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)	**D6.9**
Deliverable name:	Final version of the VERIFY database and data-management infrastructure
WP / WP number:	6
Delivery due date:	Month 43 (31/08/2021)
Actual date of submission:	Month 53 (07/06/2022)
Dissemination level:	Public
Lead beneficiary:	CEA
Responsible scientist/administrator:	Philippe Peylin
Contributor(s):	Matthew McGrath
Internal reviewer:	/

Changes with respect to the DoA
This deliverable was delayed. Original submission date was month 43. Given delays due to the COVID-19 pandemics and the need to have the final set of data/plots from the 2021 synthesis, we had to postpone the deliverable.
Dissemination and uptake (Who will/could use this deliverable, within the project or outside the project?)
This deliverable describes how data is managed and stored both internally and made available to external users. It is thus an invaluable tool to data access for both internal and external users. Web addresses for data access are given in the core of the document.
Short Summary of results (<250 words)
The VERIFY project currently stores **3900 GB of data, including 150 GB** of output fluxes. This has grown during the final year of the project, as more groups took advantage of high-resolution forcing data provided by the project. In order to share this data among partners, a series of mechanisms were put in place, focused on a dedicated data server that allows users to download any dataset. The data is separated by input and output data and by primary chemical species to allow easier access. Additional services were placed on top of this server, permitting users to browse and visualize the various datasets without downloading it. Extensive work has gone into harmonizing both the input and the output datasets in terms of file formats, grids, variable names, variable units, and aggregating spatial products into time series by country/region. This work enabled the creation of three distinct tools to manipulate the data. The first is for advanced scientific users, and helps them personalize plots. The second is for regular scientific users, and helps them explore time series and maps in a user-friendly manner. The third allows policy-makers to view a set of already-completed plots summarizing the GHG fluxes of all countries and multiple regions; these plots were created in coordination with WP5 for use in the annual synthesis paper and with WP1 for use in the country summary Fact Sheets.
Evidence of accomplishment (report, manuscript, web-link, other)
The data management system can be accessed through the main VERIFY website: https://verify.lsce.ipsl.fr/index.php/products

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

Version	Date	Description	Author (Organisation)
V0	06/09/2019	Creation/Writing	Philippe Peylin, Matthew McGrath, Patrick Brockman (CEA)
V1	08/07/2022	Formatting/Delivery on the Participant Portal	Philippe Peylin and Aurelie Paquirissamy

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

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

1. Glossary

Abbreviation / Acronym	Description/meaning
TDS	THREDDS Data Server
THREDDS	Thematic Real-time Environmental Distributed Data Services
LSCE	Laboratoire des Sciences du Climat et de l'Environnement
IT	Information technology
NetCDF	A file data format for storing scientific information

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

2. Executive Summary

This deliverable describes the final version of a state-of-the-art database and data infrastructure developed on the CEA-LSCE premises. After the completion of three cycles of GHG flux synthesis from the different work packages the current data volume exceeds 3900 GB of data. In order to share this data among partners, a series of mechanisms were put in place, focused on a dedicated data server that allows users to download any dataset. The data is separated by input and output data and by primary chemical species to allow easier access. Additional services were placed on top of this server, permitting users to browse and visualize the various datasets without downloading them. The database can be access through the "Products" page of the main VERIFY web site: http://verify.lsce.ipsl.fr/index.php/products

Extensive work has gone into harmonizing both the input and the output datasets in terms of file formats, grids, variable names, variable units, and aggregating spatial products into time series by country/region. This work enabled the creation of the following different data services:

i) A searchable catalog that displays all datasets gathered in the database. This facility was developed for expert users who need to upload the raw data and access ancillary associated files (information files associated with each output of VERIFY). In addition, specific "Jupyter notebooks" are provided to the expert users in order to share knowledge and computational programs for specific data analysis. These notebooks are stored in a GitHub repository.

ii) A user-friendly time series visualization tool that allows the user to build his own time series of different GHG fluxes from different data sets and for different pre-defined regions (individual EU countries or groups of countries).

iii) An additional map viewer tool has been added in order to interactively view the spatial distribution of the different fluxes.

iv) A policy-makers tool to view a set of already-completed plots summarizing the GHG fluxes for all countries and multiple regions; these plots were created in coordination with WP5 for use in the annual synthesis paper and with WP1 for use in the country summary Fact Sheets. The figure below provides an illustration of the policy-maker oriented visualization tool. It provides the main interface on the left for selecting the GHG and regions and one example of such synthesis for the total CH4 emissions.

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

Figure 1 : Illustration of the GHG synthesis plot facility: left: web interface to select the specific synthesis plot; right: example of such synthesis for the anthropogenic CH4 emissions for the EU27+UK.

3. Introduction

Data management ensures the proper dissemination of products from projects, as well as facilitating harmonization across different work packages and research groups. Sharing large data files presents numerous challenges. Email is a poor method to share large files, to the point that every email provider caps the maximum file size to prevent email traffic from clogging network connections. In addition, the lack of coordination among groups can quickly lead to certain groups using different versions of the same file, sabotaging attempts at harmonization. To overcome these two major challenges to data management in VERIFY, we selected a centralized system known as the THREDDS Data Server (TDS) which allows the project coordination team to upload data files to a dedicated data server and then users to download them to their local machines. THREDDS catalogs enable one to mark datasets with meta-data, making browsing more convenient.

TDS was designed to work with the NetCDF data format of files, offering a variety of services. These services include direct download of the whole dataset (via HTTPS), or download, access and display part of the data (Web Coverage Service, WCS; NetCDF Subset Service, NCSS; WMS; OpeNDAP; DAP4). NetCDF files are the preferred file format for representing time-varying spatial data in the climate community, and many tools have been developed over the years to construct, modify, and manipulate these files. NetCDF provides an abstraction layer that allows users to access individual data elements without knowing the details of how they are stored, which makes them faster to access and easier to transfer across different computing systems. This is a major advantage in fields like climate science where large amounts of data are generated on different supercomputer architectures. In addition, metadata can be stored with the variables for more transparent documentation. Finally, the NetCDF format offers the same compression level as classical binary files.

We recognized that many people in the policy community have no knowledge of NetCDF file formats. Therefore, we worked to put in place structures which take advantage of the file format but hide the complexity of the underlying data layer. This permits us to disseminate our results as widely as possible. This report explains these structures in more detail, including relevant steps in the data processing chain to make results from research groups in VERIFY available to both internal and external users. This report is also the final version of two reports on the database and data management; it thus strongly relies on the first version (D6.8, accessible from https://verify.lsce.ipsl.fr/index.php/repository/public-deliverables/wp5-wp6-synthesis-and-products-policy-relevant-ghg-monitoring-and-verification-system-design) providing updates where relevant.

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

4. Setup

The data management infrastructure in VERIFY is housed at the Laboratoire des Sciences du Climat et de l'Environnement (LSCE) outside of Paris, France. The main data storage takes place on a hard drive attached to a large computer cluster maintained by the local information technology (IT) team, which ensures professional care and maintenance. The local IT team has also configured the publicly-accessible web page that permits users to access the resources, in conjunction with a specialist in data visualization at the LSCE and members of the VERIFY coordination team. The computing cluster also ensures sufficient computational power for harmonizing (processing) the datasets as described in Section 4.

The data management infrastructure is accessible through the main VERIFY website on the "Products" page: **https://verify.lsce.ipsl.fr/index.php/products**. The layout of this page is illustrated in the figure below.

