ECMWF COPERNICUS REPORT

Copernicus Atmosphere Monitoring Service

Description of the CH_4 Inversion Production Chain

Issued by: TNO / Arjo Segers

Date: 24/01/2020

Ref: CAMS73_2018SC1_D73.5.2.2-2019_202001_production_chain_v1

IMPLEMENTED BY ECMWF

Contributors

TNO
Arjo Segers

VU AMSTERDAM
Sander Houweling

1 Table of Contents

Introduction

This report describes the CH_4 Flux Inversion system used in CAMS.

The system is to a large extent similar to the system used by the JRC in the pilot projects (MACC, MACC-III, and MACC-III). For a description of the original system we refer to (P. Bergamaschi, Alexe, & Segers, 2014).

Compared to the pilot projects, the configuration has been adjusted related to increased resolution and use of latest satellite products. In summary, the main changes are:

1) Resolutions have been changed to either global 6°×4° (without local zoom) on 25 layers for a low-resolution product or global 3°×2° on 34 layers for a high-resolution product.
2) Time windows for the inversions have been extended to 1 year (coarse resolution) or 3 years (high resolution), plus 6-month spin-up and spin-down.
3) Increased update frequency of meteorological input (ERA-Interim), and use of archived convective fluxes.
4) A bias-correction between GOSAT XCH_4 columns and their TM5 representations is computed from the surface-only inversion, which is then used during the surface+satellite inversion.

The description of the inversion system starts with an overview of the 4D-var method (Chapter 1). This is followed by a description of the emission inventories (Chapter 2), the chemistry transport model TM5 and its configuration (Chapter 3), and the observation processing (Chapter 4). These components are finally combined in the actual production chain, which is described in Chapter 5.

1. Flux inversion using 4D-var

The CAMS flux inversions of CH_4 are based on the TM5-4DVAR inverse modelling system (P. Bergamaschi et al., 2010, 2013; Peter Bergamaschi et al., 2009; Meirink, Bergamaschi, & Krol, 2008). The inversion system estimates monthly global fields of CH_4 surface fluxes that, when provided as input to a chemistry-transport model, provide the best possible match between simulated and observed CH_4 concentrations. The various components and configuration of this system are described below. Notations follow Table 1 from Rayner, Michalak, & Chevallier (2019).

1.1 State vector

The inversion system for CH_4 fluxes estimates the optimal value of a state vector $\boldsymbol{x}$ that consists of emission deviation factors and initial concentrations. As described in Chapter 5, the initial concentrations are part of the state vector only in the low-resolution inversion but not in the high-resolution inversion that provides the end product. An overview of the elements of the state is given in Table 1; the associated properties are described in following subsections.

1.1.1 Emission deviation values

The emission deviations are the most essential part of the state vector: this is the main output of the whole system. The deviations are monthly time series of global 2D fields (longitude, latitude), for a number of four methane source category groups:

- wetlands;
- rice fields;
- biomass-burning;
- other sources (anthropogenic).

An element $\gamma_{i,j,t}$ for grid cell (i,j) and time t represent the deviation from the background emissions $E^b_{i,j,t}$; the actual emission $E_{i,j,t}$ is computed from:

$$E_{i,j,t}(\gamma_{i,j,t}) = \begin{cases} E^b_{i,j,t}\ exp(\gamma_{i,j,t}) & ,\gamma_{i,j,t} < 0 \\ E^b_{i,j,t}\ (1+\gamma_{i,j,t}) & ,\gamma_{i,j,t} \geq 0 \end{cases}$$

(1)

This *semi-exponential* formulation is needed to ensure that actual emissions are always positive, since $\gamma_{i,j,t}$ could also be strongly negative (decreased emissions).

1.1.2 Initial concentrations

The second part of the state vector is a 3D field with initial concentrations. As will be described later on, this part is not present in all 4D-var runs.

1.2 Background state

The background state $\boldsymbol{x}^b$ is the initial estimate. The '$\boldsymbol{x}^b$' column in Table 1 indicates the chosen values.

For the emission deviation part, the background values are simply zero, indicating no deviation from the emission inventories.
The background for the initial concentrations is taken from an available model run, preferably the result of a previous 4D-run such that concentrations are already in good agreement with observations.

1.3 Cost function

In the inversion system, the optimal state $\boldsymbol{x}$ is defined as the state where the following cost-function reaches a minimum value:

$$J(\boldsymbol{x}) = \frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{x}^b\right)^T \boldsymbol{B}^{-1}\left(\boldsymbol{x}-\boldsymbol{x}^b\right) + \frac{1}{2}(\boldsymbol{H}(\boldsymbol{x})-\boldsymbol{y})^T \boldsymbol{R}^{-1}(\boldsymbol{H}(\boldsymbol{x})-\boldsymbol{y})$$

(2)

In summary, the cost-function assigns a penalty when $\boldsymbol{x}$ differs from a background state $\boldsymbol{x}^b$, and when the simulation $\boldsymbol{H}(\boldsymbol{x})$ differs from the observations $\boldsymbol{y}$. The observation operator $\boldsymbol{H}$ consists of the chemistry-transport model that simulates concentrations given all kinds of input data (meteorological fields etc.) and the state estimate $\boldsymbol{x}$, and a sampling operator. The background covariance matrix $\boldsymbol{B}$ and the observation error covariance matrix $\boldsymbol{R}$ define the relative weight of the two penalties in the total sum. The covariance matrices are described in detail in section 1.7 and Chapter 4.

The actual definition of the various elements of the cost function is given in the following sections.

Table 1 - Elements of the 4D-Var state and associated properties.

state		**units**	**shape**	$\boldsymbol{x}^b$	**sigma**	**correlation length scales**		
main element	**sub element**					**horizontal (Gaussian)**	**vertical**	**temporal (exp.)**
emission deviation values (semi-exponential)	wetlands	1	(lon,lat,time)	0	100%	500 km	-	0
	rice	1	(lon,lat,time)	0	100%	500 km	-	0
	biomass burning	1	(lon,lat,time)	0	100%	500 km	-	0
	other	1	(lon,lat,time)	0	50%	500 km	-	9.5 month
initial concentrations		ppb	(lon,lat,layer)	first guess	NMC	500 km	NMC	-

1.4 Minimization algorithm

The state $\boldsymbol{x}$ for which Eq. (2) reaches a minimum is obtained using the M1QN3[1] algorithm (Gilbert & Lemaréchal, 1989). The minimum is searched for in an iterative procedure that provides a more optimal state in every step. The input for each step is formed by:

- the current state estimate $\boldsymbol{x}_i$;
- the evaluated costs $J(\boldsymbol{x}_i)$ for this estimate;
- the gradient $\nabla_x J(\boldsymbol{x}_i)$ (described below) of the cost function with respect to the elements of the state, evaluated in the current estimate.

The result of an iteration step is a new estimate $\boldsymbol{x}_{i+1}$ of the optimal state. The iteration process is stopped when convergence is reached.

1.5 Cost-function gradient

The minimization procedure requires evaluation of the gradient of the cost function towards the elements of the state. The gradient can be computed from:

$$\nabla_x J(\boldsymbol{x}) = \boldsymbol{B}^{-1}(\boldsymbol{x} - \boldsymbol{x}^b) + \boldsymbol{H}^T \boldsymbol{R}^{-1}(\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{y}) \quad (3)$$

In here, $\boldsymbol{H}^T$ represents the adjoint observation operator, which mainly consists of the adjoint of the chemistry-transport model towards the state variables. The input for the adjoint observation operator is the departure vector:

$$\boldsymbol{\delta y} = \boldsymbol{R}^{-1}(\boldsymbol{H}(\boldsymbol{x}) - \boldsymbol{y}) \quad (4)$$

which consists of the differences between observations $\boldsymbol{y}$ and their simulations $\boldsymbol{H}(\boldsymbol{x})$, weighted with the assumed observation representation error covariance $\boldsymbol{R}$.
The inverse $\boldsymbol{B}^{-1}$ of the background covariance is not explicitly computed; instead, a pre-conditioner is applied as described below.

