Variational regional inverse modeling of reactive species emissions with PYVAR-CHIMERE-v2019

Audrey Fortems-Cheiney[1], Isabelle Pison[1], Grégoire Broquet[1], Gaëlle Dufour[2], Antoine Berchet[1], Elise Potier[1], Adriana Coman[2], Guillaume Siour[2], and Lorenzo Costantino[2]

[1]Laboratoire des Sciences du Climat et de l'Environnement, LSCE-IPSL (CEA-CNRS-UVSQ), Université Paris-Saclay, 91191 Gif-sur-Yvette, France.

[2]Laboratoire Interuniversitaire des Systèmes Atmosphériques, UMR CNRS 7583, Université Paris Est Créteil et Université Paris Diderot, Institut Pierre Simon Laplace, Créteil, France.

Abstract

Up-to-date and accurate emission inventories for air pollutants are essential for understanding their role in the formation of tropospheric ozone and particulate matter at various temporal scales, for anticipating pollution peaks and for identifying the key drivers that could help mitigate their emissions. This paper describes the Bayesian variational inverse system PYVAR-CHIMERE, which is now adapted to the inversion of reactive species, in addition to greenhouse gases. Complementarily with bottom-up inventories, this system aims at updating and improving the knowledge on the high spatio-temporal variability of emissions of air pollutants and their precursors. The system is designed to use any type of observations, such as satellite observations or surface station measurements. The potential of PYVAR-CHIMERE is illustrated with inversions of both CO and NO_x emissions in Europe, using the MOPITT and OMI satellite observations, respectively.

1. Introduction

The degradation of air quality is a worldwide environmental problem: 91% of the world's population have breathed polluted air in 2016 according to the World Health Organization (WHO), resulting in 4.2 millions of premature deaths every year [WHO, 2016]. The recent study of Lelieveld et al. [2019] even suggests that the health impacts attributable to outdoor air pollution are substantially higher than previously assumed (with 790,000 premature deaths in the 28 countries of the European Union against the previously estimated 500,000 [EEA, 2018]). The main regulated primary (i.e. directly emitted in the atmosphere) anthropogenic air pollutants are carbon monoxide (CO), nitrogen oxides (NO_x =NO+NO_2), sulfur dioxide (SO_2), ammonia (NH_3), volatile organic compounds (VOCs), and primary particles. These primary air pollutants are precursors of secondary (i.e. produced in the atmosphere through chemical reactions) pollutants such as ozone (O_3) and Particulate Matter (PM), which are also threatening to both human health and ecosystems. Monitoring concentrations and quantifying emissions are still challenging and limit our capability to forecast air quality to warn population and to assess i) the exposure of population to air pollution and ii) the efficiency of mitigation policies.

Bottom-up (BU) inventories are built in the framework of air quality policies such as The Convention on Long-Range Transboundary Air Pollution (LRTAP, http://www.unece.org) for air pollutants. Based on national annual inventories, research institutes compile gridded global or regional, monthly inventories (mainly for the US, Europe and China) with a high spatial resolution (currently regional or city scale inventories are typically finer than 0.1°x0.1°). These inventories are constructed by combining available (economic) statistics data from different detailed activity sectors with the most appropriate emission factors (defined as the average emission rate of a given species for a given source or process, relative to the unit of activity). It is important to note that the activity data (often statistical data) has an inherent uncertainty and that its reliability may vary between countries or regions. In addition, the emission factors bear large uncertainties in their quantification [Kuenen et al., 2014; EMEP/EEA, 2016; Kurokawa et al., 2013]. Moreover, these inventories are often provided at the annual or monthly scale with typical temporal profiles to build the weekly, daily and hourly variability of the emissions. The combination of uncertain activity data, emission factors and emission timing can be a large source of uncertainties, if not errors, for forecasting or analyzing air quality [Menut et al., 2012]. Finally, since updating the inventories and gathering the required data for a given year is costly in time, manpower and money, only a few institutes have offered estimates of the gaseous pollutants for each year since 2011 (i.e, EMEP updated until the year 2017, MEIC updated until the year 2017 to our knowledge). Nevertheless, using knowledge from inventories and air quality modeling, emissions have been mitigated. For example, from 2010 to nowadays, emissions in various countries have been modified and/or regional trends have been reversed (e.g., the decrease of NO_x emissions over China since 2011 [de Foy et al., 2016]), leading to significant changes in the atmospheric composition. Consequently, the knowledge of precise and updated budgets, together with seasonal, monthly, weekly and daily variations of gaseous pollutants driven, amongst other processes, by the emissions are essential for understanding their role in the formation of tropospheric ozone and PMs at various temporal scales, for anticipating pollution peaks and for identifying the key drivers that could help mitigate these emissions.

In this context, complementary methods have been developed for estimating emissions using atmospheric observations. They operate in synergy between a chemistry-transport model (CTM) - which links the emissions to the atmospheric concentrations-, atmospheric observations of the species of interest, and statistical inversion techniques. A number of studies using inverse modeling were first carried out for long-lived species such as greenhouses gases (GHGs) (e.g., carbon dioxide CO_2 or methane CH_4) at the global or continental scales [Hein et al., 1997; Bousquet et al. 1999], using surface measurements. Later, following the development of monitoring station networks, the progress of computing power, and the use of inversion techniques more appropriate to non-linear problems, these methods were applied to shorter-lived molecules such as CO. For these various applications (e.g., for CO_2, CH_4, CO), the quantification of sources was solved at the resolution of large regions [Pétron et al., 2002]. Finally, the growing availability and reliability of observations since the early 2000s (in-situ surface data, remote sensing data such as satellite data), the improvement of the global CTMs, of the computational capacities and of the inversion techniques have increased the achievable resolution of global inversions, up to the global transport model grid cells, i.e. typically with a spatial resolution of several hundreds of square kilometers [Stavrakou and Muller, 2006; Pison et al., 2009; Fortems-Cheiney et al., 2011; Hooghiemstra et al., 2012; Yin et al., 2015; Miyazaki et al., 2017, Zheng et al., 2019].