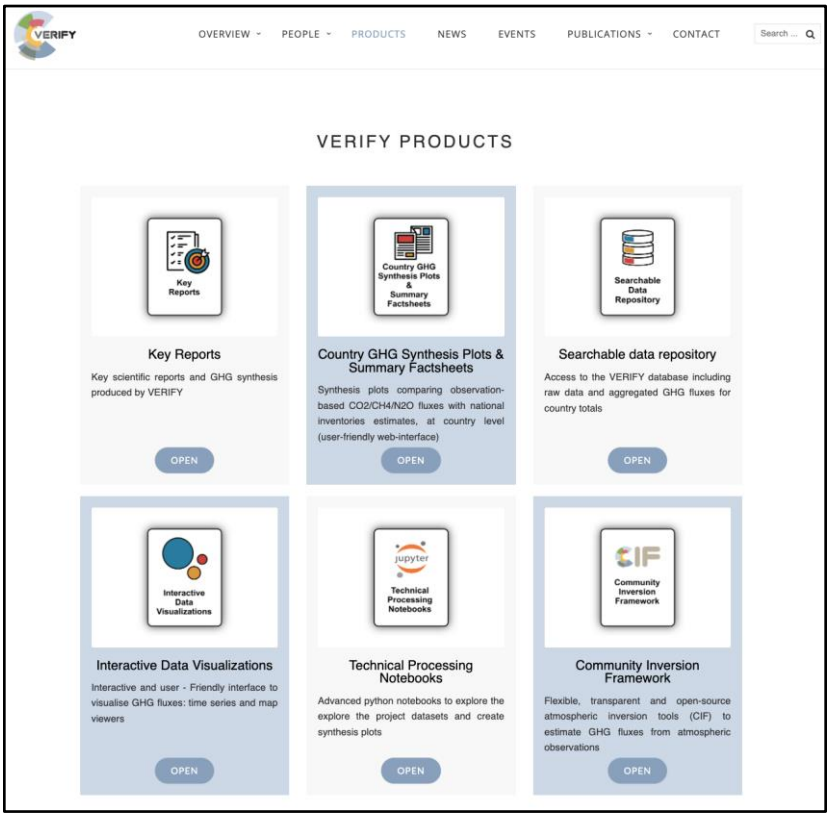


Figure 2: Main display from the "Products" page of the VERIFY web site.

From this page, the user has the choice of accessing several services:

1. An access to key reports of the project (mainly deliverables)
2. An access to country specific GHG synthesis plots and summary factsheets
3. A searchable catalog of all the data products (icon on the top right of Figure 2);
4. An access to tools for real-time visualization, manipulation, and comparison of maps and time-series at the country and regional levels (lower left icon);
5. A selection of scripts in the form of Jupyter notebooks for creating customizable plots (lower central icon);
6. An access to the Community Inversion Framework developed in VERIFY.

All of these options were regularly updated over the course of the project, especially for the pre-made synthesis plots as the project completed several operational cycles. The different data exploration tools correspond to different types of usage: from expert scientific users who need to access the raw data (service 3) to policy makers who need to be informed by the key results of the project (service 2). All these services are further described in the remainder of the document.

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

5. Data processing

The act of making data available to internal and external users begins with receiving the data from research groups, both inside and outside the project. For research groups kind enough to provide us with data who are outside VERIFY, we inform them of our data management system and obtain explicit consent before we share their products through our infrastructure. Research groups internal to VERIFY are already aware of their obligations under the project.

Research groups use a variety of different formats for their data, which makes developing generally-applicable tools a challenge. Our first step was to design a simulation protocol for two of the major work packages (WP3 and WP4) which specified a number of features to make analysis easier (Milestone 13). We have also been working with the other work packages to spread this same information, largely concerning how time is represented in the data, the spatial grid, the units of the variables, and full transparency and documentation about what the data includes (the so-called meta-data).

Data used for input into models is typically loaded immediately onto the server after making sure documentation is available. As input data is not visualized by the tools described below, and as each group has different needs for the data driving their individual models, there is no reason to drastically change the formatting.

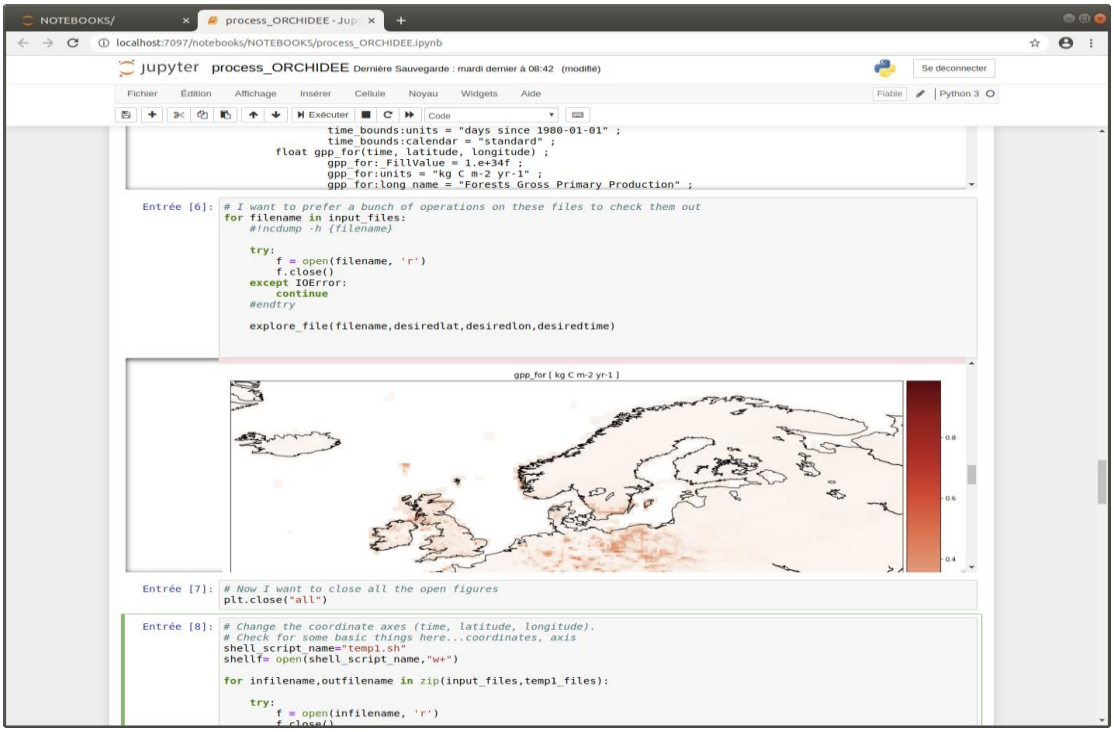


Figure 3: An example of a Jupyter notebook used to process data files before placement on the TDS.