1.6 Pre-conditioner

The pre-conditioner is a state transformation that is used to avoid computations with the full covariance matrix $\boldsymbol{B}$ and its inverse. The pre-conditioned state is defined as:

$$\boldsymbol{w} = \boldsymbol{B}^{-1/2}(\boldsymbol{x} - \boldsymbol{x}^b) \quad (5)$$

with the reverse transformation:

$$\boldsymbol{x} = \boldsymbol{x}^b + \boldsymbol{B}^{1/2}\boldsymbol{w} \quad (6)$$

Formulated in the new state, the minimization should be applied for the cost function:

[1] An update of the production is planned that uses a minimization algorithm based on the conjugate-gradient technique (Lanczos algorithm). This will also provide an estimate of the uncertainty in the end product.

$$J(w) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + \frac{1}{2}(\boldsymbol{H}(\boldsymbol{x}(\boldsymbol{w})) - \boldsymbol{y})^T\boldsymbol{R}^{-1}(\boldsymbol{H}(\boldsymbol{x}(\boldsymbol{w})) - \boldsymbol{y})$$

(7)

and gradient:

$$\boldsymbol{\nabla}_w J(\boldsymbol{w}) = \boldsymbol{B}^{1/2}\boldsymbol{\nabla}_x J(\boldsymbol{w}) = \boldsymbol{w} + \boldsymbol{B}^{T/2}\boldsymbol{H}^T\boldsymbol{\delta y}$$

(8)

The iteration loop of the optimization starts with $\boldsymbol{w} = \boldsymbol{o}$, which is the pre-conditioned equivalent of $\boldsymbol{x} = \boldsymbol{x}^b$.

1.7 Background covariance

The background covariance $\boldsymbol{B}$ statistically describes how the true state $\boldsymbol{x}$ deviates from the background value $\boldsymbol{x}^b$. In this application, it describes the uncertainty in emissions and in the initial concentrations. The full $\boldsymbol{B}$ matrix is not computed but is only present in the algorithm as a parameterization; the operations with $\boldsymbol{B}$ (or its inverse, or square root of the inverse) are implemented using these parameterizations.
The uncertainties in the emission deviation and the initial concentrations are assumed to be uncorrelated. The $\boldsymbol{B}$ matrix has therefore a block-diagonal form:

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_e & \boldsymbol{O} \\ \boldsymbol{O} & \boldsymbol{B}_c \end{pmatrix}$$

(9)

where $\boldsymbol{B}_e$ is the background covariance for the emission deviations, and $\boldsymbol{B}_c$ for the initial concentrations.

1.7.1 Background covariance for emission deviations

The emission deviations are assumed to be uncorrelated between the four categories; their covariance is therefore block-diagonal:

$$\boldsymbol{B}_e = \begin{pmatrix} \boldsymbol{B}_{e1} & \cdots & \boldsymbol{O} \\ \vdots & \ddots & \vdots \\ \boldsymbol{O} & \cdots & \boldsymbol{B}_{e4} \end{pmatrix}$$

(10)

Each of the four diagonal blocks describes the covariance in the monthly time series of 2D emission deviation fields. This covariance is parameterized as the product of a standard deviation and a correlation:

$$\boldsymbol{B}_{ei} = \boldsymbol{S}_{ei}\,\boldsymbol{C}_{ei}\,\boldsymbol{S}_{ei}$$

(11)

where the $\boldsymbol{S}$ matrix is a diagonal matrix formed from a standard deviation field, and the $\boldsymbol{C}$ matrix holds the correlations. The assumed standard deviations are listed in the 'sigma' column of Table 1. A standard deviation of 1.0 (100%) is assumed, except for category 'other' for which a standard deviation of 0.50 (50%) is assumed.

The correlations within the time series of emission deviation fields are assumed to be separable in a spatial (horizontal) and temporal component, and can therefore be written as a Kronecker product:

$$\boldsymbol{C}_{ei} = \boldsymbol{C}_{ei,h} \otimes \boldsymbol{C}_{ei,t}$$

(12)

The horizontal correlations are parameterized as being homogeneous (the same everywhere) and isotropic (independent of direction). The correlation between the emission deviations in two grid cells therefore only depends on the distance between the grid cells; here a Gaussian decay is assumed:

$$c_{h;i,j} = \exp\left(-\frac{1}{2}\left(\frac{d_{i,j}}{L}\right)^2\right)$$

(13)

where $d_{i,j}$ is the distance between the centers of cell *i* and *j* in km, and *L* is the correlation length scale. Table 1 shows the chosen length scales, which are in this application 500 km for each of the emission categories. For the wetlands, rice, and biomass burning categories, no temporal correlations are prescribed ($\boldsymbol{C}_t = \boldsymbol{I}$). Thus, for these categories large fluctuations in time are allowed, which is in line with the strong seasonal pattern in the corresponding emissions. For category 'other' (mainly anthropogenic), the deviations are assumed to be smoother in time, and therefore a temporal correlation is prescribed with an exponential decay that depends on the distance in months:

$$c_{t;m,n} = \exp\left(-\frac{|m-n|}{\tau}\right)$$

(14)

where *m* and *n* are the month numbers and τ is the temporal correlation scale of 9.5 months (Table 1).

1.7.2 Background covariance for initial concentrations

The background covariance for the initial concentrations is assumed to be separable in a horizontal and vertical component. It is written as a Kronecker product between a horizontal correlation and a vertical covariance:

$$\boldsymbol{B}_c = \boldsymbol{C}_{c,h} \otimes \boldsymbol{B}_{c,v}$$

(15)

Similar as for the emissions, the horizontal correlations are assumed to be homogeneous and isotropic, and are parameterized using a Gaussian decay:

$$c_{h;i,j} = \exp\left(-\frac{1}{2}\left(\frac{d_{i,j}}{L}\right)^2\right)$$

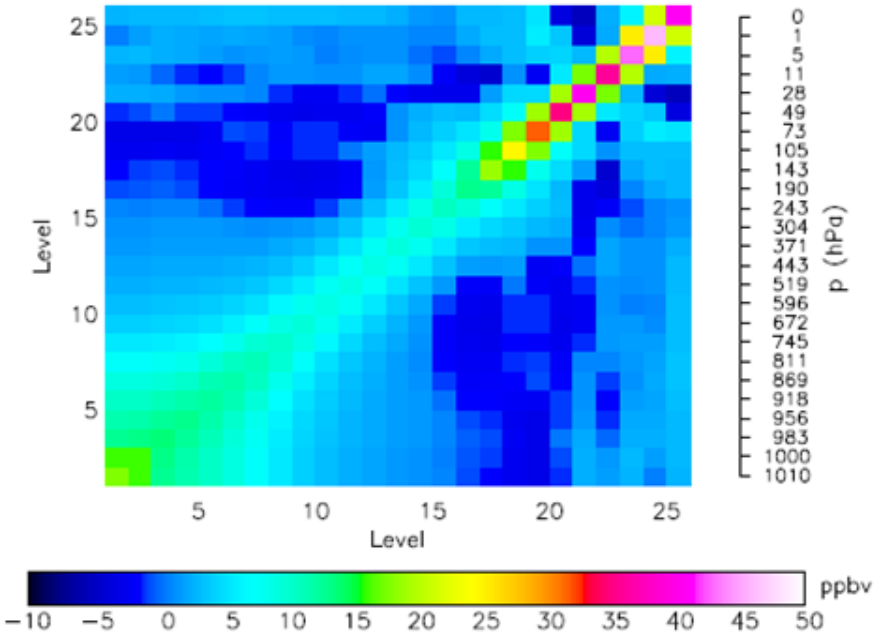
(16)

where $d_{i,j}$ is the distance between the centers of cell *i* and *j*, and *L*=500 km is the correlation length scale (Table 1).

The vertical covariance is shown in

Figure 1. This covariance was constructed in an earlier study (Meirink, Eskes, & Goede, 2006) by sampling a covariance from the difference between two model runs driven by different meteorological inputs (NMC method). Note that this vertical covariance is defined on the 25 vertical layers used for the coarse resolution version of theTM5 model; since in this project also a model version on 34 layers is used, the vertical covariance has been projected on the new layer definition too.

Figure 1 - Vertical covariance prescribed for the background covariance of the initial concentrations. Taken from Fig 1. in (Meirink, Eskes, and Goede 2006).