Today, the scientific and societal issues require an up-to-date quantification of pollutant emissions at a higher spatial resolution than the global one and imply to widely use regional inverse systems. However, although they are suited to reactive species such as CO and NO_x, and their very large spatial and temporal variability, they have hardly been used to quantify pollutant emissions. Some studies inferred NO_x [Pison et al., 2007; Tang et al., 2013] and VOC emissions [Koohkan et al., 2013] from surface measurements. Konovalov et al. [2006, 2008, 2010], Mijling et al. [2012, 2013], van der A et al. [2008], Lin et al. [2012] and Ding et al. [2017] have also shown that satellite observations are a suitable source of information to constrain the emissions of NO_x. These regional inversions using satellite observations were often based on Kalman Filter (KF) schemes [Mijling et al., 2012, 2013; Van der A et al., 2008; Lin et al., 2012; Ding et al., 2017].

Here, we present the Bayesian variational atmospheric inversion system PYVAR-CHIMERE for the monitoring of anthropogenic emissions at high spatial resolutions. It takes advantage of the previous developments for the quantification of fluxes of long-lived GHG species such as CO_2 [Broquet et al., 2011] and CH_4 [Pison et al., 2018] at the regional to the local scales, but now solves for reactive species such as CO and NO_x. It has also a better level of robustness, clarity, portability, and modularity than these previous systems .It is based on the Bayesian variational assimilation code PYVAR [Chevallier et al. 2005] and on the regional state-of-the-art CTM CHIMERE, dedicated to the study of regional atmospheric pollution events [Menut et al., 2013, Mailler et al., 2017]. Variational techniques require the adjoint of the model to compute the sensitivity of simulated atmospheric concentrations to corrections of the fluxes. CHIMERE is one of the CTMs possessing its adjoint code (e.g., for global models: GEOS-CHEM [Henze et al., 2007], IMAGES [Stavrakou and Muller, 2006], TM5 [Krol et al., 2008], GELKA [Belikov et al., 2016] and LMDz [Chevallier et al., 2005; Pison et al., 2009] ; for limited-area models: CMAQ [Hakami et al., 2007], EURAD-IM [Elbern et al., 2007], RAMS/CTM-4DVAR [Yumimoto et Uno, 2006], WRF-CO2 4D-Var [Zheng et al., 2018]).

The principle of variational atmospheric inversion and the configuration of PYVAR-CHIMERE are described in Section 2 and in Section 3, respectively. Details about the forward**,** tangent-linear and adjoint codes of CHIMERE are also given. Then, the potential of PYVAR-CHIMERE is illustrated in Section 4 with the optimization of European CO and NO_x emissions, constrained by observations from the Measurement of Pollution in the Troposphere (MOPITT) and from the Ozone Monitoring Instrument (OMI) satellite instruments, respectively.

2. Principle of Bayesian variational atmospheric inversion

The Bayesian variational atmospheric inversion method adjusts a set of control parameters in input of the CTM, including parameters related to the emissions whose estimate is the primary target of the inversion. The control vector **x** contains these variables to be optimized during the inversion process (surface fluxes but it may also include initial or boundary conditions for example, see Section 3.3). The adjustments are applied to prior values, usually taken, for the emissions, from pre-existing BU inventories. The principle is to minimize, on the one hand, the departures from the prior estimates of the control parameters, which are weighted by the uncertainties in these estimates (called hereafter "prior uncertainties"), and, on the other hand, the differences between simulated and observed concentrations, which are weighted by all other sources of uncertainties explaining these differences (called hereafter all together "observation errors"). In statistical terms, the inversion searches for the most probable estimate of the control parameters given their prior estimates, the observations, the CTM and the associated uncertainties. The solution, which will be called posterior estimate in the following, is found by the iterative minimization of a cost function J [Talagrand et al., 1997], defined as:

$$J(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x_b})^T \boldsymbol{B}^{-1} (\boldsymbol{x} - \boldsymbol{x_b}) + (H(\boldsymbol{x}) - \boldsymbol{y})^T \boldsymbol{R}^{-1} (H(\boldsymbol{x}) - \boldsymbol{y}) \qquad \text{(Eq. 1)}$$

H is the non-linear observation operator that projects the state vector **x** onto the observation space. In most of the variational atmospheric inversion cases (such as those described in Section 4), the observation operator includes the CTM and an interpolation or an extraction and averaging of the simulated concentration fields (see Section 3.4). The observations in **y** could be surface measurements and/or remote sensing data such as satellite data. The prior uncertainties and the observation errors are assumed to be centered and to have a Gaussian distribution. Consequently, the prior uncertainties are characterized by their covariance matrix **B** and the observation errors are characterized by their covariance matrix **R.** By definition, the observation errors combine errors in both the data and the observation operator, in particular measurement errors and errors in the conversion of satellite measurement into concentration data, errors from the CTM, representativity errors due to the comparison between point measurements and gridded models or due to the representation of the fluxes as gridded maps at a given spatial resolution, and aggregation errors associated with the optimization of emissions at a given spatial and/or temporal resolution (as specified in the control vector) that is different from (usually coarser than) that of the CTM [Wang et al., 2017].

For inversions with observation and control vectors having a high dimension, the minimum of J cannot be found analytically due to computational limitations. It can be reached iteratively with a descent algorithm. In this case, the iterative minimization of J is based on a gradient method. J is calculated with the forward observation operator (including the CTM) and its gradient relative to the control parameters **x**: $\nabla J(\boldsymbol{x}) = B^{-1}(\boldsymbol{x} - \boldsymbol{x_b}) + H^T\boldsymbol{R}^{-1}(H(\boldsymbol{x}) - \boldsymbol{y})$ (Eq. 2) is provided by the adjoint of the observation operator (including the adjoint of the CTM). As shown in Figure 1, the minimization algorithm repeats the forward-adjoint cycle to seek an optimal solution for the control parameters.