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

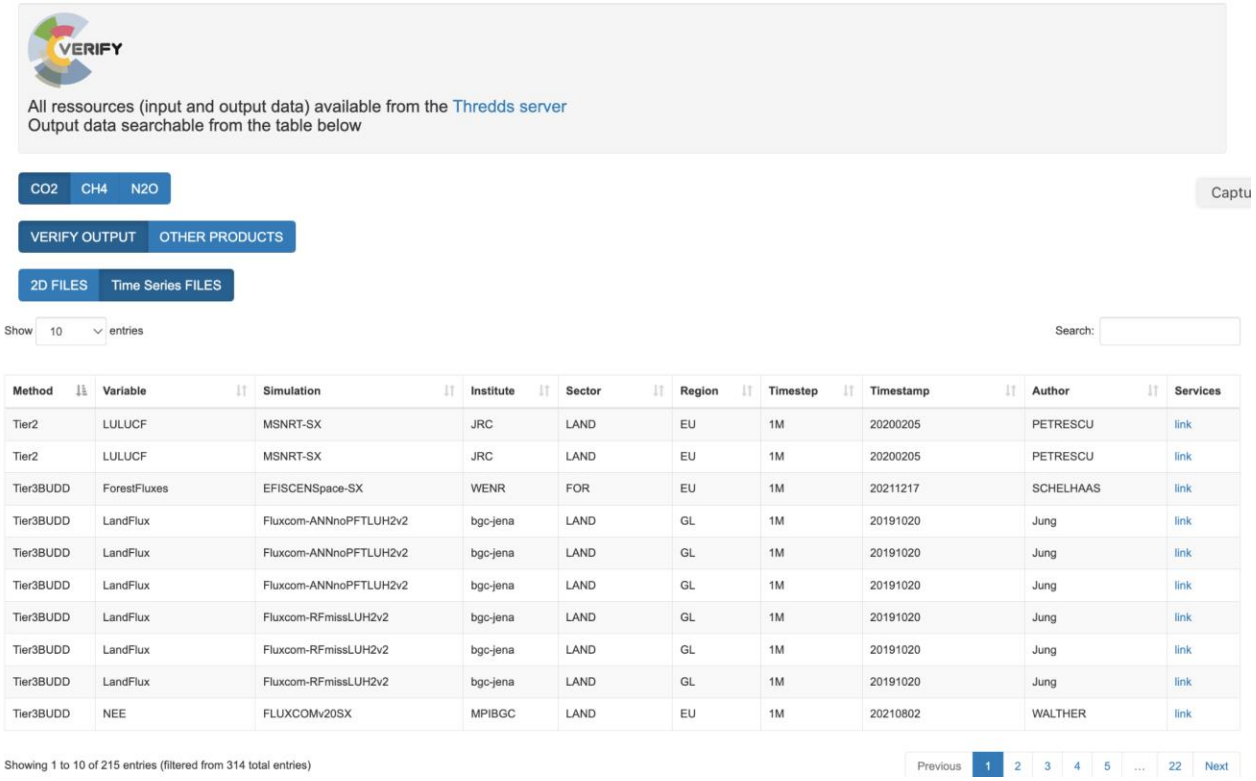
However, we have set in place a specific “quality check” procedure for all output data that will be distributed under the VERIFY database. Computational tools are notoriously sensitive to minor modifications and typographical errors, refusing to work with simple spelling mistakes. Despite our efforts to promote a single file format, our first action upon receiving output data from groups is therefore to examine it with a script known as a Jupyter notebook. Figure 3, above, illustrates the content of a Jupyter notebook which is an open-source web application that permits one to combine coding language, mark-up text, equations, and visualizations all in the same package script. The main objectives were to:

- One notebook, based on the Python scripting language, was created for most output dataset received due to the uniqueness of every data file, though the overall flow and some of the individual components can remain the same.
- The notebook opens the file, checks the header information, plots some of the variables, and calculates some simple statistics.
- It then runs through a series of actions to harmonize the time axis, variable names, and variable units, before copying the processed file to the TDS.

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

6. Data access

The THREDDS Data Server (TDS) permits access to all VERIFY harmonized data products via the "Searchable data repository" button on the official VERIFY website: https://verify.lsce.ipsl.fr/index.php/products. The content of this page is illustrated on Figure 4.



Method	Variable	Simulation	Institute	Sector	Region	Timestep	Timestamp	Author	Services
Tier2	LULUCF	MSNRT-SX	JRC	LAND	EU	1M	20200205	PETRESCU	link
Tier2	LULUCF	MSNRT-SX	JRC	LAND	EU	1M	20200205	PETRESCU	link
Tier3BUDD	ForestFluxes	EFISCENSpace-SX	WENR	FOR	EU	1M	20211217	SCHELHAAS	link
Tier3BUDD	LandFlux	Fluxcom-ANNnoPFTLUH2v2	bgc-jena	LAND	GL	1M	20191020	Jung	link
Tier3BUDD	LandFlux	Fluxcom-ANNnoPFTLUH2v2	bgc-jena	LAND	GL	1M	20191020	Jung	link
Tier3BUDD	LandFlux	Fluxcom-ANNnoPFTLUH2v2	bgc-jena	LAND	GL	1M	20191020	Jung	link
Tier3BUDD	LandFlux	Fluxcom-RFmissLUH2v2	bgc-jena	LAND	GL	1M	20191020	Jung	link
Tier3BUDD	LandFlux	Fluxcom-RFmissLUH2v2	bgc-jena	LAND	GL	1M	20191020	Jung	link
Tier3BUDD	LandFlux	Fluxcom-RFmissLUH2v2	bgc-jena	LAND	GL	1M	20191020	Jung	link
Tier3BUDD	NEE	FLUXCOMv20SX	MPIBGC	LAND	EU	1M	20210802	WALTHER	link

Figure 4: VERIFY data product catalog accessible via the official VERIFY website

Several facilities have been included since the first version and the product catalog now offers:

1. A direct access to the TDS web catalog by clicking on the link "Thredds server" at the top of the page, close to the VERIFY logo.

2. A catalog / table that uses metadata about all output files to allow them to be sorted. These meta-data are:
   a. "Variable": a descriptor of the kinds of variables included in the file,

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

b. "Simulation": describes the model or technique used,
c. The "institute" that produced the dataset,
d. "Sector" for land, forest, grass, crop, ocean,
e. "Region" for EU (Europe) or GL (global) product,
f. "Timestep" for monthly (1M) or annual (YR),
g. "Timestamp" to indicate when the file was created,
h. "Author" for the researcher who produced the dataset,

Access to the raw data themselves is then done by selecting the link in the "Services" column for the product. Note that two other columns in the catalog will be added in June 2022 to display the period (years) covered by the products and the email address of the data owner.
Finally, we added above the catalog itself three "menus" in order to filter the output data following:

- The GHG species: CO2, CH4 and N2O
- The VERIFY specific data or those from other projects
- The type of files: 2D files and Time series files for the regional totals.

Direct access to all data on the TDS:
The TDS web catalog can be also accessed directly with the following link: https://verifydb.lsce.ipsl.fr/thredds/verify/catalog.html. It provides the raw version of the catalog, which is less presentable to the ordinary viewer than the catalog. However, it is convenient for expert users to navigate through the different sub-directories of the TDS in order to find all data files. Figure 5 below illustrates the main directories of the TDS server.