1.7.3 Eigenvalue decomposition

The actual operations involving the background covariance are actually not using the B matrix itself, but only (transposed) square roots. These are implemented using an eigenvalue decomposition of the correlation matrices; in a general notation:

$$\boldsymbol{C} = \boldsymbol{L}\,\boldsymbol{\Lambda}\,\boldsymbol{L}^{T}$$

(17)

where $\boldsymbol{\Lambda}$ is the diagonal matrix with (strictly positive) eigenvalues. In terms of these matrices, the transformation from pre-conditioned to normal state then becomes:

$$\boldsymbol{x} = \boldsymbol{x}^{b} + \boldsymbol{S}\,\boldsymbol{L}\,\boldsymbol{\Lambda}^{1/2}\,\boldsymbol{w}$$

(18)

and the pre-conditioned gradient:

$$\nabla_{w} J(\boldsymbol{w}) = \boldsymbol{w} + \boldsymbol{\Lambda}^{1/2}\boldsymbol{L}^{T}\,\boldsymbol{H}^{T}\boldsymbol{\delta y}$$

(19)

The expression of the covariances in block-diagonal matrices and Kronecker products remains for the eigenvalue decomposition. It is therefore sufficient to decompose only the individual horizontal and temporal correlation matrices, and the vertical covariance.

1.8 Observations

The observation vector $\boldsymbol{y}$ collects all measured values that should be used to optimize the state. The observations are collected for the time window of the 4D-var optimization, which is here either 24 or 48 months.

In this application, the observations are ground-based observations from the NOAA network, and satellite column-mean mixing ratios from the GOSAT satellite. The measurement network and the satellite data are described in more detail in Chapter 4 on observations.

1.9 Observation simulations

The observation operator $\boldsymbol{H}(\boldsymbol{x})$ simulates the observations in $\boldsymbol{y}$ given a state vector $\boldsymbol{x}$.

The core element of $\boldsymbol{H}$ is the chemistry transport model TM5, which is here used to simulate methane concentrations given an estimate of the actual emissions, initial concentrations, and other input data such as meteorology and methane sink rates. The properties and configuration of the model are described in more detail in Chapter 3.

The other element of $\boldsymbol{H}$ is the algorithm that computes an equivalent of an observation from the model fields. Typically, this consists of spatial interpolations from the model grid to the observation locations, aggregation of concentrations into total columns, and averaging in time. Since this strongly depends on the observation type, the detail will be discussed later in Chapter 4.

The observation representation error $\boldsymbol{R}$ describes the uncertainty in the observation representation. It should quantify how much the simulations in $\boldsymbol{H}(\boldsymbol{x})$ are expected to differ from the observations $\boldsymbol{y}$. Differences are typically due to instrumental errors and model representation errors that arise from the limited horizontal and vertical resolutions. The exact formulation of the observation representation errors is left for the description of the observation data.

2. Emissions

The end of the production chain is made of the optimized surface fluxes of CH_4. The optimized fluxes are deviations from a priori fluxes to provide better agreement with observations when used as input in the chemistry transport model TM5 described in Chapter 3.

2.1 Emission categories

The CH_4 fluxes will be estimated for four different emission categories:

- wetlands;
- rice fields;
- biomass burning;
- other sources, mainly anthropogenic.

These categories have a different spatial and/or temporal distribution such that the inversion system is able to distinguish their individual impact on concentrations.

2.2 Emission inventories

A priori estimates for the defined categories are composed from different inventories. Table 2 and Figure 2 illustrate which inventories are used throughout the inversion period.

2.2.1 Wetland emissions

Wetland emissions are taken from a simulation with LPJ-wsl (Zhang et al. 2018). This set provides emissions for 1990-2017, for earlier years the 1990 simulations are used too, and for later years the 2017 simulations.

2.2.2 Anthropogenic emissions

Rice fields and the 'other' anthropogenic sources are taken from the EDGAR v4.3 inventory (Janssens-Maenhout, 2017). This inventory provides emissions up to 2012. The inventory consists of gridded maps of yearly total emissions for a number of source categories. Two postprocessings are made to this set.

- For years after 2012, an extrapolation in time is made using yearly growth factors that are applied to the gridded fields valid for 2012. The growth factors are based on proxies for global fossil fuel and agricultural production, taken from BP and FAO respectively. If these statistics are not available for the most recent years, a linear extrapolation is used. An illustration of this procedure is shown in Figure 3.
- For the year 2010, the inventory also includes gridded maps per month. The relative time profile per grid cell for this year is used for all years of the inventory, including the temporal extrapolation for recent years.

Note that for rice fields the monthly time provided with the 2010 emissions was found to be incorrect, since it actually represents the activity of agricultural soil management (Margarita Choulga, personal communication). Therefore, for rice emissions the seasonality from "Matthews" that was used before in combination with EDGAR v4.2 also applied with the new inventory.

2.2.3 Biomass burning

Biomass burning emissions are taken from GFAS (Kaiser et al., 2012), as available from the CAMS fire emission service from 2003 onwards. For the 1990-2002 period the ACCMIP/MACCity inventory is used (Granier et al., 2011).

2.2.4 Remaining sources and soil sink

Climatologies are used for the remaining sources (oceans, wild animals, and termites), as well as for the soil sinks.

Table 2 - Overview of *a priori* emission inventories used for the CH_4 inversions. The colors represent the different emission super-categories that are optimized by the inversion.

Category	Period	Source
wetlands	*climatology*	Kaplan
rice	1989-2008	EDGAR v4.2 with Matthews seasonality
	2009-2017	EDGAR v4.2 with Matthews seasonality (valid for 2008)
biomass burning	1989-1996	ACCMIP-MACCity
	1997-2011	GFED v3.1
	2012-2017	GFED v3.1 climatology 1997-2010
other anthropogenic	1989-2008	EDGAR v4.2
	2009-2017	EDGAR v4.2 (valid for 2008)
oceans	*climatology*	Lambert
wild animals	*climatology*	Olson
soil sink	*climatology*	Ridgwell
termites	*climatology*	Sanderson

Figure 2 – Timeline of *a priori* emission inventories used during for the CH_4 inversions.

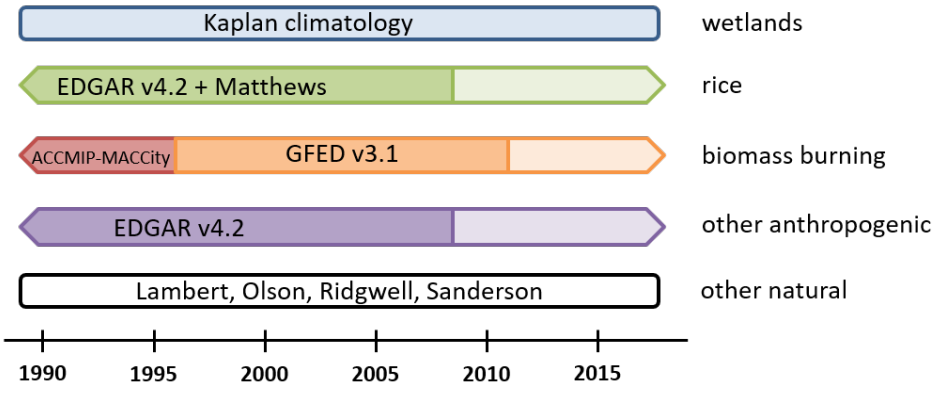
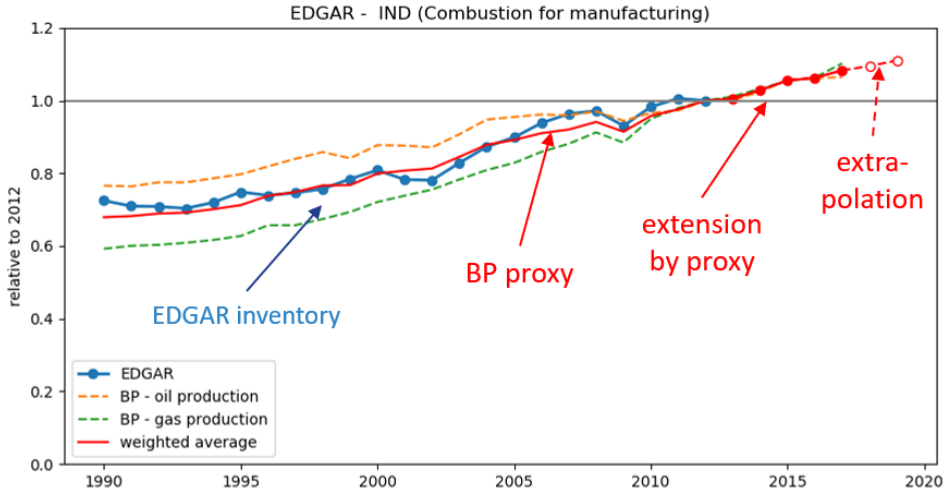


Figure 3 – Illustration of extrapolation of EDGAR v4.3 emissions using yearly growth factors.