The high-non linearity of the chemistry for reactive species makes it difficult to use its tangent-linear to approximate the actual observation operator (e.g. as in Chevallier et al. [2010]who use the conjugate gradient algorithm of Fisher and Courtier [1995]), and, more generally, it makes the inversion problem highly non linear. Therefore, in PYVAR-CHIMERE, we use the M1QN3 limited memory quasi-Newton minimization algorithm [Gilbert and Lemaréchal, 1989], which relies on the actual CHIMERE non-linear model to compute J at each iteration of the minimization. As most quasi-Newton methods, it requires an initial regularization of **x**, the vector to be optimized, for better efficiency. We adopt the most generally used regularization, made by minimizing in the space defined by $\chi = \boldsymbol{B}^{\frac{1}{2}}(\boldsymbol{x} - \boldsymbol{x_b})$ instead of the control space defined by **x**. Although more advanced regularizations can be chosen, the minimization with χ is preferred for its simplifying the equation to solve. In the χ-space, Equation 2 can be re-written as follows: $\nabla J\chi = \chi + \boldsymbol{B}^{\frac{1}{2}}H \times \left(\boldsymbol{R}^{-1}(H(\boldsymbol{x}) - \boldsymbol{y})\right)$. The criterion for stopping the algorithm is based on a threshold set on the ratio between the final and initial gradient norms or on the maximum number of iterations to perform. Due to the non-linearity of the problem, the minimization may reach only a local minimum.

Finally, the calculation of the uncertainty in the estimate of emissions from the inversion, known as "posterior uncertainty", is challenging in a variational inverse system. Even though the posterior uncertainty can be explicitly written in various analytical forms, it requires the inversion of matrices that are too large to invert given the current computational resources in our variational approach. As a trade-off between computing resources and comprehensiveness, the analysis error may be evaluated by an approach based on a propagation of errors through sensitivity tests (e.g., as in Fortems-Cheiney et al., [2012]). It can also be estimated through a Monte Carlo Ensemble [Chevallier et al., 2007], implemented in PYVAR.

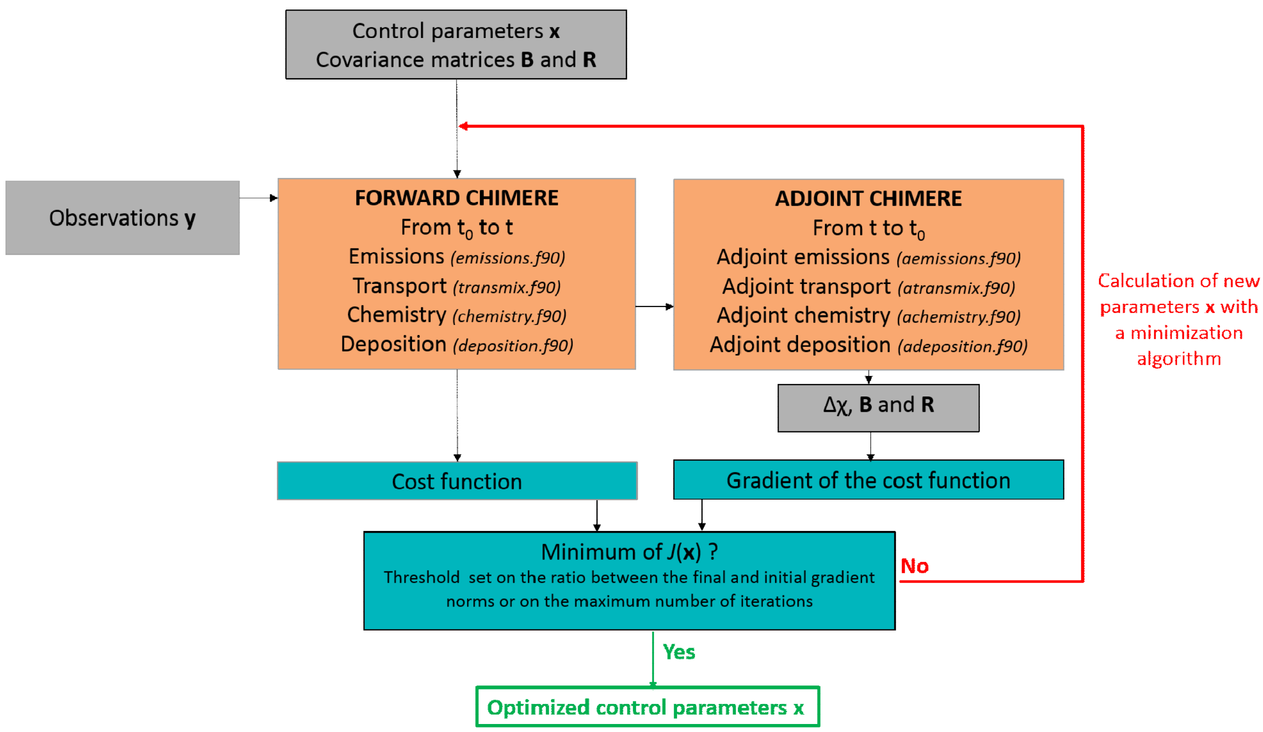

Figure 1. *Simplified scheme of the iterative minimization in PYVAR-CHIMERE. PYVAR, CHIMERE and text sources are* ***displayed in blue, in orange and in grey, respectively***.

3. The PYVAR-CHIMERE configuration

3.1. PYVAR adapted to CHIMERE

The PYVAR-CHIMERE inverse modeling system is based on the Bayesian variational assimilation code PYVAR [Chevallier et al. 2005] and on a previous inversion system coupled to CHIMERE [Pison et al., 2007]. PYVAR is an ensemble of Python scripts, which deals with preparing the vectors and the matrices for the inversion, drives the required Fortran codes of the transport model and computes the minimization of the cost function to solve the inversion. Previously used for global inversions with the LMDz model (e.g., Pison et al., 2009; Chevallier et al., 2010; Fortems-Cheiney et al., 2011; Yin et al., 2015; Locatelli et al., 2015; Zheng et al., 2019), PYVAR has been adapted to CHIMERE with an adjoint code without chemistry a first time by Broquet et al. [2011].

In order to couple PYVAR to the new state-of-the-art version of CHIMERE (see Section 3.2), to include chemistry, and to increase its modularity, flexibility and clarity, the new system described here has been developed. It includes elements of the inversion system (coded in Fortran90) of [Pison et al., 2007].