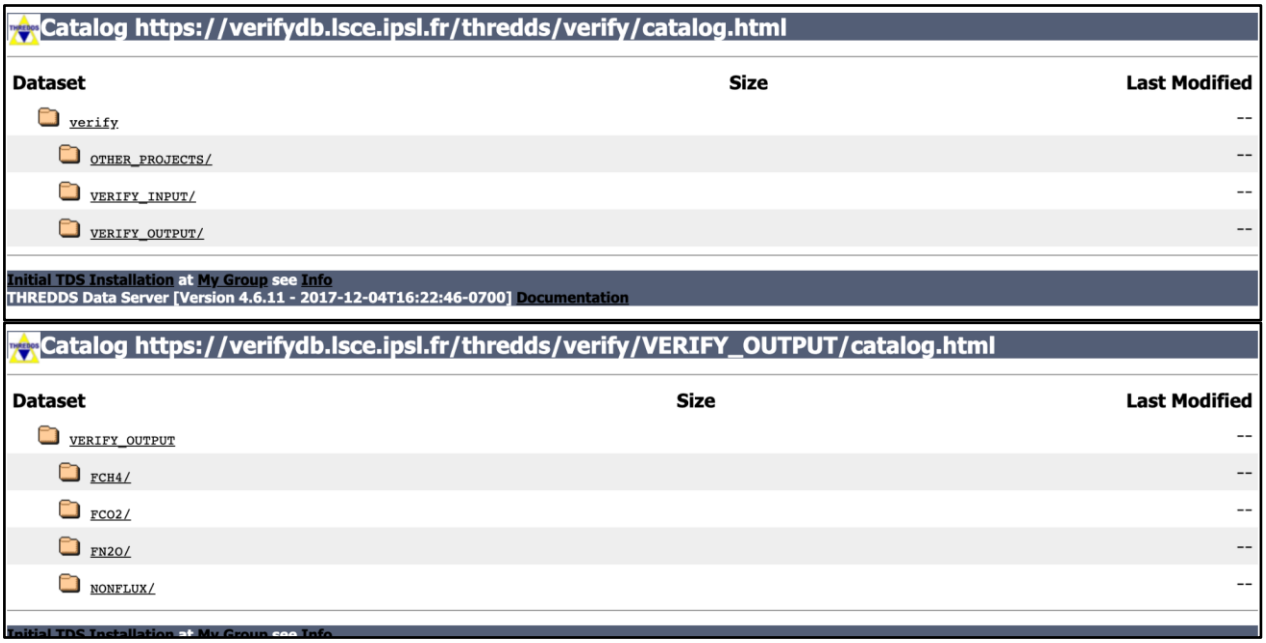


Figure 5: Illustration of the Thredds Data Server structure; top: root structure with the access to different types of products; bottom: organization by species of the VERIFY output data.

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

The TDS contains more information on each file than displayed in the catalog, including access to the processing scripts (Python scripts and Jupyter notebooks) used to process many of the VERIFY output fluxes. These scripts used to process a given FILENAME are contained in the FILENAME.process folder under the different VERIFY output sub-directories.

The TDS server contains different data sets including not only those produced by the VERIFY project but also numerous data sets gathered for the GHG flux synthesis of VERIFY. These data sets are grouped under the "OTHER_PROJECTS" sub-directory and their use follows specific data policy / data access.
The main VERIFY output data sets can be found under the "VERIFY_OUTPUT" directory and subsequent sub-directories (one per GHG). Different files can be found corresponding to:

- Spatially explicit 2D NetCDF files containing the raw output data from model simulations defined over a global / regional spatial grid. These files end in _2D.nc.
- Country total time series created from the 2D files that correspond to the aggregated fluxes (annual or monthly time steps) per region. The ensemble of regions included initially all European countries and a pre-selected set of groups of countries; it has been updated to include all countries of the world with a final set of 230 regions (countries and groups of countries). Two types of country total files were created depending on the masks that were used:
    - files ending with _CountryTotWithEEZ.nc, that correspond to country masks including the Extended Economic Zone (EEZ) for each country, i.e. the coastal ocean under that country's influence.
    - files ending with _CountryTotWithOutEEZ, that correspond to the country masks but without the EEZ. For simplicity these later sets are not shown directly in the catalog but are accessible through direct search on the Thredds server.

These files are of less interest to the average user, but very useful to the advanced user, depending on the analysis, the type of GHG and the sector that is envisaged.

Access to country totals for each data product:
Given the main objectives of the VERIFY project, a specific emphasis is given to files containing the country's total fluxes.
The link found in the "Services" column of the catalog (see Figure 4) enables the user to download all or part of the "country-total data", using the following techniques: HTTPS, Web Coverage Service (WCS), NetCDF Subset Service (NCSS), WMS, OpeNDAP, and DAP4. Figure 6 illustrates the proposed services associated with these country total files on the TDS.

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

Figure 6: Illustration of the proposed services associated with a "country total" file, accessible through the "Services" link on the data product catalog (see Figure 4)

7. Interactive data visualization tools

This section describes the final version of an ensemble of tools to interactively visualize the data that are produced by the VERIFY project (including non-VERIFY data). The objective is to provide advanced users with some tools to create specific plots for the analysis of the GHG fluxes produced in VERIFY. As detailed in section 3 (Setup) two major facilities are provided:

I. a technical processing toolbox (based on notebooks) for expert users and
II. a user-friendly data visualization tool for advanced users. The visualization tools consist of a "time series viewer" and a "map visualization" facility that was recently added.

7.1 Toolbox of data processing notebooks

The page is accessible from the main "Products" page by selecting the "Technical processing notebooks" icon: http://verify.lsce.ipsl.fr/index.php/products
It displays an ensemble of notebooks (Jupyter notebooks) archived under a worldwide software development platform, GitHub. The GitHub repository contains currently (March 2022) an ensemble of notebooks to i) process the data (see for instance, section 4) and ii) plot CO2 fluxes.
Such advance user facilities will be still enriched during the last months of the project with new plotting scripts that will be provided both by the coordinator (CEA-LSCE)

7.2 Time series visualization tool

The time series and mapping facilities are accessible from an intermediate menu (see figure 7) that is obtained by selecting the "Interactive Data Visualization" box of the "PRODUCTS" page (see figure 2).

