CO_2 Human Emissions

Protocol defining harmonized input & output datasets

Liesbeth Florentie,
Wouter Peters,
Maarten Krol

Co-ordinated by ECMWF

D1.1 Protocol defining harmonized input and output datasets for T1.1, T1.2 and T1.3

Dissemination Level:	Public
Author(s):	Liesbeth Florentie (WU), Wouter Peters (WU), Maarten Krol (WU)
Date:	25/09/2018
Version:	1.0
Contractual Delivery Date:	30/09/2018
Work Package/ Task:	WP1/ T1.1
Document Owner:	WU
Contributors:	WU
Status:	Final

CHE: CO2 Human Emissions Project

Coordination and Support Action (CSA)
H2020-EO-3-2017 Preparation for a European capacity to monitor CO2 anthropogenic emissions

Project Coordinator: Dr Gianpaolo Balsamo (ECMWF)
Project Start Date: 01/10/2017
Project Duration: 39 months

Published by the CHE Consortium

Contact:
ECMWF, Shinfield Park, Reading, RG2 9AX,
gianpaolo.balsamo@ecmwf.int

The CHE project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776186.

Table of Contents

1 Executive Summary

The current document outlines minimum requirements for input and output datasets as emerging from WP1 of the CO_2 Human Emissions (CHE) project. The aim of this protocol consists of harmonizing the different datasets to ensure compatibility and easy accessibility for subsequent integral analyses. Special attention is given to both the set-up and data format of CO_2 surface flux inversions, which will be used to assess the effectivity of proposed methodological innovations.

2 Introduction

2.1 Background

To improve monitoring of European anthropogenic CO_2 emissions, the CHE project targets methodological advances that include representation of anthropogenic CO_2 variability in space and time, and separation of anthropogenic emissions from biogenic fluxes at country to global scales. To this end it is desired to integrate observational data from both in situ measurements as well as remote sensing (i.e. 'top-down' information) with inventory information and modelling capacities for CO_2 fluxes and transport (i.e. 'bottom-up' data). Within WP1 of CHE, the effort is focussed on extending currently available capabilities, for example to make improved use of satellite observations. Facilitation of such integrated analyses, for example using inverse modelling approaches, requires compatibility and easy accessibility of the various data sets involved.

2.2 Scope of this deliverable

This document describes a protocol that aims to ensure harmonization of input datasets and output datasets for use in, and as resulting from, tasks T1.1, T1.2 and T1.3 of the CHE project.

2.2.1 Objectives of this deliverables

The objective of this protocol is to ensure compatibility and facilitate easy accessibility of the datasets involved in tasks T1.1, T1.2 and T1.3 of the CHE project. To this end minimum requirements and formatting guidelines are formulated for the involved data products.

Additionally, we aim to facilitate possible future continuation of the analyses performed in the context of WP1 of CHE by maximizing accessibility for external parties as well as consistency with other climate monitoring projects. To this end we adhere to current community's practices as much as possible.

2.2.2 Work performed in this deliverable

After consultation with the partners involved in tasks T1.1, T1.2 and T1.3, an overview is constructed that contains minimum requirements on the related datasets. Templates are constructed to facilitate easy compliance with these requirements. A selection of input datasets has been made available for general use, both within the consortium and for external parties.

2.2.3 Deviations and counter measures

Not applicable.

3 Input datasets

Within WP1, a selection of data products for use as input in integrated analyses is provided/generated. This selection contains datasets originating from a wide variety of sources, including remote sensing, in situ measurements and inventory products. For newly generated products, specific minimum requirements hold depending on the type of data considered.

3.1 Specific requirements

All datafiles are to be provided in netCDF4 format. In addition, specific requirements depending on the type of input data hold, as discussed below.

3.1.1 Description/metadata

All input datasets are to be provided accompanied by a description file which, apart from product relevant information, should contain at least the following metadata:

- Filename
- Contributor
- Description (including the source of the data)
- Date of generation
- Spatial coverage and resolution
- Temporal coverage and resolution

3.1.2 Remotely-sensed XCO_2

Column-averaged dry-air mole fractions of CO_2, i.e. XCO_2, originating from satellite measurements are to be provided in parts per million dry air mole fraction (ppm), including a specification of the used column averaging kernel. The data should conform with the greenhouse gas climate change initiative (GHG-CCI) convention, the specific requirements for which can be found in [1].