3.2. Development and parallelization of the adjoint and tangent-linear codes of CHIMERE

To compute the sensitivity of simulated atmospheric concentrations to corrections to the fluxes, the adjoint of CHIMERE has been developed. Originally, the sequential adjoint was coded [Menut et al., 2000; Menut et al., 2003; Pison et al., 2007].The adjoint has been coded by hand line by line, following the principles formulated by Talagrand [1997]. It contains exactly the same processes as the CHIMERE forward model. Then, it has been parallelized. This work required a redesigning of the whole code, associatedwith a full testing scheme.Furthermore, the tangent-linear (TL) code has been developed and validated at LSCE.Changes have been implemented in the forward CHIMERE code embedded in PYVAR-CHIMERE to match requirements of the studies lead with PYVAR-CHIMERE. These changes have been implemented in both the adjoint and the TL codes. Compared to the CHIMERE 2013 version [Menut et al., 2013], the most important of these changes are:

•For the geometry, the possibility of polar domains and the use of the coordinates of the corners of the cells instead of only the centers

•For the transport, the non-uniform Van Leer transport scheme on the horizontal,

•For chemistry, various switches have been added to avoid going into the chemistry, deposition and wet deposition routines when no species requires them (e.g. no chemistry for methane at a regional scale).

PYVAR-CHIMERE is currently operational for the full module of gaseous chemistry. As a compromise between the robustness of the method for reactive species, the time required coding the adjoint and the computational cost with a full chemical scheme, the aerosols modules of CHIMERE have not been included in the adjoint of CHIMERE yet and are therefore not available in PYVAR-CHIMERE. The development and maintenance of the adjoint means that the version used is necessarily one or two versions behind the distributed CHIMERE version (http://www.lmd.polytechnique.fr/chimere/). It should also be noted that PYVAR-CHIMERE only infer anthropogenic emissions at this stage. The optimization of biogenic emissions, which are linearly interpolated at the sub-hourly scale in CHIMERE, is currently under development.

As an example, Figure 2 presents a simplified scheme of how PYVAR scripts are used to drive this version of CHIMERE for forward simulations and inversions using satellite observations. A mode is also available to test the adjoint: it runs the TL code.

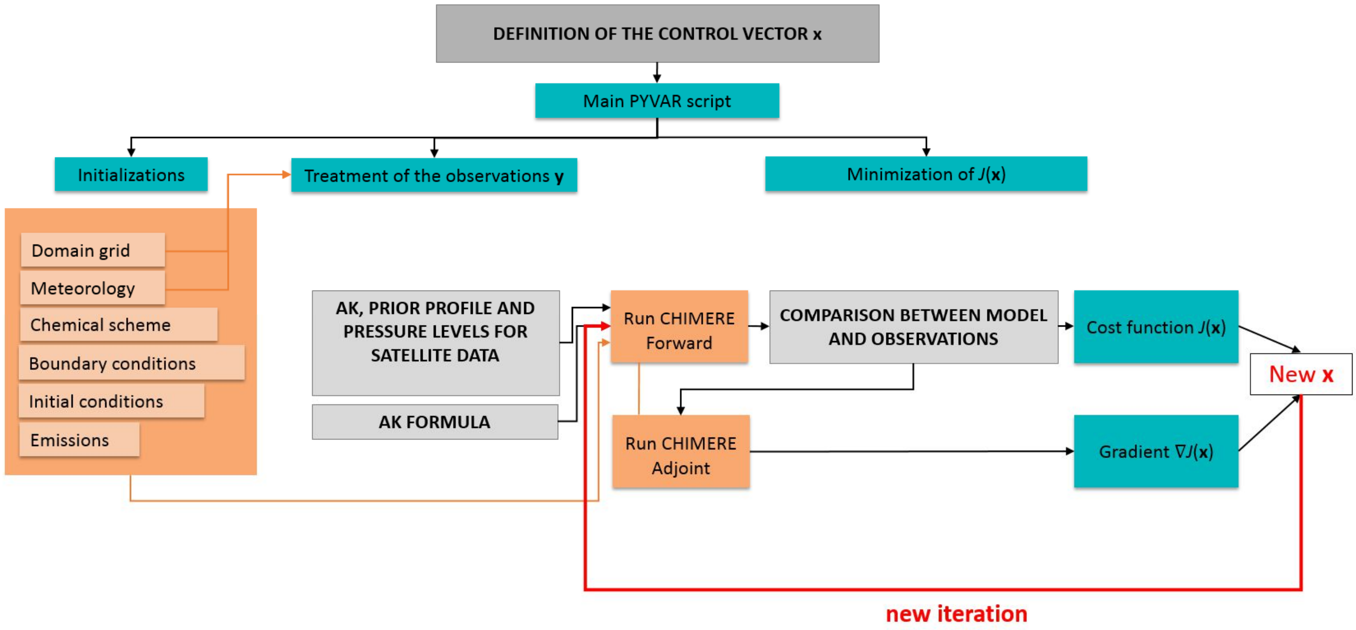

Figure 2. *Simplified scheme of how PYVAR scripts are used to drive CHIMERE for an inversion using satellite observations. PYVAR, CHIMERE and text sources are displayed in blue, in orange and in grey, respectively. "AK" refers to Averaging Kernels as detailed in Section 3.4.*

3.3. Definition of the control vector

The control vector is specified by the user in a text file. This file is formatted following Table 1.The parameters to constrain could be fluxes and/or initial conditions and/or boundary concentration conditions, at the grid-cell resolution or for one region encompassing up to the whole domain. Several types of corrections can be applied, they are defined in the code as "add", "mult" or "scale". Both the corrections "add" and "mult" are applied to gridded control variables. For correction type "add" the control variables are increments added to the corresponding components of the model inputs. For correction type "mult", the control variables are scaling factors multiplying the corresponding components of the model inputs. The difference between the two options "add" and "mult" plays a role when inverting fluxes which can switch from positive to negative values (like CO_2 natural fluxes). For type "scale", the corrections consist in applying scaling factors to activity maps and/or masks for regions (which is similar to the control of budgets for different regions, types of activities, and/or processes in inversions where the control vector is not gridded [Wang et al., 2018]) and adding the obtained values to the corresponding components of the model inputs.