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

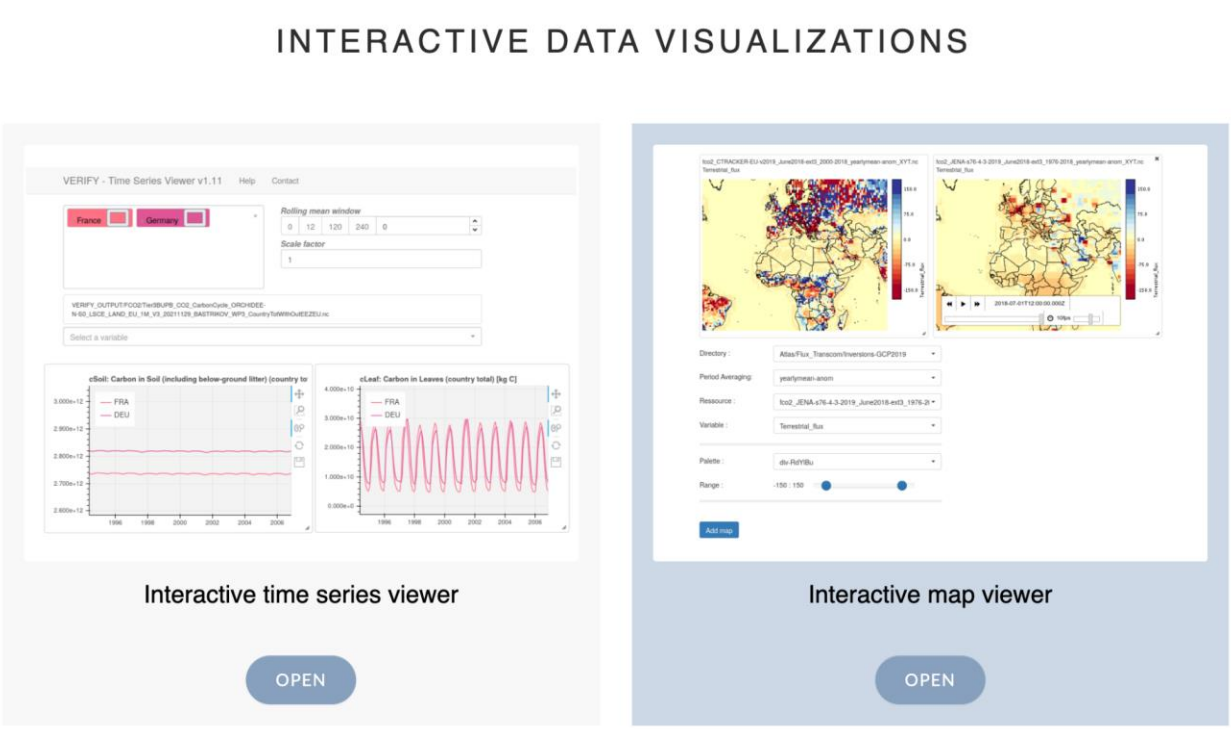


Figure 7: Intermediate page obtained from the "PRODUCTS" page of the web site, when selecting the "Interactive Data Visualization": provide access to the time series viewer and the interactive map viewer.

Selecting the "Interactive time series viewer" (figure 7) opens up a similar **catalog** with all data datasets as described above for accessing the raw data (figure 4). However this catalog includes a new column the **"Time series viewer" link** (second-to-last column) and has a smaller set of columns (metadata). This link will open up a window allowing the user to explore time series for different countries and regions in an interactive way.

Selecting products:

The first step consists of selecting an ensemble of products from the catalog. Standard selection mechanisms are available (i.e. selecting different lines/products by clicking on them) and the selected data sets are highlighted in blue.

Analysis of common variables:

Users can then check which variables are contained in the different data sets and which ones are common to all files, using the button labeled "Find common variables" below the catalog. After clicking this button, a list of variables appears at the bottom of the screen, either in green or in blue. Those variables in green are variables which appear in all the selected datasets. If variables appear in green, users can click the "Compare time series" button to display the time series for any of the common variables for any country/region with labels indicating which dataset they come from.

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

Inter-active time series plot:

Figure 8 provides an illustration of the interactive time series viewer. Several features are available to the user:

- A window in the upper right allows the user to select countries and regions to plot, which are then displayed in the bottom window.
- Multiple variables can be chosen, which will open more windows at the bottom: one variable per window, with all the variables in the file accessible.
- The viewing windows can be dynamically resized, zoomed in, and closed down according to the user's whims.
- The viewing windows can also be saved as a "png" file.
- Rolling means can be imposed to make the data more smooth (upper right box)
- A scale factors can be applied to change units (which are displayed in the title of each window along with the variable).

All features are "dynamic" and the plots are thus updated automatically depending on the selected options. Note that on the top of the viewing window there are a "Help" and "Contact" buttons that allows to get more information on how to use this interactive time series viewer.

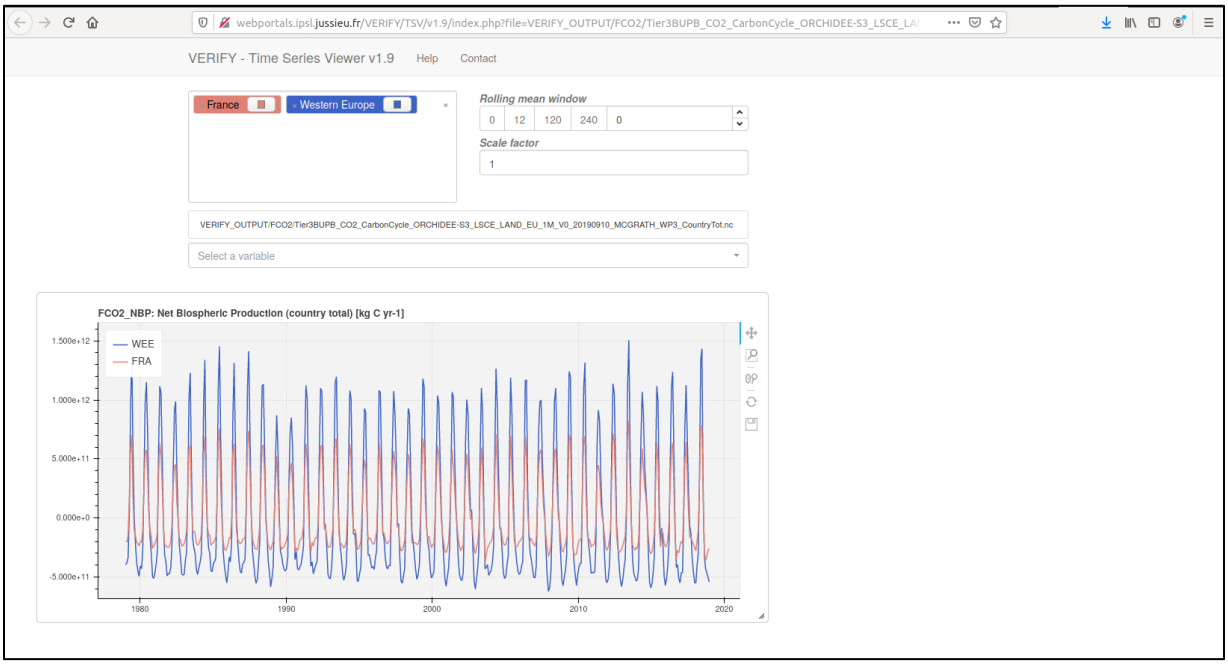


Figure 8: An example of the time series viewer available through the "Searchable data repository" button on the "Products" page on the VERIFY website

7.3 Mapping visualization tool

We recently added an “interactive map viewer” that can be accessible from the page described above in figure 7. When entering the map viewer, we first have a menu, as illustrated in figure 9, to select the first map to plot. A set of selection options are proposed:

- Directory: a menu to select a directory on the VERIFY database;
- Period Averaging: a menu to select the averaging period (this feature is not valid for all products);
- Rossources: a menu to select the ressources (i.e. the specific 2D file) to be used;
- Variable: a menu to select the variable from the product;
- Palette: a menu of predefined color palettes to select the one that will be used;
- Range: a menu to dynamically change the range (min and max) of the color palette.

Once all options are selected, we should click on the “Add map” button at the bottom of the page in order to visualize the map corresponding to the selected product and variable (see figure 9).