3.1.3 Fossil fuel emissions

Grid-level fossil-fuel emission products are requested to comply with the annual global total values reported by the global carbon project (GCP), which can be found in [2] or downloaded from the integrated carbon observation system (ICOS) carbon portal website via https://meta.icos-cp.eu/objects/mtuoxtXq4VhQaZmS4hPJuoQZ. They should be provided on a resolution of at least 1x1 degree, with unit kgC/m^2/s.

3.2 Provided datasets

An overview of the input datasets that will at least be made available within the framework of tasks T1.1, T1.2 and T1.3 of the CHE project (but not limited to) is provided below. The datasets should be provided open access, the latest after publication of the corresponding research study. Links to the datasets will be put on the CHE data portal (https://www.che-project.eu/data-portal), alongside with links to external input datasets.

- Gridded fossil-fuel emissions (Wageningen University):
    - Monthly CO_2 emissions as resulting from the burning of fossil fuels, reported in kgC/m^2/s. The product has global coverage with a spatial resolution of

1x1degree, monthly temporal resolution, and it covers the years from 2000 till 2017.
    - This data product was created for use with CarbonTracker Europe in order to perform inversion analyses for the 2018 global carbon budget. Original emissions for the years 2000-2012 are as resulting from the GEOCARBON/CARBONES ('30-year re-analysis of CARBON fluxES and pools over Europe and the globe') project, under the EU-funded 7th Framework Program (contract number 242316). The fossil-fuel emission fluxes for subsequent years are adjusted by extrapolating and scaling the regional trends such that the annual global totals are conform the values reported by the GCP [2].

- OCO-2 XCO_2 (University of Bremen & University of Leicester)
    - Re-analysis products based on the official U.S. National Aeronautics and Space Administration (NASA) v8 remote-sensing product of the Orbiting Carbon Observatory 2 (OCO-2) satellite.
    - Global coverage, temporal coverage at least year 2015

In addition, links are provided to external datasets that are considered relevant for the integrated analyses that shall be performed within the context of CHE. These include:

- Observation package 'GLOBALVIEWplus v4.0', provided by the U.S. National Oceanic and Atmospheric Administration (NOAA), (https://www.esrl.noaa.gov/gmd/ccgg/obspack/data.php)
- Level 2 native OCO-2 remote sensing product, version 8, provided by NASA, (https://disc.gsfc.nasa.gov/datasets?processingLevel=2&project=OCO)
- Solar-Induced Fluorescence (SIF) from GOME-2A and GOME-2B, provided by the Koninklijk Nederlands Meteorologisch Instituut (KNMI), (http://www.temis.nl/surface/sif.html)

4 CO_2 flux inversion results

Output datasets containing carbon fluxes as resulting from inversion analyses are to be generated in line with the following requirements.

4.1 Inverse analysis requirements

4.1.1 Required runs

The focus of WP1 of the CHE project is to investigate innovations in CO_2 flux inversion analyses, with a main focus on the use of remote sensing data. To allow studying the feasibility of new approaches (e.g. to identify both the effect of remote sensing data on the inversion results, and future requirements on remote sensing products), inversion output datasets should at least include results for the following runs:

A. **Baseline run** using currently best set-up and according to specifications in section 4.1.2
B. **Innovation run** which adds a single innovation to the baseline run (e.g. the use of XCO_2 satellite data)

If you want to introduce multiple innovations, please do so sequentially, so using one innovation run as baseline for another innovation.

To facilitate an intercomparison study, all inversion results should include posterior fluxes for the year 2015.

4.1.2 Input data for inversion analyses

Parties are free to use their preferred set-up for the inversions, with the general input datasets discussed in section 3 provided for general purpose.

For ease of intercomparison, however, the flux inversion output datasets generated in the context of WP1 are requested to comply with the following minimum requirements w.r.t. the used input datasets:

- **CO_2 Observation package: GLOBALVIEWplus v4.0 dataset** (obspack_co2_1_GLOBALVIEWplus_v4.0_2018-08-02) as provided by NOAA via their website (https://www.esrl.noaa.gov/gmd/ccgg/obspack/data.php). The choice of which stations to assimilate is left free. Note, however, that this ObsPack also contains aircraft data, which shall be used in a later stadium for validation purposes. Therefore please make sure to not include all aircraft data in the assimilation.