Different simple but efficient ways of building the error covariance matrix B are implemented in PYVAR-CHIMERE. The variances and correlations are defined independently. The variances are specified by the user through standard deviation coefficient (Table 1), which can be a fixed value ("fx") or a percentage ("pc") to define the diagonal standard deviation matrix $\sum$. For correction types "mult" and "scale", as well as for correction type "add" with a fixed value, the value is directly used as the standard deviation of the uncertainty in the corresponding components of the control vector. For correction type "add" with a percentage provided, maps of standard deviation of uncertainty are built by applying this percentage to the matching input fields (fluxes, initial conditions, boundary conditions). The user may also provide a script to build personalized maps of variances.

Potential correlations between uncertainties in different types of control variables, e.g. between fluxes and boundary conditions, and correlations between uncertainties in different species, e.g. between fluxes of CO and NO_x, are not coded yet. Only correlations for a given type of control variable and a given species are so far taken into account so that the **B** matrix is block diagonal. For a given type of control variable and a given species (in the illustration in section 4.2.2: CO, NO or NO_2 fluxes), spatial and temporal correlations can be defined using correlation lengths through time Lt and space Ls. Those lengths are used to model temporal and/or spatial auto-correlations using an exponentially decaying function: the correlation r between parameters and at a given location but separated by duration $d(x_i, x_j)$, or at a given time but distant by $d\left(x_i, x_j\right)$ is given by $r\left(x_i, x_j\right) = exp\left(\frac{-d(x_i,x_j)}{L}\right)$where $L\ (= L_T\ or\ L_S)$ is the corresponding correlation length. There is no correlation between uncertainties in land and ocean flux. Note that the spatial correlations are computed for each vertical level independently when dealing with control variables with vertical resolution (3D fields of fluxes when accounting for emission injection heights, or boundary/initial conditions). Vertical correlations in the uncertainties in such variables have not been coded yet. Apart from this, the system assumes that temporal correlations and spatial correlations depend on the time lag and distance but not on the specific time and location of the corresponding parameters. It also assumes that the correlation between uncertainties at different locations and different time can be derived from the product of the corresponding autocorrelation in time and space.

Each block of **B** can thus be decomposed based on Kronecker products: $\mathbf{B}=\sum C_t \otimes C_s \sum$ where $\otimes$ is the Kronecker product, C_t and C_s are the temporal and spatial correlations, respectively. The calculations involving $\mathbf{B}^{1/2}$ are simplified in PYVAR-CHIMERE using the Eigen-decomposition of Ct and Cs. Its square root can be calculated according to: $C_t^{1/2}= V_{Ct} D_{Ct}^{1/2} V_{Ct}^T$ (and similarly for C_s) (Eq 4) where V_{Ct} is the matrix with the Eigenvectors as columns, and D_{Ct} is the diagonal matrix of

Eigenvalues of C_t. It is possible to chose a threshold under which the eigenvalues are truncated when computing the spatial correlations in order to save computation and memory, but not when computing the temporal correlations.

Possible ways to define the control vector and to construct the error covariance B matrix								
Constrained species	Correction type : - Add - Mult - Scale	Spatial resolution - at the grid-cell resolution - for one region	Temporal resolution (in hours)	Input to constrain: -Fluxes -Initial conditions -Lateral Boundary conditions -Top Boundary conditions	B variance coefficient: -fx -pc	Decorrelation time (in hours)	Decorrelation length on land (in km)	Decorrelation length on sea (in km)
Examples of the definition of the control vector and of the construction of the B matrix for the experiments presented in Section 4								
CO	add	0.5°x0.5°	24	Fluxes	100 %	-	-	-
CO	add	0.5°x0.5°	24	Initial conditions	15%	-	-	-
CO	add	0.5°x0.5°	24	Lateral Boundary conditions	15%	-	-	-
CO	add	0.5°x0.5°	24	Top Boundary conditions	15%	-	-	-
NO	add	0.5°x0.5°	24	Fluxes	30 %	-	50	50
NO	add	0.5°x0.5°	24	Initial conditions	15%	-	-	-
NO_2	add	0.5°x0.5°	24	Fluxes	30 %	-	50	50
NO_2	add	0.5°x0.5°	24	Initial conditions	15%	-	-	-

Table 1. *Possible ways to define the control vector and to construct the error covariance B matrix in PYVAR-CHIMERE and examples of the configuration for the experiments presented in Section 4.*

3.4 Equivalents of the observations

The individual data given as constraints in the system are first formatted into a text file described in Figure 4. During forward simulations, the equivalents of the components of **y** (i.e, the equivalents of the individual data) are calculated by PYVAR-CHIMERE. It includes the CTM and an interpolation (see below the vertical interpolation from the model's grid to the satellite levels) or an extraction and averaging (e.g. extracting the grid cell matching the geographical coordinates of a surface station and averaging over one hour). As a compromise between technical issues such as the time required for reading/writing files, the observation operator *H* that generates the equivalent of the observations by the model (i.e. $H(\mathbf{x})$) has been so far partly embedded in the code of CHIMERE. It makes it easier to use finer time intervals than available in the usual hourly outputs of CHIMERE to compute the required information (e.g., within the finer CTM physical time steps).