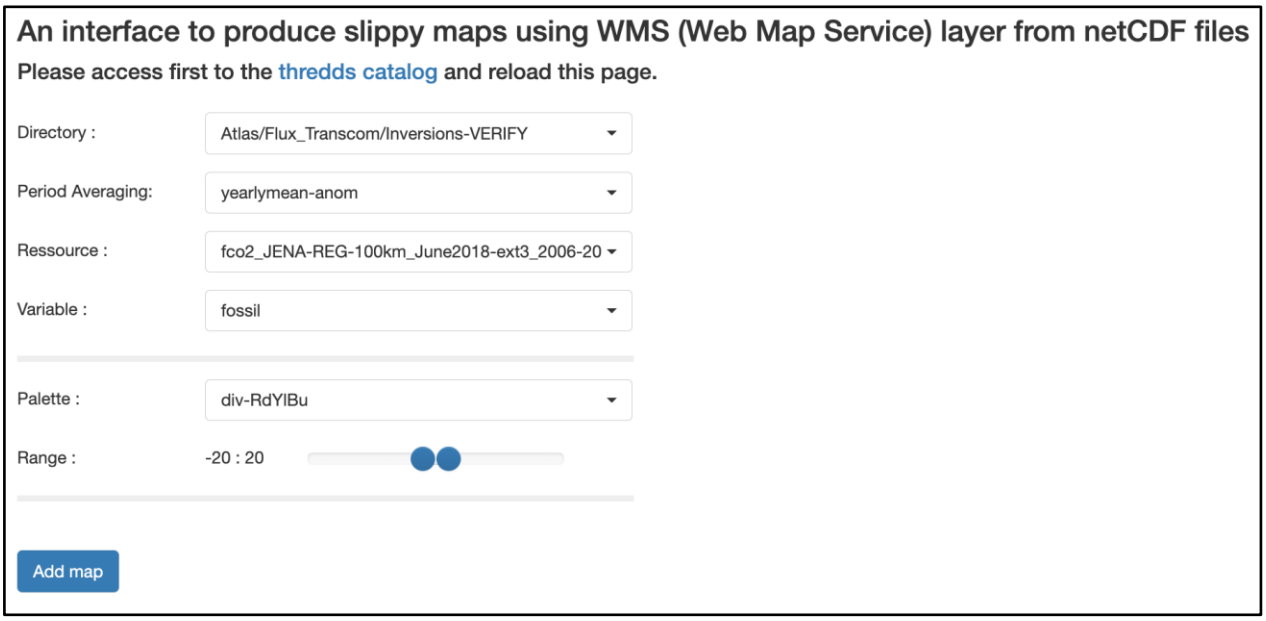


Figure 9: Illustration of the menu obtained from the map viewer, when first entering.

The map that has been produced is interactive:

- The viewing windows can be dynamically resized, zoomed in, and closed down according to the user's whims.
- The geographical area that is visualized can be changed dynamically, moving (with the "moose") the region of interest and zooming in and out.
- The different time steps associated with the selected variable can be visualized: when the mouse is on one map a small menu appears at the bottom of the map allowing to browse through the available time steps.
- The "range" and "palette" options are also dynamically linked with the map and thus can be changed interactively.

The menu for selecting additional products is available below the previously selected maps and the user can choose any other additional product (illustration with two products in figure 10). A few features (being currently tested) will be added before the end of the project:

- A synchronization of the different maps: all zooming and time stepping action will be synchronized between the different maps.
- A export function to save the selected maps as a "png" or "pdf" file

Note finally that a specific tutorial will be added (link to be put at the top of the page, see figure 8) through a video explaining how to use the map viewer.

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

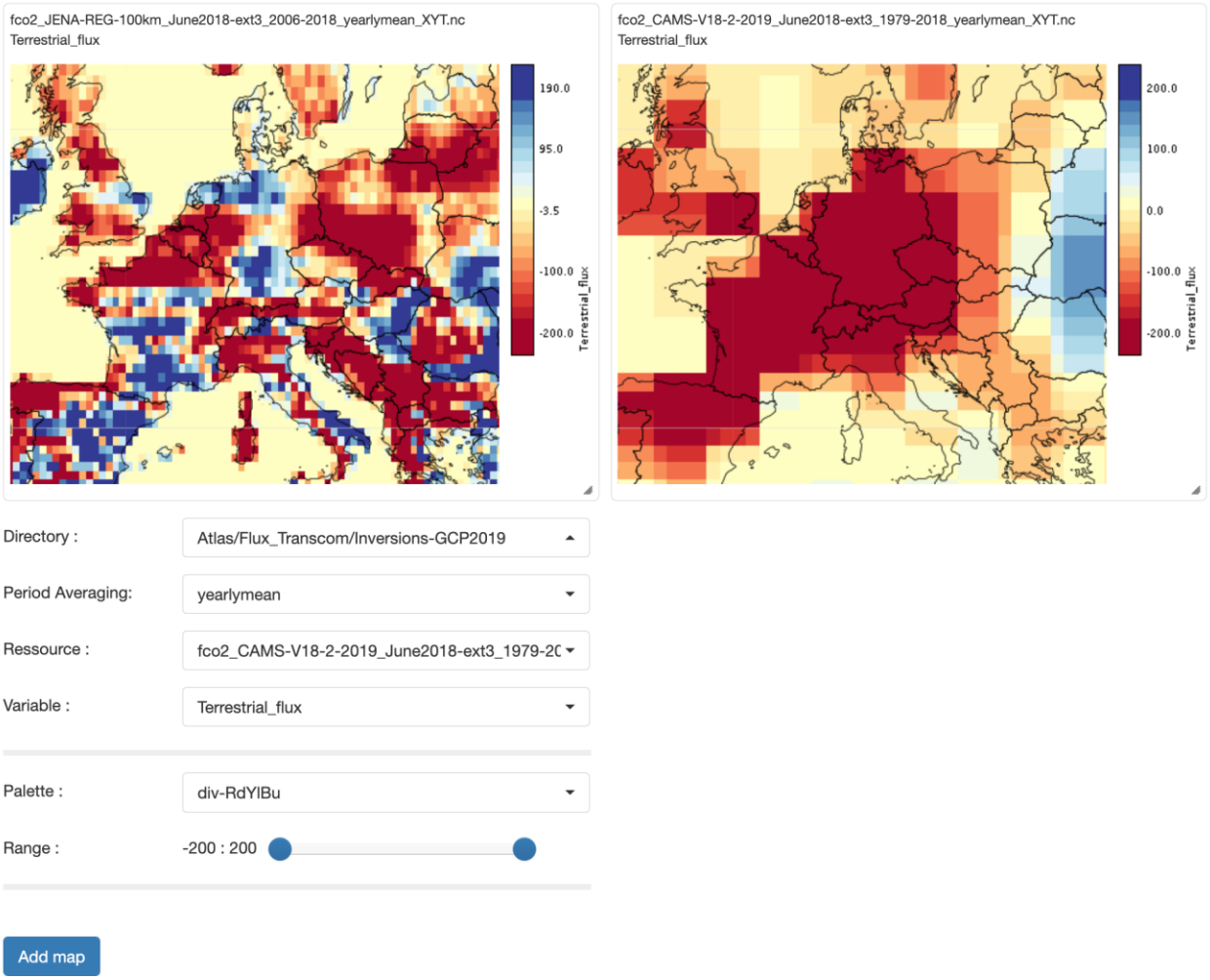


Figure 10: Illustration of the map viewer with two selected products for the terrestrial CO2 flux (the Regional inversion from VERIFY and the global inversion from CAMS).

8. GHG synthesis plots and summary fact-sheets

This section describes a key feature of the VERIFY project, linked to the production of the GHG synthesis at country scale. These synthetic plots summarize the core work of several work packages, WP1-2-3-4 with a comparison of the UNFCCC flux estimates (WP1) to the observation-based fluxes derived in WP2-3-4.

These plots, also used in the annual scientific synthesis report, correspond to the "Country GHG synthesis plots and summary factsheets" icon of the main products page (see Figure 2) or can also be access directly with the link: http://webportals.ipsl.jussieu.fr/VERIFY/FactSheets/. The tool presents several options to choose and display an ensemble of "pre-made" plots and summary factsheets. Figure 11 illustrates the main page for selecting the different synthetic plots and summary factsheets.