2.3 Resolutions

The horizontal resolution of the inverted emission fluxes is globally 3°×2° (high resolution product, an intermediate inversion on 6°×4° is performed too), with monthly temporal resolution. These resolutions are bounded by the information content of the observations and the computational resources available; the a priori emission estimates might have a higher spatial (anthropogenic) or temporal resolution (biomass burning). The sub-monthly variations remain those of the a priori emissions.

2.4 Emission uncertainties

The optimization of emissions requires specification of uncertainties. In the current application, uncertainties are assigned to the emissions using *semi-exponential* factors as defined in Eq. (1). The amplitude and spatial and temporal correlations are listed in Table 1. In summary:

- For wetlands, rice, and bio-mass burning an uncertainty of 100% is defined; for the other sources (anthropogenic), this is 50%.
- Spatial correlations have a Gaussian-shaped decay with a correlation length scale of 500 km.

Uncertainties in the 'other' sources (anthropogenic) are assumed to be smooth in time, with a temporal correlation with exponential decay and a length scale of 9.5 months. Emissions in wetlands, rice, and biomass burning are allowed to vary from month to month, without temporal correlation.

3. Simulation model

The global atmospheric transport model used is TM5 (Huijnen et al., 2010; Krol et al., 2005). The components of the model relevant for the production chain are described in the following sections, except for the emission input that has been described in Chapter 2.

3.1 Horizontal resolution

The final product is obtained using a model version configured on a global 3°×2° horizontal grid, and 34 vertical layers (see below). Near the poles, the number of cells is reduced to avoid presence of small grid cells and associated requirement of small time steps.
An intermediate coarse-resolution product is defined on a global resolution of 6°×4° and 25 vertical layers. This grid is similar to what was used in the *pilot projects* (MACC), but without a local zoom.

3.2 Vertical layers

The vertical layers in the model are coarsened from the original 60 of the meteorological input data.

Figure 6 shows the distribution of the original layers (ERA-Interim, L60) and the model versions. Compared to the *pilot projects*, the product also includes a high(er) resolution result defined on 34 vertical layers, while an intermediate product on a coarse resolution with 25 layers is used. The high-resolution configuration (34 layers) coarsens the original layers especially in the stratosphere and the lower troposphere. The coarse-resolution configuration (25 layers) is a further coarsening of the 34 layers, with layers taken together in the upper troposphere too.

3.3 Meteorological input

The meteorological input for the model is taken from ECMWF's ERA5 re-analysis (Hersbach et al., 2018). Compared to the previous production chain based on ERA-Interim, the resolution of the meteorological data has increased from ~60 to ~30 km, and the number of layers has increased from 60 to 137; note that the TM5 model uses a lower resolution. The input frequency of the meteorological data is increased to hourly, which is only used for the surface field to limit storage and reading time of 3D fields; the fields are anyway interpolated to the model time step. The configuration of the meteorological input is summarized in Table 3.

Table 3 - Configuration of meteorological input.

Configuration	**Setting**
Originating center	ECMWF
Dataset	ERA-5
Horizontal resolution	~ 30 km, regridded to model resolutions
Number of layers	137, regridded to model levels
Datatype	Forecasts from 06:00 and 18:00 over 1-12 hours
Temporal resolution	1 hourly for surface fields and 3 hourly for model level fields, linear interpolation to model time step
Remarks	Data includes archived convective fields (updraft and downdraft fluxes, entrainment and detrainment rates)

The change in meteorological data from ERA-Interim to ERA-5 has been validated by model simulations of CH_4 over the entire 1990-2017 period. Figure 4 shows the observed and simulated concentrations in 2017 for South Pole station. The difference between the simulations with either ERA-Interim or ERA-5 is small, and anyway smaller than the difference with the observations. Similar small differences are seen in Figure 5 which shows a comparison between XCH_4 columns observed by the TCCON network and the simulations. In general, the latitudinal gradient is too strong, with simulations that are too high at the northern hemisphere and too low at the southern hemisphere. This gradient is slightly improved when using ERA-5, but remains present. From this and other comparisons it was concluded that the change in meteorological data was implemented correctly.

Figure 4 – Concentration of CH4 at South Pole station after 28 year of TM5 simulations with either ERA-Interim or ERA-5 meteorological data.

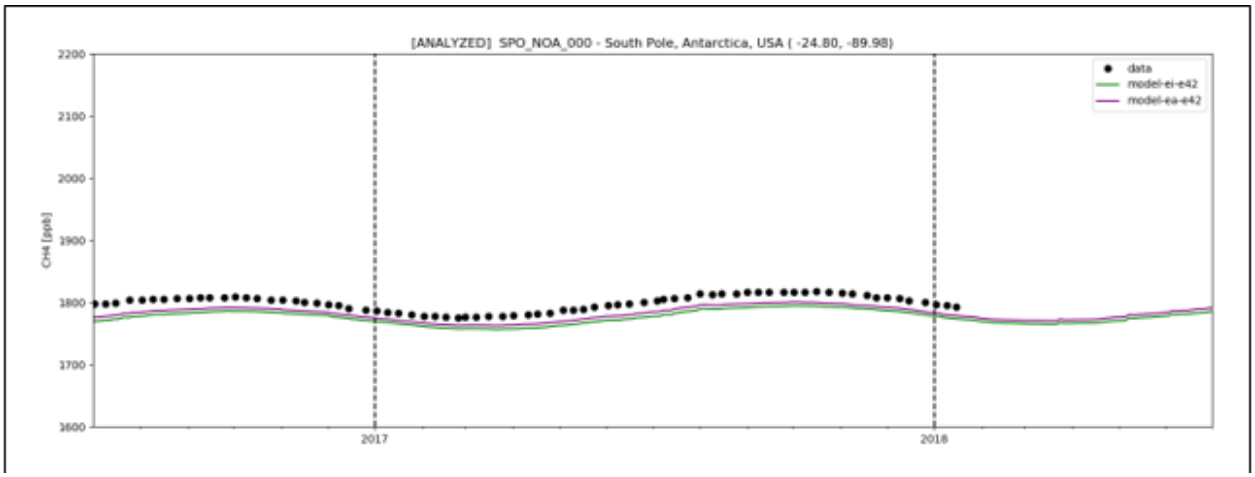
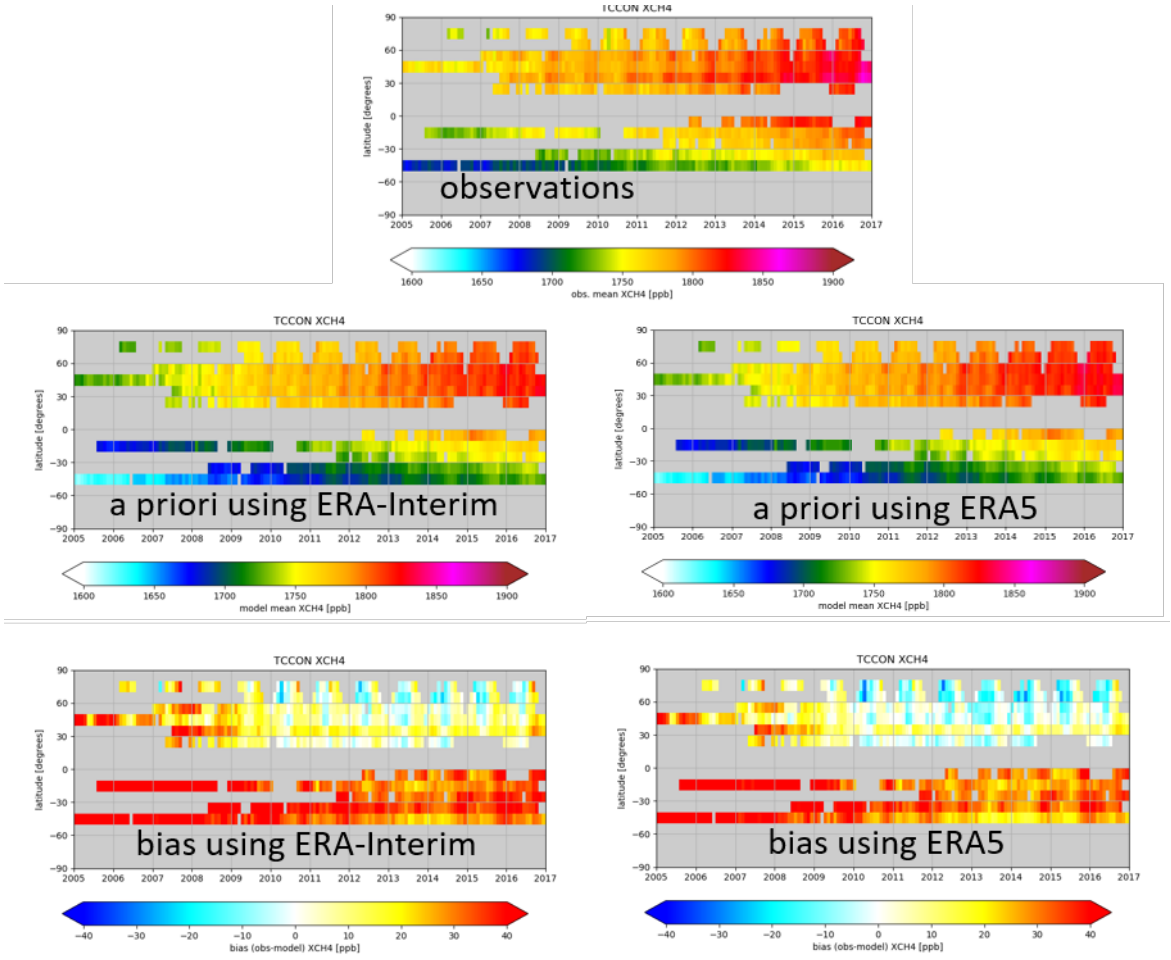