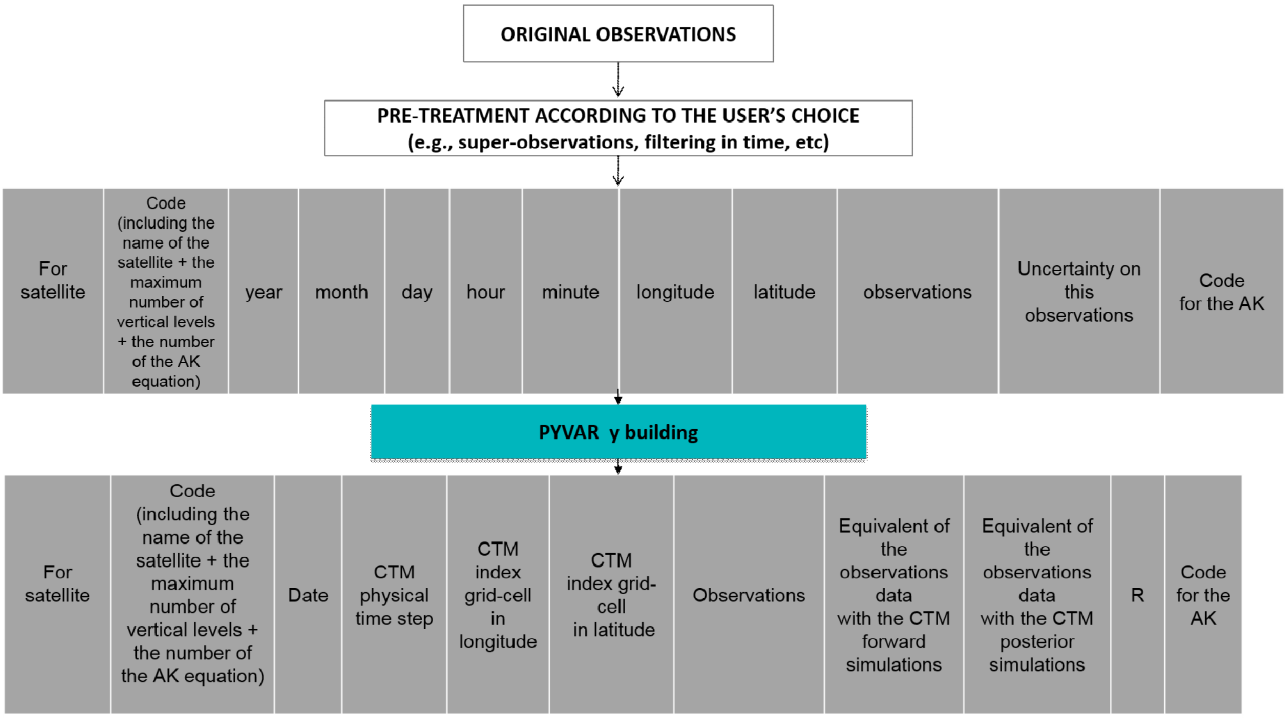

Figure 3. *Simplified scheme of how PYVAR scripts prepares the observations, using satellite data*

To make comparisons between simulations and satellite observations, the simulated vertical profiles are first interpolated on the satellite's levels (with a vertical interpolation on pressure levels) **in CHIMERE**. Then, the averaging kernels (AKs), when available, are applied to represent the vertical sensitivity of the satellite retrieval. Two types of formula, depending on the satellite observations used, have been detailed in PYVAR-CHIMERE for the use of AKs: $C_m = AK.C_{m(o)}$ or $C_m = x_a + AK\left(c_{m(o)} - x_a\right)$ where C_m is the modeled column, AK contains the averaging kernels, x_a is the prior profile (provided together with the AKs when relevant) and $C_{m(o)}$ is the vertical distribution of the original model partial columns interpolated to the pressure grid of the averaging kernels.

3.5. Numerical language

The PYVAR code is in Python 2.7, the CHIMERE CTM is coded in Fortran90. The CTM requires several numerical tools, compilers and libraries. The PYVAR-CHIMERE system was developed and tested using the software versions as described in Table 2.

	URL	Version

Software	Python	https://www.python.org/downloads/	2.7
	Fortran compiler ifort	https://software.intel.com/en-us/fortran-compilers	Composer-xe-2013.2.146
Libraries or packages	UnidataNetCDF	https://www.unidata.ucar.edu/	3
	Open MPI	https://www.open-mpi.org/	1.10.5
	GRIB_API	https://confluence.ecmwf.int/display/GRIB/Releases	1.14
	nco	http://nco.sourceforge.net/#Source	4.6.3

Table 2. *URL addresses for the development and the use of the PYVAR-CHIMERE system and its modules.*

PYVAR-CHIMERE's computation time for one node of 10 CPUs is about 4h for 1 day of inversion (with ~10 iterations) for the European domain size of 101 (longitude) x 85 (latitude) x 17 (vertical levels) used in Section 4.2.3. The model parallelism results from a Cartesian division of the main geographical domain into several sub-domains, each one being processed by a worker process. To configure the parallel sub-domains, the user has to specify two parameters in the model parameter file: the number of sub-domains for the zonal and meridian directions. The total number of CPUs used is therefore the product of these two numbers plus one for the master process.

4. Potential of PYVAR-CHIMERE for the inversion of CO and NO_x emissions

The potential of the PYVAR-CHIMERE system to invert emissions of reactive species is illustrated with the inversion of CO and NO_x anthropogenic emissions in Europe respectively based on MOPITT CO data andOMI NO_2 data. We have chosen to present illustration of CO inversion over a 7-day window, the first week of March 2015. Considering the short lifetime of NO_x of a few hours [Valin et al., 2013; Liu et al., 2016], we have chosen to present illustration of NO_x inversion over a 1-day window, the 19th February 2015. These particular periods have been chosen as they present **a representative** number of super-observations during winter, and as the emissions are high during that period. All the information required by the system to invert CO and NO_x emissions are listed in Table 1.

4.1. Data and model description

4.1.1. Observations y

We use CO data from the MOPITT instrument [Deeter et al., 2019]. MOPITT has been flown onboard the NASA EOS-Terra satellite, on a low sun-synchronous orbit that crosses the equator at 10:30 and 22:30 LST. The spatial resolution of its observations is about 22x22 km^2 at nadir. It has been operated nearly continuously since March 2000. MOPITT CO products are available in three variants: thermal-infrared TIR only, near-infrared NIR only and the multispectral TIR-NIR product, all containing total columns and retrieved profiles (expressed on a ten-level grid from the surface to 100 hPa). We choose to constrain CO emissions with the MOPITT surface product for our illustration. Among the different MOPITTv8 products, we choose to work with the multispectral MOPITTv8-NIR-TIR one, as it provides the highest number of observations, with a good evaluation against in situ data from NOAA stations [Deeter et al., 2019]. The MOPITTv8-NIR-TIR surface concentrations are sub-sampled into "super-observations" in order to reduce the effect of errors that are correlated between neighboring observations*:* we selected the median of each subset of MOPITT data within each 0.5°×0.5° grid-cell and each physical time step (about 5-10 minutes). After this screening, 8437"super-observations" remain in the 7-day inversion (from 10667 raw observations). These super-observations are provided to PYVAR-CHIMERE as constraints **y**, and treated as described in Section 3.4. It is important to note that the potential of MOPITT to provide information at a high temporal resolution, up to the daily scale, is hampered by the cloud coverage (see the blanks in Figure 6b).