- **Fossil fuel emissions:** The choice for a specific grid-level fossil-fuel-emission product is left free, however, annual global totals should be in line with the values reported by the GCP.

Input datasets not discussed here, such a prior biosphere fluxes and their uncertainties, are free to choose as fits.

4.2 Minimum requirements for the output datasets

Output files are to be created as netCDF4 files, with variables names conforming the CF (Climate and Forecasting) convention (http://cfconventions.org/) where applicable. For each inversion analysis, please provide simulated CO_2 mole fractions of the assimilated and non-assimilated data in addition to the optimized surface fluxes.

4.2.1 Description/metadata

All output datasets are to be provided accompanied from a description file which contains at least the metadata as specified in section 3.1.1, and a description of the innovation introduced in the innovation run.

Inversion output datasets should be provided globally (unless the inversion set-up does not allow for global results), with posterior fluxes on a 1x1 degree grid and with monthly time resolution. All output datasets should include posterior fluxes for the year 2015.

4.2.2 Mole fractions

Simulated CO_2 mole fractions as resulting from the inversion analysis are to be reported for all locations in the ObsPack, so including both the assimilated data and independent locations. It should be indicated for every location whether or not the location was included in the assimilation. This can be done either in a separate text file, or by copying the ObsPack files and adding the additional data as separate variables.

4.2.3 Surface fluxes

An empty netCDF4 file which is to be filled with at least the below variables is available on the CHE data portal. A detailed overview of the included variables can be found in Appendix A.

- **Prior and posterior monthly averaged total CO_2 surface fluxes**, including fossil fuel emissions, on a 1x1 degree grid (ranging from -180° to 180° degrees East longitude, and from -90° to 90° degrees North latitude), in unit mol/m^2/s.

- **Posterior uncertainty covariance matrix for the total CO_2 fluxes, for the TRANSCOM regions**, on a monthly basis for at least the year 2015.

- General variables:
  - **Date**: the time-interval center, as days since 2000-01-01 00:00:00 UTC.
  - **Latitude**: center of the grid cell, as degrees North
  - **Longitude**: center of the grid cell, as degrees East
  - **Cell area:** horizontal area of a gridcell
  - **Land area fraction:** fraction of horizontal area occupied by land surface
  - **Region mask:** numerical mask to aggregate TRANSCOM regions

Additionally, groups are encouraged to also report the breakdown of their calculated fluxes by separately reporting fossil fuel, fire, ocean and biosphere fluxes, in a similar fashion as for the total fluxes. These variables are also included in the empty netCDF4 template file.

4.3 Publication of the output datasets

Links to the datasets are to be shared on the CHE data portal, with each partner preferably arranging storage of the actual data themselves. For small datasets storage on the CHE data portal can be arranged if desired.

All partners participating in the CHE project agree to share the data with the other partners involved in the project, and to make the data publicly available, the latest after publication of the related research study.

5 Conclusion

The current document provides an overview of minimum requirements for input and output products to be generated in the scope of WP1 of the CHE project, with the aim to facilitate subsequent integral analyses and intercomparison studies. This protocol is constructed after consultation with the involved partners, and includes both format guidelines as well as set-up specification for CO_2 flux inversion runs.

6 References

[1] ESA Climate Office, Data Standards Requirements for CCI Data Producers, CCI-PRGM-EOPS-TN-13-0009, issue 1, revision2, 09/03/2015 (http://cci.esa.int/sites/default/files/CCI_Data_Requirements_Iss1.2_Mar2015.pdf)

[2] C. Le Quéré et al., Global Carbon Budget 2017, Earth Syst. Sci. Data, 10, 405–448, 2018 (https://doi.org/10.5194/essd-10-405-2018)

Appendix A

The netCDF4 template consists of:

- **Description file: CHE_inversion_fluxes.cdl**
  This file contains all variables and attributes, as well as latitude and longitude coordinates of the centers of 1x1 deg grid cells and a mask to identify the TRANSCOM regions. Please complete the global attributes of this file with the requested information.