Figure 5 – Comparison between TCCON XCH_4 observations and TM5 simulations as function of latitude and time.

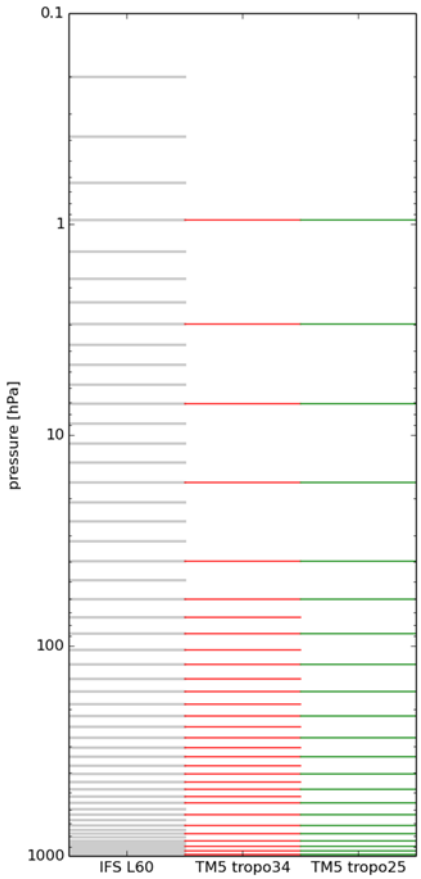


3.4 Chemical loss of CH_4

For the loss of CH_4 by chemistry the system uses loss rates computed off-line by full-chemistry transport models.

- For the tropospheric sink, loss rates are based on monthly OH fields simulated by the TM model, which were optimized to be in agreement with methyl chloroform (P. Bergamaschi et al., 2005; Houweling, Dentener, & Lelieveld, 1998).
- Stratospheric loss rates due to OH, Cl, and $O(^1D)$ are taken from ECHAM5/MESSy1 (Jöckel et al., 2006).

Figure 6 - Vertical layers in meteorological input (ERA-Interim, L60) and TM5 high resolution (34 layers) and coarse resolution (25 layers).



3.5 Dry air mixing ratios

To compare the model simulations with observations it is necessary to know how to interpret simulated mixing ratios with respect to dry or wet air. To ensure mass conservation, TM5 uses a constant global total air mass. This should be interpreted as a dry air mass, since the actual air mass is not constant but fluctuates with the water load.

4. Observations

The observation data used to optimize methane fluxes includes surface observations (NOAA) and satellite total columns (GOSAT). An additional set of independent observations is used for validation; these include surface and satellite observations that are not used in the inversion, total column observations, and aircraft observations.

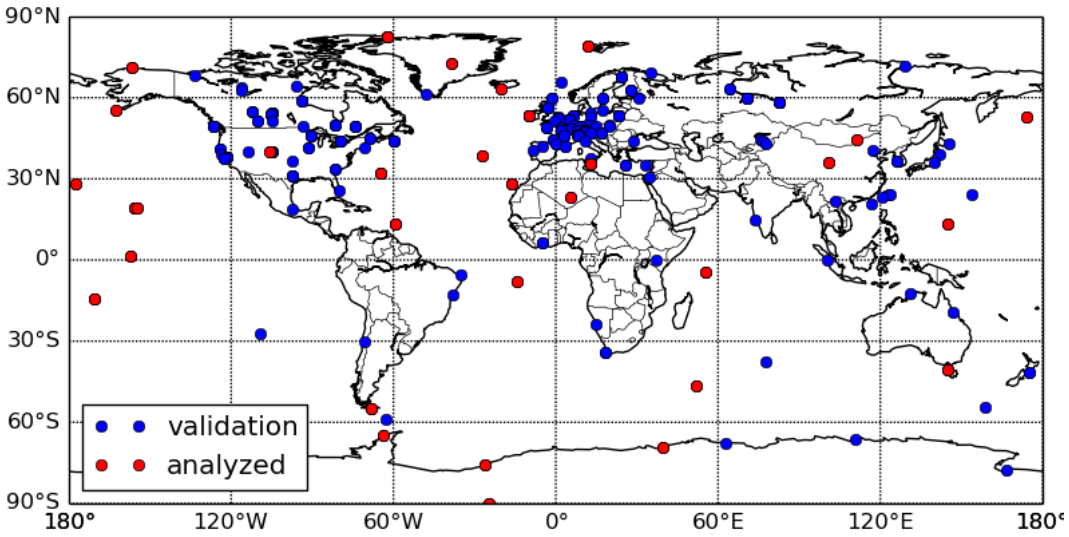
4.1 NOAA surface observations

Surface observations are available from the NOAA Earth System Research Laboratory (ESRL) global cooperative air sampling network (Dlugokencky et al., 2009). These concern CH_4 dry air mole fractions measured at the surface or on towers. The observations are acquired preliminary from the official release through personal communication with the providers, within 2-3 months after observation.

4.1.1 Station selection

The observation stations selected for this application are similar to those used in (P. Bergamaschi et al., 2013); Locations are shown in Figure 7. The selection excluded observation sites with significant local influence as well as observations from towers. In addition, sites are only selected if long time series without significant data gaps are available, to minimize temporal variations in the network geometry.

Figure 7 - Locations of ground based stations included in the inversion (analyzed) and other locations used for validation.



4.1.2 Observation pre-processing

A pre-processing is applied to the raw data in order to average the samples within 3-hourly windows. Multiple observations at these small temporal scales are present from sites with duplicate sampling or high temporal frequency.
For an estimate of the uncertainty, the NOAA product comes with an *analysis error* value; this is currently not used however, and a constant error standard deviation of 3 ppb is assigned to each (averaged) observation.

4.1.3 Observation representation

Surface observations are simulated in TM5 from (dry air) mixing ratios. In the horizontal, the concentrations are interpolated from the model grid to the station location using bi-linear interpolation. In the vertical, the concentrations are interpolated to the altitude of the site (in m above sea level); this is not necessarily the lowest model, since the model orography is at coarse resolution and therefore a smooth version of the true orography. In case the site is located below the model orography (e.g., in a valley), the concentrations at the surface are used.

The observation representation error is computed following (P. Bergamaschi et al., 2010). The base is the assumed measurement uncertainty of 3 ppb, with additional contributions for the model representation error based on 3-D gradients of simulated CH_4 mixing ratios, local emissions, and boundary layer height development.

4.2 Satellite XCH_4 columns

From 2009 onwards, satellite observations from GOSAT are available to constrain the surface fluxes too. The total columns XCH_4 from this instrument have been proven to be a useful source of information on methane concentrations (P. Bergamaschi et al., 2013). The inversion setup follows (Pandey et al., 2016) with respect to the treatment of XCH_4.

4.2.1 GOSAT data

For the optimization of the CH_4 fluxes, the production chain uses GOSAT XCH_4 columns. The product used is the RemoteC XCH_4 PROXY, as retrieved by SRON for the ESA/CCI[2] project (Detmers & Hasekamp, 2016). The PROXY product is based on the ratio between the CH_4 and CO_2 signal, and assume proper knowledge of the CO_2 field (Schepers et al., 2012).
The product includes a bias correction as function of the retrieved albedo, based on comparison of the retrievals with TCCON XCH_4 observations.
For the v18r1(s) releases data version v2.3.9 was used for 2009-2018; at production time, no data was available for the final 6-month spin-down in 2019.