The observational constraint on NO_2 emissions comes from the OMI QA4ECV tropospheric columns [Muller et al., 2016; Boersma et al., 2016, Boersma et al., 2017]. The Ozone Monitoring Instrument (OMI), a near-UV/Visible nadir solar backscatter spectrometer, was launched onboard EOS Aura in July 2004. It has been flown on a 705 km sun-synchronous orbit that crosses the Equator at 13:30 LT. Our data selection follows the criteria of the OMI QA4ECV data quality statement. As the spatial resolution of the OMI data is finer than that of the chosen CHIMERE model grid (13x24 km^2 against 0.5°×0.5°, respectively), the OMI tropospheric columns are sub-sampled into "super-observations" (median of the OMI data within the 0.5°×0.5° grid-cell and each physical time step and its corresponding AK).

4.1.2 CHIMERE set-up

CHIMERE is run over a 0.5°×0.5° regular grid (about 50x50km^2) and 17 vertical layers, from the surface to 200hPa (about 12km), with 8 layers within the first two kilometers. The domain includes 101 (longitude) x 85 (latitude) grid-cells (15.5°W-35°E; 31.5°N-74°N, see Figure 5). CHIMERE is driven by the European Centre for Medium-Range Weather Forecasts (ECMWF) meteorological forecast [Owens and Hewson, 2018]. The chemical scheme used in PYVAR-CHIMERE is MELCHIOR-2, with more than 100 reactions [Lattuati, 1997; CHIMERE 2017], including 24 for inorganic chemistry. The prior anthropogenic emissions for CO and NO_x emissions come from the TNO-GCHco-v1 inventory [Super et al., 2019], the last update of the TNO-MACCII inventory [Kuenen et al., 2014]. The prior anthropogenic emissions for VOCs come from the EMEP inventory [Vestreng et al., 2005; EMEP/CEIP website]. Different climatological values from the LMDZ-

INCA global model [Szopa et al., 2008] or from a MACC reanalysis are used to prescribe concentrations at the lateral and top boundaries and the initial atmospheric composition in the domain. Full access to and more information about the MACC reanalysis data can be obtained through the MACC-II web site(http://www.copernicus-atmosphere.eu). In order to ensure realistic fields of simulated CO and NO_2 concentrations from the beginning of the inversion period, runs have been preceded with a 10-day spin-up.

4.1.3. Sensitivity to emissions and to initial and boundary conditions

With its lifetime of about two months, CO could be strongly driven by the initial and lateral boundary conditions prescribed in the CTM. In fact, as seen in Figure 4b, initial and boundary conditions provide a relatively flat background and the patterns which appear clearly over the background are linked to surface emissions (Figure 4a). To characterize the uncertainties in the concentration fields due to the initial and lateral boundary conditions, we performed a sensitivity test by using either climatological values from LMDZ-INCA or a MACC reanalysis: the results were not significantly different, with relative differences in concentrations of less than 15% over continental land (Figure 5c).

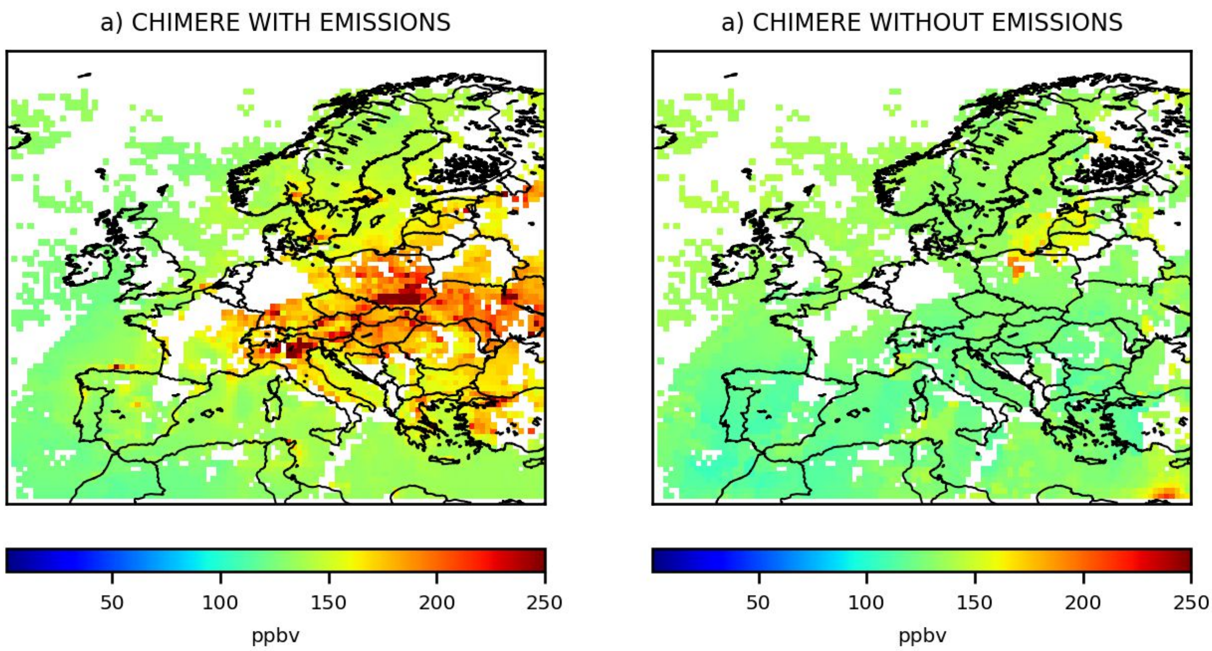

Figure 4. *CO surface concentrations simulated by CHIMERE a) with anthropogenic and biogenic emissions, and b) without emissions, in ppbv, at the 0.5°x0.5° grid-cell resolution, over Europe averaged from the 1st to the 7th, March 2015.*