- **Template netCDF4 file: CHE_inversion_fluxes.nc**
  This is just a template, as resulting from converting the description file by running
  `ncgen -o CHE_inversion_fluxes.nc CHE_inversion_fluxes.cdl`

- **Python script: CHE_fill_netCDF.py**
  This script can be adjusted to write local variables to the netCDF output file

The following dimensions, variables and attributes are included in the template:

- global attributes:
  - title
  - contributor
  - description
  - source
  - history
  - comment
  - references
  - project_ID = “CHE”
  - convention = “CF-1.6”

- dimensions:
  - lon = 360
  - lat = 180
  - time = UNLIMITED
  - regions_tc = 23

- general variables:
  - time (time)
  - lat (lat)
  - lon (lon)
  - region_mask (lat,lon)
  - cell_area (lat)
  - land_area_fraction (lat,lon)

- specific variables requested for CHE:
  - total_flux_prior (time,lat,lon)
  - total_flux_opt (time,lat,lon)
  - total_flux_prior_covariance (time,regions_tc,regions_tc)
  - total_flux_opt_covariance (time,regions_tc,regions_tc)

- optional variables desired for CHE:
  - bio_flux_prior (time,lat,lon)
  - bio_flux_opt (time,lat,lon)
  - bio_flux_prior_covariance (time,regions_tc,regions_tc)
  - bio_flux_opt_covariance (time,regions_tc,regions_tc)
  - land_flux_prior (time,lat,lon)
  - land_flux_opt (time,lat,lon)
  - land_flux_prior_covariance (time,regions_tc,regions_tc)
  - land_flux_opt_covariance (time,regions_tc,regions_tc)
  - oce_flux_prior (time,lat,lon)
  - oce_flux_opt (time,lat,lon)
  - oce_flux_prior_covariance (time,regions_tc,regions_tc)
  - oce_flux_opt_covariance (time,regions_tc,regions_tc)
  - fire_flux_imposed (time,lat,lon)
  - fossil_flux_imposed (time,lat,lon)

Document History

Version	Author(s)	Date	Changes
0.5	Liesbeth Florentie (WU)	07/09/2018	Original document
1.0	Liesbeth Florentie (WU)	26/09/2018	Implementation of internal review comments: • general wording, • changed 'flux covariance matrix' to 'posterior uncertainty covariance matrix' (section 4.2.3), • dimensions of cell_area changed from [lat,lon] to [lat]

Internal Review History

Internal Reviewers	Date	Comments
Andre Butz (DLR)	24/09/2018	Wording: • Check "XCO2" and "CO2" occurrences: either all or none with a subscript "2". • The unit kilogram writes "kg" not "Kg". • Check definition of acronyms at first occurrence (GHG, CCI, GCP, ...) • First sentence, section 4.2.3, has awkward fontsize. • Section 4.2.3, first bullet: Why are the fluxes in units "mol/m2/s" not "kg/m2/s"? The latter would be consistent with the emission units.
Thomas Kaminski & Michael Vossbeck (iLab)	25/09/2018	Good job colleagues! Just a few comments: • 3.1.1: "accompanied from" → "accompanied by"? • 3.1.2: "averaging kernels": probably you don't mean the full matrix but just the "column averaging kernels" • 3.1.3: Without scaling to GCP total you might get a more representative spread of results • 3.2: Will the gridded fossil fuel emissions field come with uncertainty ranges, so it can be used as prior for an inversion? • 4.1.2: It is good to keep the aircraft profiles for validation. Maybe it would be

		good to agree on a few sites that are also used for validation? • 4.2.1 (and thereafter): “inverted fluxes” is bad wording as the flux is not inverted (i.e. its sign was changed or so) but the posterior flux from an inversion, so you might want to call it “posterior flux” or “optimised flux” • 4.2.3: Maybe provided link to the empty netCDF4 file? • 4.2.3: The flux covariance can be computed from the fluxes, so maybe you don’t mean the “flux covariance matrix” but the “posterior uncertainty (or error) covariance matrix”? • Appendix A: “CH-1.6” → “CF-1.6”? • Appendix A: “cell area” should be independent of lon for the regular grid, i.e. you could save some space

Estimated Effort Contribution per Partner

Partner	Effort
WU	0.6
CEA	0.05
UB	0.05
UEA	0.05
ULEIC	0.05
Total	**0.8 Person/Months**