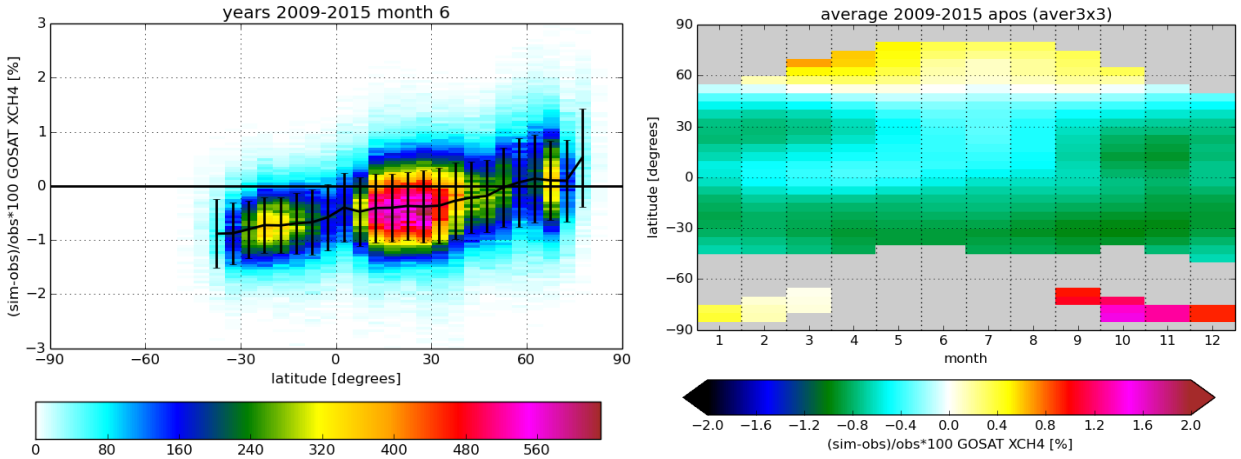
4.2.2 Observation representation

In the 4DVAR optimization, TM5 is sampled according to the time and location of each valid GOSAT retrieval, using the corresponding averaging kernel. Following (Pandey et al., 2016), a bias correction is applied to the GOSAT XCH_4. The correction is based on comparison of the TM5 inversion using NOAA surface observations only ("S1" stream) and the original retrieval product. It accounts for inconsistencies between inversions using in situ and satellite data, caused most likely by a combination of transport model and spectroscopic uncertainties. The current bias correction is computed per month and per latitude band of 5 degrees. The left panel of Figure 8 shows an example of the relative bias for the month of June, which shows that the TM5 model underestimates the

[2] www.esa-ghg-cci.org

GOSAT XCH_4 values towards southern latitudes. The right panel of the same figure shows the collected relative bias as a function of month and latitude band; the data from this panel is used for the bias correction. The bias correction is positive above 50N for each month with values reaching +0.5%; below 50N, the bias correction becomes negative with values down to minus 0.8%. At Antarctic latitudes the bias is strongly positive and is rather scattered; for this, all GOSAT retrievals below 60S are ignored.

Figure 8 - Illustration of bias correction between GOSAT XCH_4 columns and TM5 inversions from the "S1" stream (surface observations only). The left panel shows the relative bias as function of latitude as a density plot (number of samples per latitude/percentage box), computed for June over the years 2009-2015. The mean value per latitude band is computed if at least 500 samples are present for a band. The right panel shows the collected mean relative biases for all months, smoothed with a moving average over 3x3 values.



The uncertainty assigned to satellite retrievals in the inversion is a combination of the GOSAT retrieval uncertainty and a model representation error. The latter is calculated as the standard deviation of modeled XCH_4 in grid boxes surrounding the location of a single GOSAT sounding, calculated from a simulation using a priori emissions.

4.3 Validation data

Observations from other sources are available for validation. Here we briefly summarize the data sets that could be used.

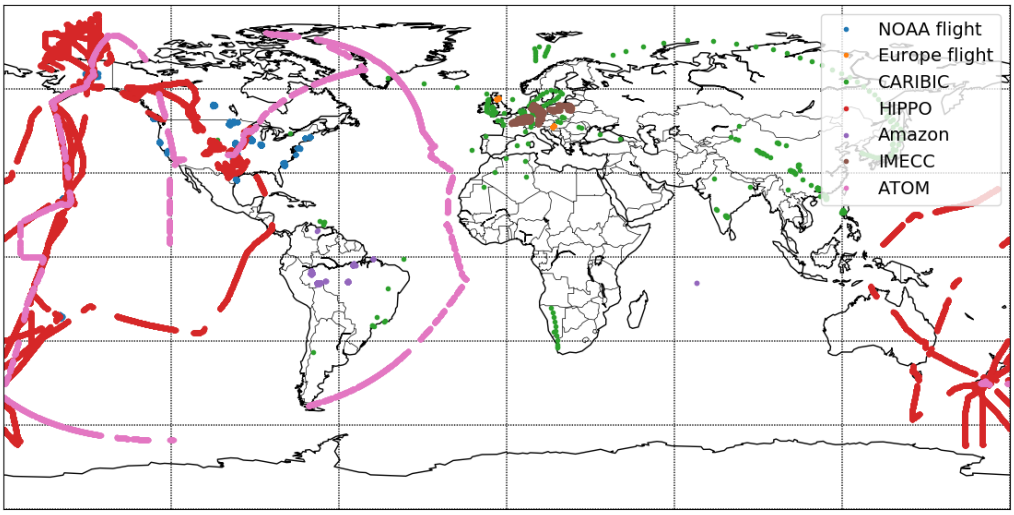
4.3.1 Surface observations

Apart from the NOAA observations used to constrain the CH_4 fluxes, additional observations are available that were excluded from the inversion. These include locations with significant local influence, as well as observations made at tower sites.

4.3.2 Aircraft observations

Aircraft measurements including CH_4 are available from various campaigns. For the current CH_4 inversions (reanalysis 2000-2014, extension with 2015), flight observations have been collected from various campaigns. For computation of the Key-Performance-Index-2 as defined in section 6.2 only observations for the free troposphere are needed, which is here defined as all observations taken between the boundary layer and an altitude of 10 km; locations where such data is available is shown in Figure 9..

Figure 9 - Locations of aircraft observations collected for validation of the CH_4 flux inversions.



4.3.3 Ship cruises

Ship cruises provide tracks of mixing ratios observations over the oceans. These observations are available from NOAA.

4.3.4 Ground based total columns

The Total Carbon Column Observing Network (TCCON) provides observations of the total CH_4 column (and other tracers) using ground-based Fourier Transform Spectrometer (FTS) instruments. The reported total columns are smoothed versions of the actual columns, since the instrument is less sensitive to concentrations in the top of the atmosphere. For comparison with model profiles it is therefore essential to take the averaging kernels provided with the product into account. Also the water content of the atmosphere should be taken into account. An extensive description of the how to compare the TCCON observations with model simulations is available on the TCCON wiki[3] and in (Wunch et al., 2010). The observed columns are expressed as mixing ratio's (XCH_4) in ppb.

[3] tccon-wiki.caltech.edu/Network_Policy/Data_Use_Policy/Auxiliary_Data

4.3.5 Satellite products

The following satellite products are available for validation purposes.

- For the years prior to 2012, total CH_4 columns of the SCIAMACHY instrument are available. These have been analyzed in the inversion system in the past, but were shown to require rather large bias corrections (Monteil et al., 2013), and have therefore not been considered for the inversion. However, for comparison with previously obtained results, it might still be informative to use this data in the validation phase.
- Tropospheric columns from the IASI satellite instrument (Razavi et al., 2009).

5. Operational system

5.1 Introduction

The inversion system is used to perform multi-year re-analyses for the period 1990-2018. To speed up the inversion, the multi-year time series is analyzed in blocks that can be processed in parallel. However, the number of blocks should not be too large to limit the number of transitions, and therefore rather long time windows are used. The target inversion on high resolution (processed in parallel) is preceded by a coarse resolution inversion that provides the proper initial conditions (processed serially).

The processing with coarse/high resolution inversions is performed for two different streams: the first stream (S1) only analyses surface observations, while the second stream (S2) also analyses satellite observations.

