managed by the University Corporation for Atmospheric Research, and funded by the National Science Foundation of the United States, with the goal of helping educators and scientists obtain and use geoscience data. The TDS also supports several dataset collection services including some sophisticated dataset aggregation capabilities. This allows the TDS to aggregate a collection of datasets into a single virtual dataset, greatly simplifying user access to that data collection. The TDS also contains viewing tools to facilitate direct user browsing of stored datasets, instead of forcing the user to rely on metadata.

More details are available in Deliverable D6.8.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

6. Conclusions

This collection of input data provides an excellent basis for the work in WP3 and WP4 and allows for increased comparability between model results. Datasets were adapted to meet the demands in VERIFY (e.g., the climate data), datasets were improved (land use data) and new data sets were created (merging CO_2 flux data from open ocean and coastal areas). The highlights of the data sets are the land use/land cover data set HILDA+ and the merged coastal and ocean fluxes. HILDA+ provides land use/land cover data from 1900 to 2019 in a 1km spatial and yearly temporal resolution. This is a high-quality global data set provided to cover Europe in the context of VERIFY and include additional sub-group classes based on the requests of project partners. Through the expertise present in VERIFY, a merged dataset of European coastal and open ocean carbon fluxes was created, and will contribute to improved simulations of surface carbon fluxes by atmospheric inversions. The development of the forest management and biomass datasets advance our knowledge about these systems, in particular through an innovative approach that combines expert knowledge and digital.

The data collection revealed short-comings and data gaps that need to be addressed in future. Even though several data sets for agricultural management are provided, the collection showed limitations of the data sets (e.g., rough estimates, very little harmonization, large scale averages aggregated to spatial data sets) and were not provided to the end of the target period. While the datasets that were developed and maintained within the project were updated continuously, these additional data from outside the project remained unchanged and end before 2020, which is not suitable for an operational system.

One key aspect to the success of this work so far has been collaboration between modelling groups and data providers to tailor products for their needs (e.g., climate data, land use/land cover data, biomass data, nitrogen data, coastal fluxes). All these efforts resulted in this complete data set that provides all relevant data for the different bottom-up and top-down approaches used in VERIFY.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

7. References

Nishina, Kazuya; Ito, Akihiko; Hanasaki, Naota; Hayashi, Seiji (2017): Historical global map of NH4+ and NO3- application in synthetic nitrogen fertilizer, link to NetCDF files. PANGAEA, https://doi.org/10.1594/PANGAEA.861203.

Nabuurs, G. J., Verweij, P., Van Eupen, M., Pérez-Soba, M., Pülzl, H. & Hendriks, K., 2019, Next-generation information to support a sustainable course for European forests. Nature Sustainability. 2, 1062. https://doi.org/10.1038/s41893-019-0374-3

Rödenbeck C. ,. Bakker D. C. E, Metzl N.,. Olsen A, Sabine C., Cassar N., Reum F., Keeling R. F., and Heimann M., Interannual sea–air CO2 flux variability from an observation-driven ocean mixed-layer scheme, Biogeosciences, 11, 4599–4613, 2014 doi:10.5194/bg-11-4599-2014
Landschützer, Peter & Laruelle, Goulven & Roobaert, Alizee & Regnier, Pierre. (2020). A uniform p CO₂ climatology combining open and coastal oceans. 10.5194/essd-2020-90.

Simpson. D.. Benedictow. A.. Berge. H.. Bergström. R.. Emberson. L. D.. Fagerli. H.. Flechard. C. R.. Hayman. G. D.. Gauss. M.. Jonson. J. E.. Jenkin. M. E.. Nyíri. A.. Richter. C.. Semeena. V. S.. Tsyro. S.. Tuovinen. J.-P.. Valdebenito. Á.. and Wind. P. (2012) The EMEP MSC-W chemical transport model – technical description. Atmos. Chem. Phys.. 12. 7825-7865. doi:10.5194/acp-12-7825-2012.

Simpson, D., R. Bergström, H. Imhof, and P. Wind. Updates to the emep/msc-w model, 2016-2017. In Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components. EMEP Status Report 1/2017. The Norwegian Meteorological Institute, Oslo, Norway, 2017.

Simpson, D., R. Bergström, S. Tsyro, and P. Wind: Updates to the EMEP MSC-W model, 2018- 2019. In Transboundary particulate matter, photo-oxidants, acidifying and eutrophying components. EMEP Status Report 1/2019. The Norwegian Meteorological Institute, Oslo, Norway, 2019

De Smet PAM, Hettelingh J-P (2001) Intercomparison of Current European Land Use/Land Cover Databases, Status Report 2001 Coordination Center for Effects RIVM Report 259101010, Bilthoven, Netherlands, pp. 41-52

Oleson, K. (2010). Technical Description of the Community Land Model (CLM). NCAR Technical Note. TN-478+STR. 10.5065/D6RR1W7M.

Lawrence, David & Oleson, Keith & Flanner, Mark & Thornton, Peter & Swenson, Sean & Lawrence, Peter & Zeng, Xubin & Yang, Zong-Liang & Levis, Samuel & Sakaguchi, Koichi & Bonan, Gordon & Slater, Andrew. (2011). Parameterization Improvements and Functional and Structural Advances in Version 4 of the Community Land Model. Journal of Advances in Modeling Earth Systems. 3. M03001. 10.1029/2011MS000045.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.

Avitabile V, Herold M, Heuvelink G, Lewis SL, Phillips OL, Asner GP et al. (2016). An integrated pan-tropical biomass maps using multiple reference datasets. *Global Change Biology,* 22: 1406–1420. doi:10.1111/gcb.13139.

Santoro, M., Cartus, O. and Wegmüller, U., 2020. Estimation of Forest Above-Ground Biomass with C-Band Scatterometer Backscatter Observations. In *IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium* (pp. 4987-4990). IEEE.

Santoro, M., Cartus, O. and Wegmüller, U., 2019. Complementarity of X-, C-, and L-band SAR backscatter observations to retrieve forest stem volume in boreal forest. Remote Sensing 11 (13), 1563.

Spawn, S.A., Sullivan, C.C., Lark, T.J. and Gibbs, H.K., 2020. Harmonized global maps of above and belowground biomass carbon density in the year 2010. Scientific Data, 7(1), pp.1-22.

VERIFY is a research project funded by the European Commission under the H2020 program. Grant Agreement number 776810.