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

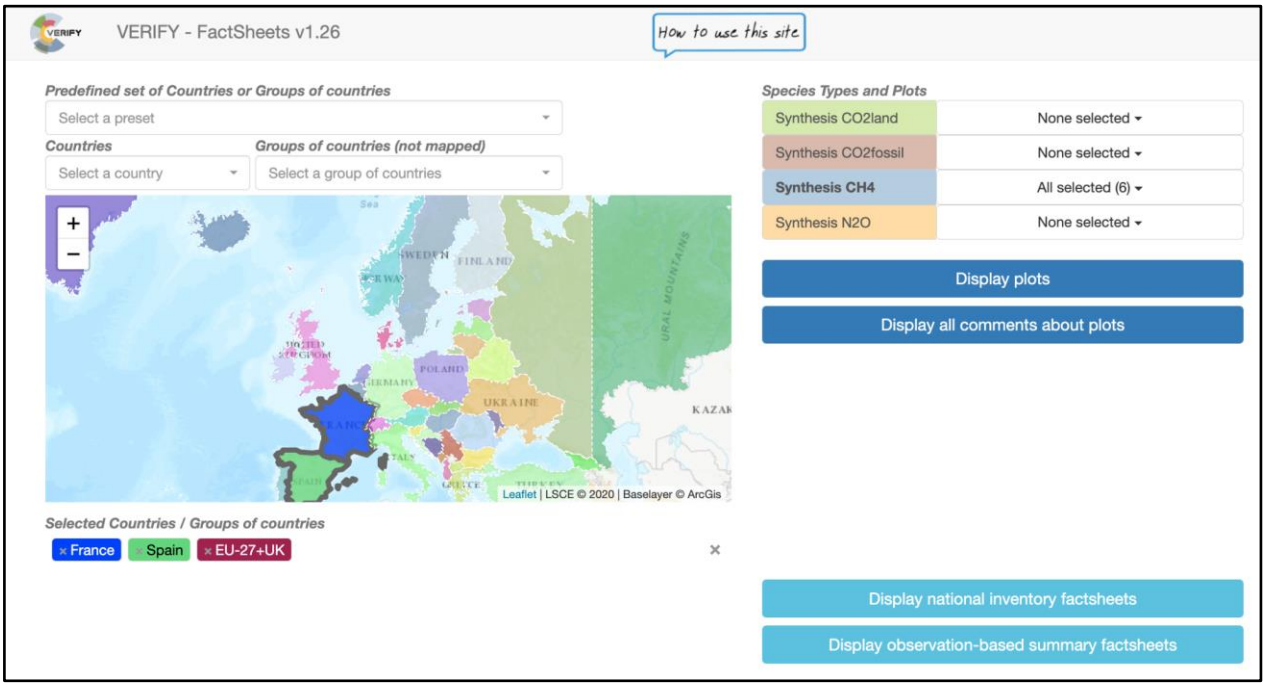


Figure 11: An example of the web tool that selects and displays synthetic plots of greenhouse gas fluxes.

The main page of this web-site has been updated since the first description and it now offers the following features:

- A map with possibilities of selecting an ensemble of countries by clicking on them. Note that the list of countries has been extended to include most countries in the world with now 231 regions (countries and groups of countries).
- Two additional drop-down menus, the top one to select a pre-defined set of regions and the bottom to select from a list different European countries or groups of countries.
- A menu on the top right to select the type of synthesis plots the user wishes to view: one for every species, except for CO_2, which is additionally broken down into “Natural” and “Anthropogenic” emissions. The different plots illustrate key concepts related to greenhouse gas fluxes, such as comparing bottom-up and top-down approaches, or comparing top-down approaches to national inventory methods.
- A button “Display plots” allows the user to view the selected synthesis plots for all selected regions.
- A button “Display comments about plots” will open a page with all plots that have been commented by any user.
- A button “Display national inventory factsheets” open a page containing a “pdf” file for each country that contains the synthetic factsheet produced from the national inventory fluxes (factsheets produced by WP1)

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

- A button “Display observation-based summary factsheets” opens a page containing a “pdf” file for each country that contains the observation-based summary factsheets produced by WP5 with a selected set of GHG synthesis plots.
- A specific button at the top of the page, “How to use this site”, provides a short video tutorial on how to use the different features of the site

When the “Display plots” button is selected a new page opens to display all plots. Figure 12 provides an example for one region (EU27+UK) of the type of plots that can be obtained for each GHG. Each plot has several icons on the top in order to:

- View short description (caption) of each type of plot (clicking on the small “I” icon)
- View the legend that explains the acronyms used across all the plots (clicking on the small “?” icon)
-
- Close the specific plot (clicking on the small “X” icon)

Additionally, the plots can be rearranged, resized, and deleted in order to facilitate viewing. By “double clicking” on a plot you open a larger view of the selected plot. With the large viewing mode additional options appears on the top of the window (see figure 13):

- Access the raw data used to create the plot (clicking on the small “grid” icon)
- Make a comment to the specific plot (clicking on the “bubble” icon); commenting all figures provides a new way to collectively share information / views associated with each country's GHG synthesis plot.
- Change version of the plot, clicking on one of the different icons: “V2019”, “V2020”, and “V2021”; these different versions correspond to the GHG synthesis that were made respectively in 2019, 2020 and 2021, respectively. Note that a final synthesis will be made in 2022 together with CoCO2.
- Browse through the different selected plots with the two arrows.