5.2 Coarse resolution inversions

The processing in different resolutions is illustrated in Figure 10. The first step in the processing is a coarse-resolution inversion (A), which provides initial concentrations for the high resolution run performed later on. The coarse resolution run is performed on $6^o \times 4^o$ and 25 layers. The inversion is split into optimizations that target at 1-year time windows, each with 6 months spin-up and spin-down. Each of these 24-month inversions optimize emissions but also the initial concentrations. The inversion is started from a first-guess initial concentration obtained from the optimized concentration from the previous year. The coarse-resolution inversions are therefore necessarily performed sequentially one after another, which is feasible due to the limited computation time needed at this resolution.

5.3 High-resolution inversion

The target flux is estimated from a high-resolution inversion ($3^o \times 2^o$, 34 layers) processed with long time windows of preferably 3 years. Similar as for the coarse-resolution inversion, a 6-month spin-up is used to minimize the impact of the initial conditions, and a 6-month spin-down constrains emissions in the 3-year target period with future observations. Observations are therefore collected for a window of 48 months.

The high-resolution inversions are processed in parallel, using initial concentrations optimized by the coarse resolution run. Only emissions are optimized in this step, leaving the initial concentrations fixed.

5.4 Stream 1: surface only

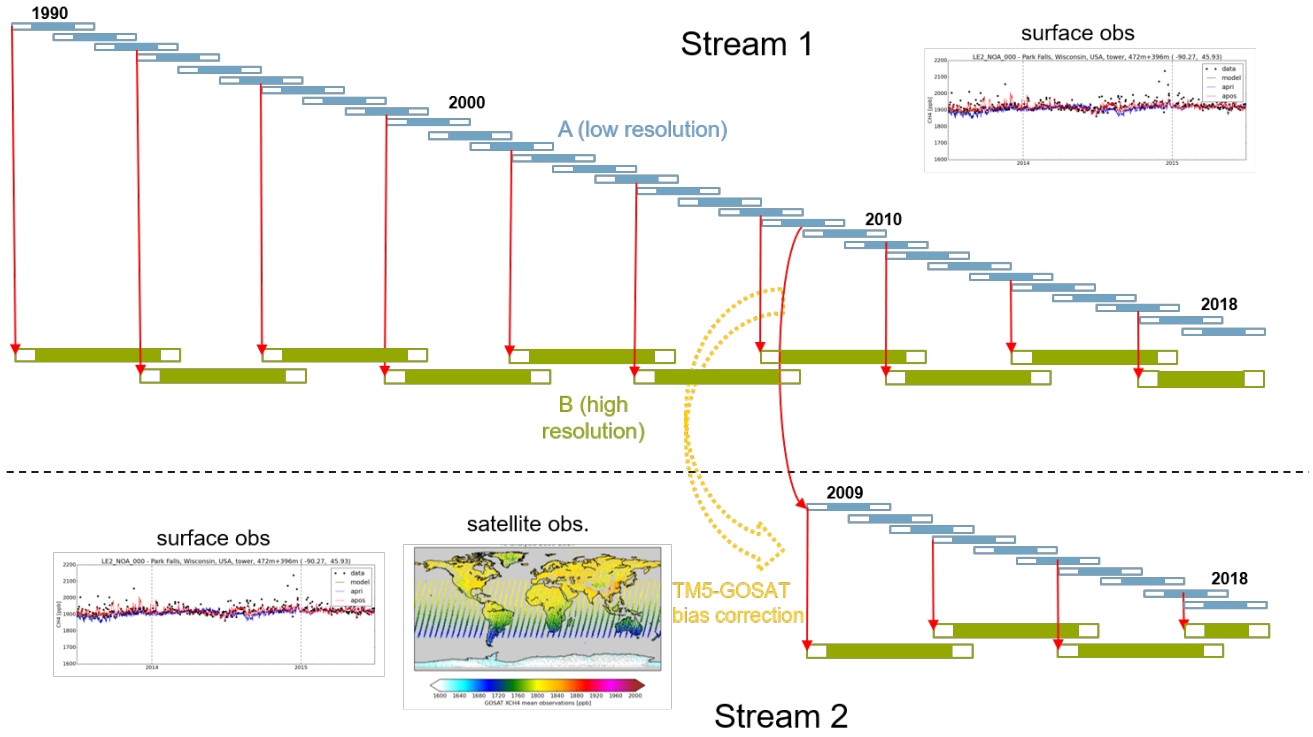
The first stream (S1) that provides optimized CH_4 fluxes only analyzes NOAA ground observations. These provide high quality long time series. A substantial amount of stations have time series from 1990 onwards. S1 uses these observations to contribute to the optimized CH_4 fluxes for the full period from 1990 until now. For the sake of consistency, the station selection for this stream is limited to sites with long time records available, although the number of observation sites in the network has increased with the years.

5.5 Stream 2: surface + satellite

The second production stream (S2) analyses not only NOAA surface observations but also satellite observations. The currently most suitable product comes from the GOSAT instrument that has been operational since 2009. The S2 production therefore covers the period 2009 onwards, with the first two 3-year high resolution inversions for the period 2009-2014. For the years prior to 2012, total CH_4 columns of the SCIAMACHY instrument are available too. These have been analyzed in the inversion system in the past, but were shown to require rather large bias corrections (Monteil et al., 2013); these have therefore not been considered for the inversion.

The first low-resolution inversion of the S2 stream (targeting 2009) is initialized with the same concentrations as used for the corresponding inversion window in the S1 stream. The S2 stream also uses a TM5-GOSAT bias correction (see subsection 4.2.2) computed from the S1 results; thus, the production of the S2 stream is only start after the S1 stream is finished.

Figure 10 - Illustration of the CH_4 inversion chain with streams using surface observations only (Stream 1) and surface plus satellite observations (Stream 2). The surface plus satellite stream is connected with the surface only stream by the first initial concentration and the TM5-GOSAT bias correction. Both streams use inversions at coarse resolution (A) and high resolution (B); the coarse resolution is performed sequentially to provide optimized initial concentrations for the high resolution inversion that is processed in parallel.



6. Key performance indices

The efficiency of this CAMS production chain, the scientific quality of its products, and the user up-take/satisfaction are characterized with key performance indicators (KPIs). In this chapter the relevant KPI's for the CH_4 inversion are described.

6.1 KPI: Delay in production time

One of KPI's is related to the timeliness of the production. For the CH_4 production chain the timeliness is determined by the time needed to release the inversion products and validation report, starting from the moment that all required input data is available. The following data needs to be acquired.

- Meteorological input data for the TM5 model. The production chain is based on meteorological data from ECMWF's ERA Interim archive (see Chapter 3), which is available after 2 months from the time for which it is valid. Since the inversion windows have a 6-month spin-down period, the required meteorological data is available in September.
- Emission inventories. As described in Chapter 2, most emission inventories are only available for periods more than 2 years ago. Inversions for the previous year are therefore simply using what is available, and no extra delay is involved with acquisition of new emissions.
- Observations that are used for the inversions are obtained directly from their providers (Chapter 4): NOAA for surface observations, and SRON for GOSAT retrievals. Both set of observations are expected to be available in the 3rd month after measuring. In view of the 6-month spin-down period of the inversions, the required observations are therefore available in September.

The acquisition of input data is therefore finished in the month of September following the target year of the inversion. Production of a single year of inversions is estimated to take 3 months, which leads to an expected delivery in November.

6.2 KPI: Bias and standard deviation in free troposphere

The KPI's related to the quality of the product can be summarized as:

> *"Mean absolute bias (standard deviation) between the posterior simulation and a large set of independent aircraft measurements in the free troposphere."*

The thresholds are currently set to 10 ppb for the absolute bias, and 20 ppb for the standard deviation.

Aircraft measurements including CH_4 are available from various past and ongoing campaigns, as described in section 4.3.2. The flight tracks of these campaigns are collected, and simulations are stored from the production chain. To compute these KPI's, a selection for the free troposphere is applied on the available aircraft measurements, keeping all measurements taken above the boundary layer height computed by ECMWF and below 10 km altitude. The bias and standard deviation between these simulations and the observed values are computed per year.

7. References

Bergamaschi, P., Alexe, M., & Segers, A. (2014). *Upgraded CH4 flux inversion system*. Retrieved from http://macc.copernicus-atmosphere.eu/documents/maccii/deliverables/ghg/MACCII_GHG_DEL_D_44.3_20140709_Bergamaschi_final.pdf

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Dentener, F., Wagner, T., … Goede, A. (2007). Satellite chartography of atmospheric methane from SCIAMACHY on board ENVISAT: 2. Evaluation based on inverse model simulations. *Journal of Geophysical Research*, *112*(D2), D02304. https://doi.org/10.1029/2006JD007268

Bergamaschi, P., Frankenberg, C., Meirink, J. F., Krol, M., Villani, M. G., Houweling, S., … Levin, I. (2009). Inverse modeling of global and regional CH 4 emissions using SCIAMACHY satellite retrievals. *Journal of Geophysical Research*, *114*(D22), D22301. https://doi.org/10.1029/2009JD012287

Bergamaschi, P., Houweling, S., Segers, A., Krol, M., Frankenberg, C., Scheepmaker, R. A., … Gerbig, C. (2013). Atmospheric CH 4 in the first decade of the 21st century: Inverse modeling analysis using SCIAMACHY satellite retrievals and NOAA surface measurements. *Journal of Geophysical Research: Atmospheres*, *118*(13), 7350–7369. https://doi.org/10.1002/jgrd.50480

Bergamaschi, P., Krol, M., Dentener, F., Vermeulen, A., Meinhardt, F., Graul, R., … Dlugokencky, E. J. (2005). Inverse modelling of national and European CH4 emissions using the atmospheric zoom model TM5. *Atmospheric Chemistry and Physics*, *5*(9), 2431–2460. https://doi.org/10.5194/acp-5-2431-2005

Bergamaschi, P., Krol, M., Meirink, J. F., Dentener, F., Segers, A., van Aardenne, J., … Dlugokencky, E. J. (2010). Inverse modeling of European CH4 emissions 2001-2006. *Journal of Geophysical Research*, *115*(D22), D22309. https://doi.org/10.1029/2010JD014180

Detmers, R., & Hasekamp, O. (2016). *Product User Guide (PUG) for the RemoTeC XCH4 PROXY GOSAT Data Product v2.3.8*.

Dlugokencky, E. J., Bruhwiler, L., White, J. W. C., Emmons, L. K., Novelli, P. C., Montzka, S. A., … Gatti, L. V. (2009). Observational constraints on recent increases in the atmospheric CH 4 burden. *Geophysical Research Letters*, *36*(18), L18803. https://doi.org/10.1029/2009GL039780

Gilbert, J. C., & Lemaréchal, C. (1989). Some numerical experiments with variable-storage quasi-Newton algorithms. *Mathematical Programming*, *45*(1), 407–435. https://doi.org/10.1007/BF01589113

Granier, C., Bessagnet, B., Bond, T., D'Angiola, A., van der Gon, H., Frost, G. J., … van Vuuren, D. P. (2011). Evolution of anthropogenic and biomass burning emissions of air pollutants at global and regional scales during the 1980--2010 period. *Climatic Change*, *109*(1), 163. https://doi.org/10.1007/s10584-011-0154-1

Hersbach, H., P. de Rosnay, B. Bell, D. Schepers, A. Simmons, C. Soci, S. Abdalla, M. Alonso Balmaseda, G. Balsamo, P. Bechtold, P. Berrisford, J. Bidlot, E. de Boisséson, M. Bonavita, P. Browne, R. Buizza, P. Dahlgren, D. Dee, R. Dragani, M. Diamantakis, J. Flemming, R. Forbes, A. Geer, T. Haiden, E. Hólm, L. Haimberger, R. Hogan, A. Horányi, M. Janisková, P. Laloyaux, P. Lopez, J. Muñoz-Sabater, C. Peubey, R. Radu, D. Richardson, J.-N. Thépaut, F. Vitart, X. Yang, E. Zsótér & H. Zuo, 2018: Operational global reanalysis: progress, future directions and synergies with NWP, ECMWF ERA Report Series 27.

Houweling, S., Dentener, F., & Lelieveld, J. (1998). The impact of nonmethane hydrocarbon

compounds on tropospheric photochemistry. *Journal of Geophysical Research: Atmospheres*, *103*(D9), 10673–10696. https://doi.org/10.1029/97JD03582

Huijnen, V., Williams, J., van Weele, M., van Noije, T., Krol, M., Dentener, F., … Pätz, H.-W. (2010). The global chemistry transport model TM5: description and evaluation of the tropospheric chemistry version 3.0. *Geoscientific Model Development*, *3*(2), 445–473. https://doi.org/10.5194/gmd-3-445-2010

Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U., and Petrescu, A. M. R.: EDGAR v4.3.2 Global Atlas of the three major Greenhouse Gas Emissions for the period 1970–2012, Earth Syst. Sci. Data Discuss., https://doi.org/10.5194/essd-2017-79, 2017.

Jöckel, P., Tost, H., Pozzer, A., Brühl, C., Buchholz, J., Ganzeveld, L., … Lelieveld, J. (2006). The atmospheric chemistry general circulation model ECHAM5/MESSy1: consistent simulation of ozone from the surface to the mesosphere. *Atmospheric Chemistry and Physics*, *6*(12), 5067–5104. https://doi.org/10.5194/acp-6-5067-2006

Kaiser, J. W., Heil, A., Andreae, M. O., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M. G., Suttie, M., and van der Werf, G. R.: Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power, Biogeosciences, 9, 527–554, doi:10.5194/bg-9-527-2012, 2012.

Krol, M., Houweling, S., Bregman, B., van den Broek, M., Segers, A., van Velthoven, P., … Bergamaschi, P. (2005). The two-way nested global chemistry-transport zoom model TM5: algorithm and applications. *Atmospheric Chemistry and Physics*, *5*, 417–432.

Meirink, J. F., Bergamaschi, P., & Krol, M. C. (2008). Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion. *Atmospheric Chemistry and Physics*, *8*(21), 6341–6353. https://doi.org/10.5194/acp-8-6341-2008

Meirink, J. F., Eskes, H. J., & Goede, A. P. H. (2006). Sensitivity analysis of methane emissions derived from SCIAMACHY observations through inverse modelling. *Atmospheric Chemistry and Physics*, *6*(5), 1275–1292. https://doi.org/10.5194/acp-6-1275-2006

Monteil, G., Houweling, S., Butz, A., Guerlet, S., Schepers, D., Hasekamp, O., … Röckmann, T. (2013). Comparison of CH 4 inversions based on 15 months of GOSAT and SCIAMACHY observations. *Journal of Geophysical Research: Atmospheres*, *118*(20), 11,807-11,823. https://doi.org/10.1002/2013JD019760

Pandey, S., Houweling, S., Krol, M., Aben, I., Chevallier, F., Dlugokencky, E. J., … Röckmann, T. (2016). Inverse modeling of GOSAT-retrieved ratios of total column CH4 and CO2 for 2009 and 2010. *Atmospheric Chemistry and Physics*, *16*(8), 5043–5062. https://doi.org/10.5194/acp-16-5043-2016

Rayner, P., Michalak, A. M., & Chevallier, F. (2016). Fundamentals of Data Assimilation. *Geoscientific Model Development Discussions*, 1–21. https://doi.org/10.5194/gmd-2016-148

Razavi, A., Clerbaux, C., Wespes, C., Clarisse, L., Hurtmans, D., Payan, S., … Coheur, P. F. (2009). Characterization of methane retrievals from the IASI space-borne sounder. *Atmospheric Chemistry and Physics*, *9*(20), 7889–7899. https://doi.org/10.5194/acp-9-7889-2009

Schepers, D., Guerlet, S., Butz, A., Landgraf, J., Frankenberg, C., Hasekamp, O., … Aben, I. (2012). Methane retrievals from Greenhouse Gases Observing Satellite (GOSAT) shortwave infrared measurements: Performance comparison of proxy and physics retrieval algorithms. *Journal of*

Geophysical Research: Atmospheres, *117*(D10), n/a-n/a. https://doi.org/10.1029/2012JD017549

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., … Zondlo, M. A. (2010). Calibration of the Total Carbon Column Observing Network using aircraft profile data. *Atmospheric Measurement Techniques*, *3*(5), 1351–1362. https://doi.org/10.5194/amt-3-1351-2010

Zhang, Zhen, Niklaus E Zimmermann, Leonardo Calle, George Hurtt, Abhishek Chatterjee, and Benjamin Poulter (2018). Enhanced Response of Global Wetland Methane Emissions to the 2015-2016 El Nino-Southern Oscillation Event. Environmental Research Letters 13 (5), 074009. doi: 10.1088/1748-9326/aac939

Copernicus Atmosphere Monitoring Service

ECMWF - Shinfield Park, Reading RG2 9AX, UK

Contact: info@copernicus-atmosphere.eu

atmosphere.copernicus.eu copernicus.eu ecmwf.int