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

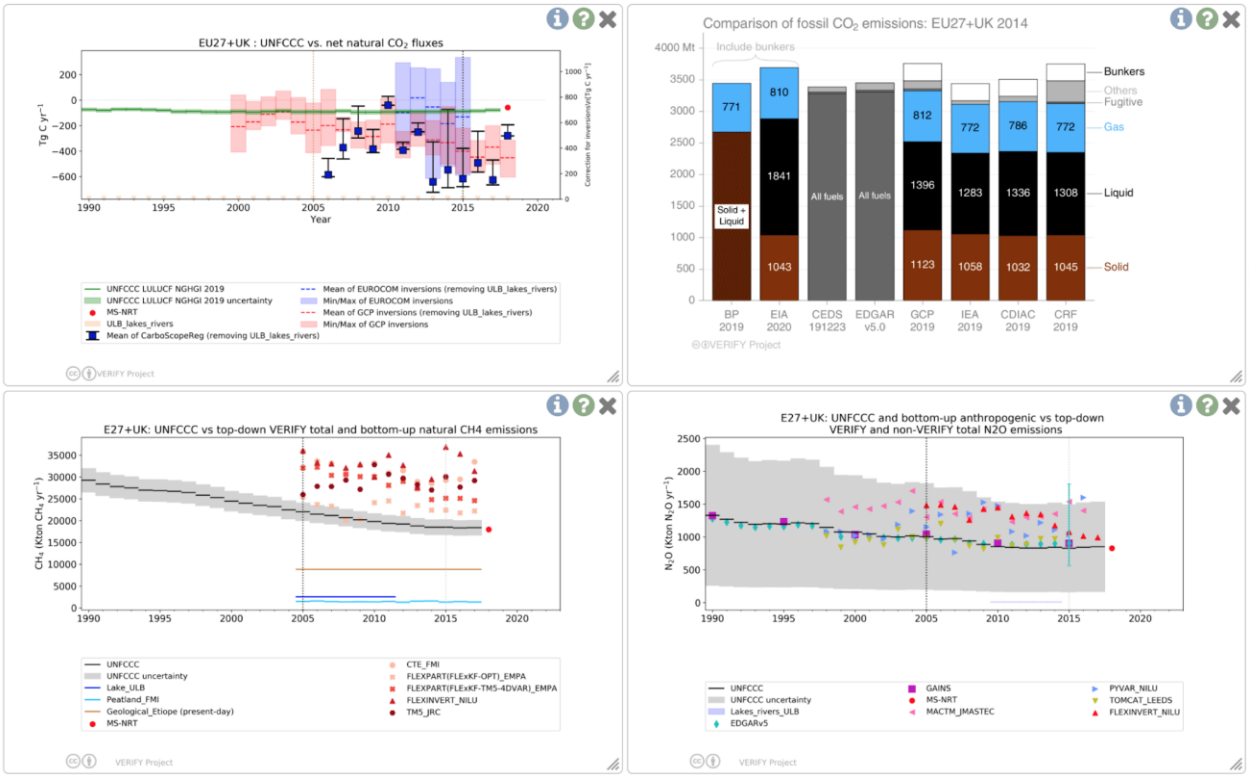


Figure 12: Illustration for one region (EU27 + UK) of typical synthesis plot for CO_2 natural (upper left), CO_2 anthropogenic (upper right), CH_4 (lower left) and N_2O (lower right). Each plot provides a synthesis combining UNFCCC flux estimates and other data-driven flux estimates.

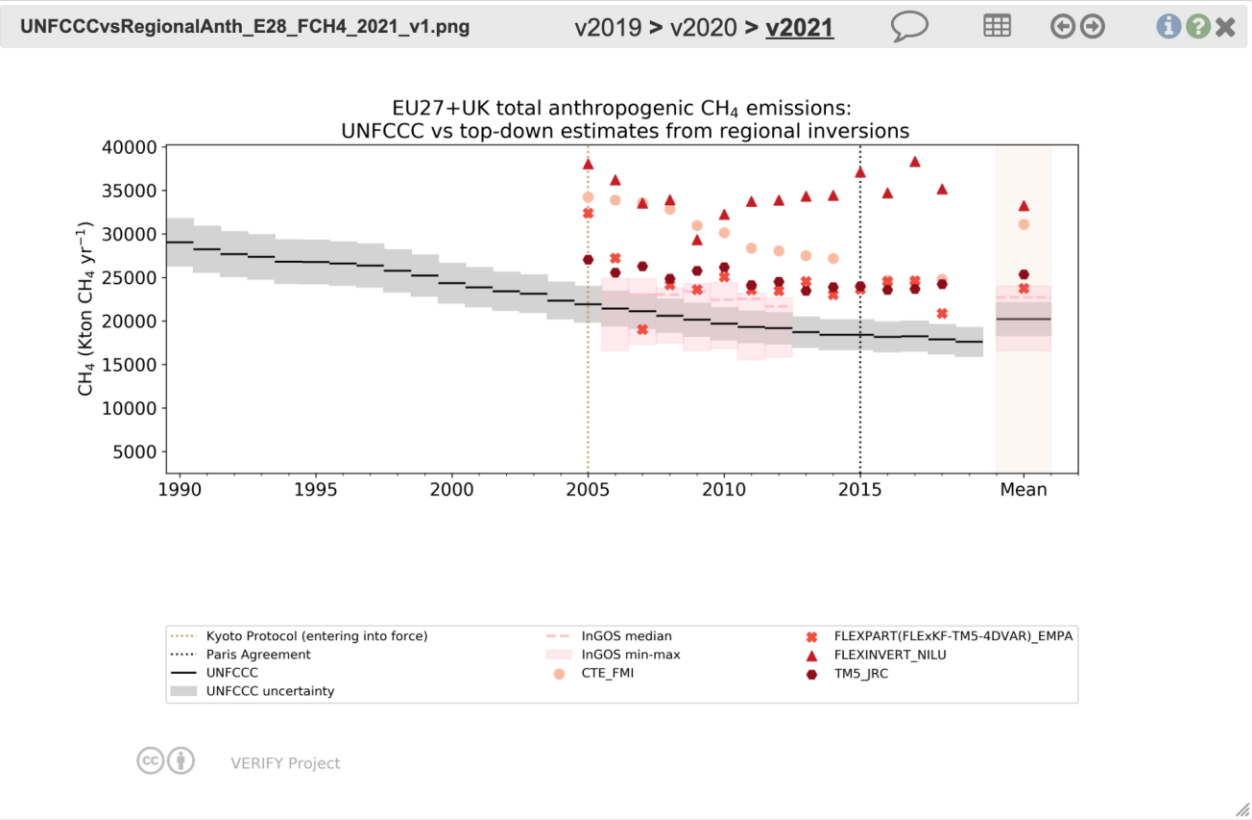


Figure 13: Illustration for one region (EU27 + UK) of a typical synthesis plot for CH4 emission, comparing the UNFCCC reported emissions with different atmospheric inversions.

9. Conclusion and perspectives

This deliverable provides the final description of the database and data management infrastructure, as implemented on the CEA-LSCE servers. The main infrastructure relies on a Thredds Data Server (TDS) and offers not only the access to the main data sets but also some visualization facilities. Our infrastructure permits us to meet most requirements of the FAIR policy with this implementation:

- **Findable:** Our dataset is directly findable through a dedicated web portal that offers a catalog with searching capabilities. The next step is to provide a DOI (Digital Object Identifier) for all key data sets of VERIFY (i.e. the main flux estimates). These steps will be done in collaboration with ICOS, during the migration of the key datasets to the ICOS carbon portal (during the second half of 2022).

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

- **Accessible:** All datasets are accessible through a standard protocol widely used to share data across the climate scientific community: a Thredds Data Server (TDS, managed by unidata) designed to work with the NetCDF data format of files and offering a variety of services (WCS, NCSS, WMS, etc).

- **Interoperable:** Most output from the VERIFY project tries to follow international standards associated with NetCDF files. The metadata output tries to follow the standard NetCDF Climate and Forecast (CF) Metadata convention, using the standard variable names, units, dimensions, axis, required 'coordinates' attribute, etc.

- **Re-usable:** The main data sets that have been produced by the VERIFY project will be kept available and accessible after the duration of the project through the ICOS carbon portal.

It should be noted that some VERIFY products were password-protected to ensure that VERIFY research groups have the first attempt at using them. Access to all raw data is now made available to everyone but still with a registration and the acceptance of a fair use policy. The visualization tools for the GHG synthesis plots and summary Factsheets, on the other hand, are not password protected.

Note finally that this last version of the deliverable provides several updates compared to first version including:

- An update of the database with a new structure to highlight more clearly the data set produced by VERIFY, including the different releases (V2019, V2020 and V2021).
- The addition of a new user-friendly mapping facility to compare the spatial gradient of the GHG fluxes between data sets.
- The addition of a few "Jupyter notebooks", especially notebooks that can be used to create the synthesis plots.
- Some changes to enhance the accessibility of the key GHG synthesis plots, Factsheets, and time-series visualization.
- The choice of the ICOS carbon portal to transfer the key datasets and part of the data infrastructure for a long-term preservation of the VERIFY outputs.

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