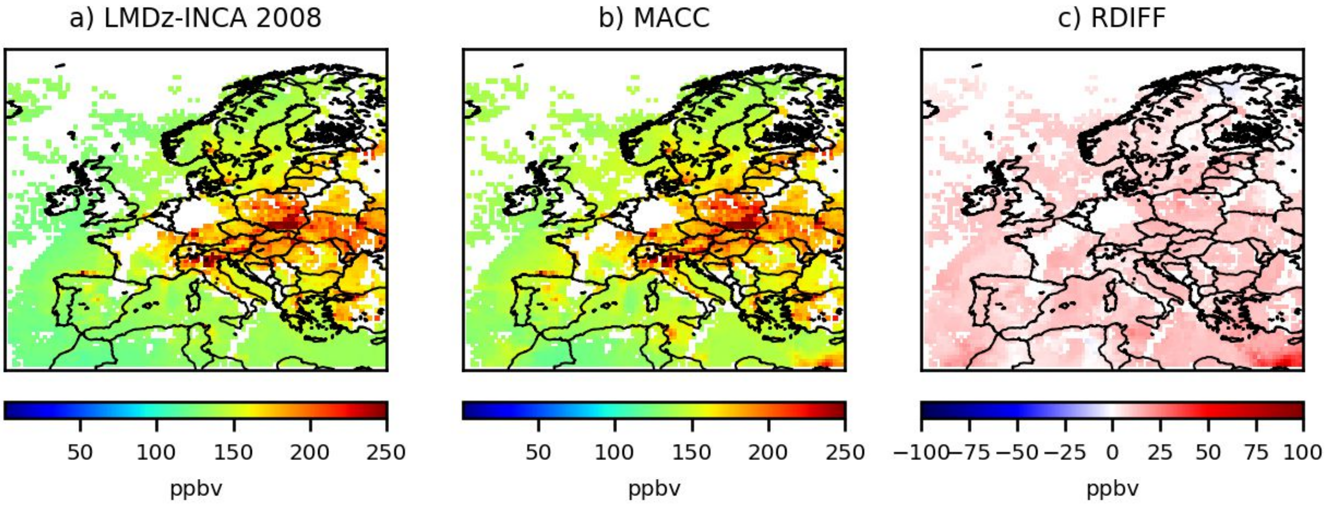

Figure 5. *CO surface concentrations simulated by CHIMERE using for initial and boundary conditions, a) the climatological values from the LMDZ-INCA global model b) the climatological values from a MACC reanalysis, in ppbv, and c) the relative differences between these two simulations , in %, at the 0.5°x0.5° grid-cell resolution, over Europe averaged from the* 1^{st} *to the* 7^{th}*, March 2015.*

4.1.4. Comparison between CHIMERE and the observations

Large discrepancies are found between the MOPITT CO observations (Figure 6b) and the prior simulation of their equivalents by CHIMERE over Europe (Figure 6a). For the first week of March 2015, CO concentrations are generally under-estimated by CHIMERE, particularly over Central and Eastern Europe (excepted in the south of Poland). On the contrary, CO concentrations seems to be over-estimated over Spain and Portugal. Large discrepancies are also found between the OMI NO_2 super-observations and the prior simulation of their equivalents by PYVAR-CHIMERE (Figure 7). Over Europe, the prior simulation strongly underestimates the tropospheric columns over industrial areas (e.g., over the Netherlands and over Po Valley).These discrepancies might be explained by an underestimation in the BU inventory due to a general trend in emissions (if the underestimation persists throughout the year) or to an underestimation regarding particular activity sectors or the time profiles at given scales (daily, monthly). This can also be explained by uncertainties from the satellite data or from the CTM (e.g., atmospheric production, chemistry with OH).

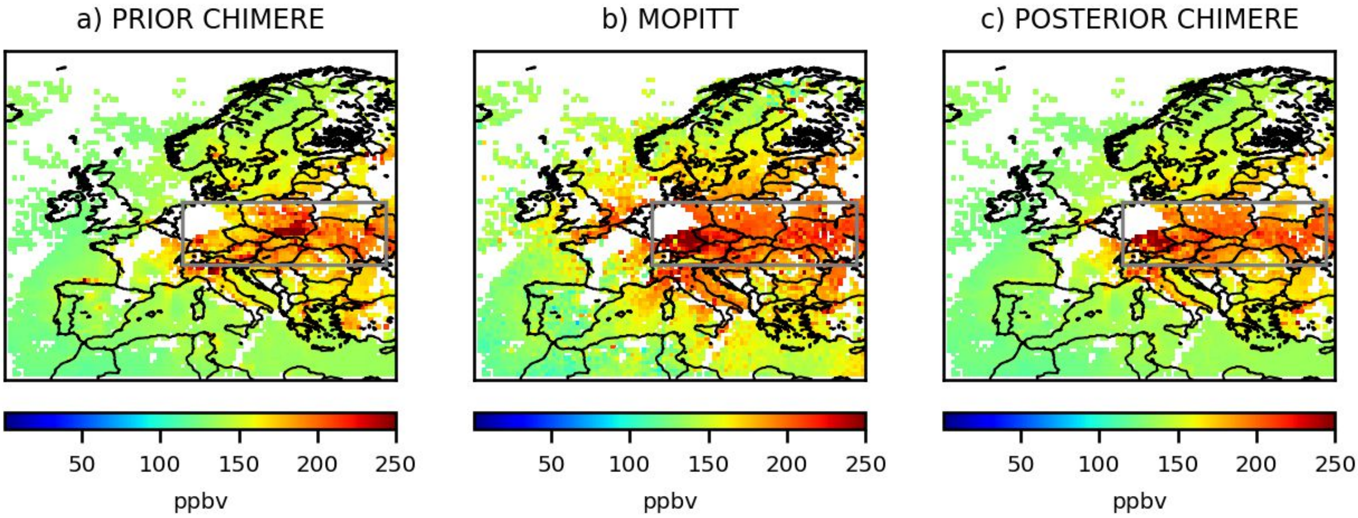

Figure 6. *CO collocated surface concentrations a) simulated by CHIMERE using the prior TNO-GHGco-v1 emissions and the climatological values from the LMDZ-INCA global model for initial and boundary conditions, b) observed by MOPITTv8-NIR-TIR and c) simulated by CHIMERE using the posterior emissions, in ppbv, at the 0.5°x0.5° grid-cell resolution, over Europe averaged from the 1st to the 7th, March 2015. Mean bias between simulations and observations are given in Section 4.2.3 for the area in the grey box.*

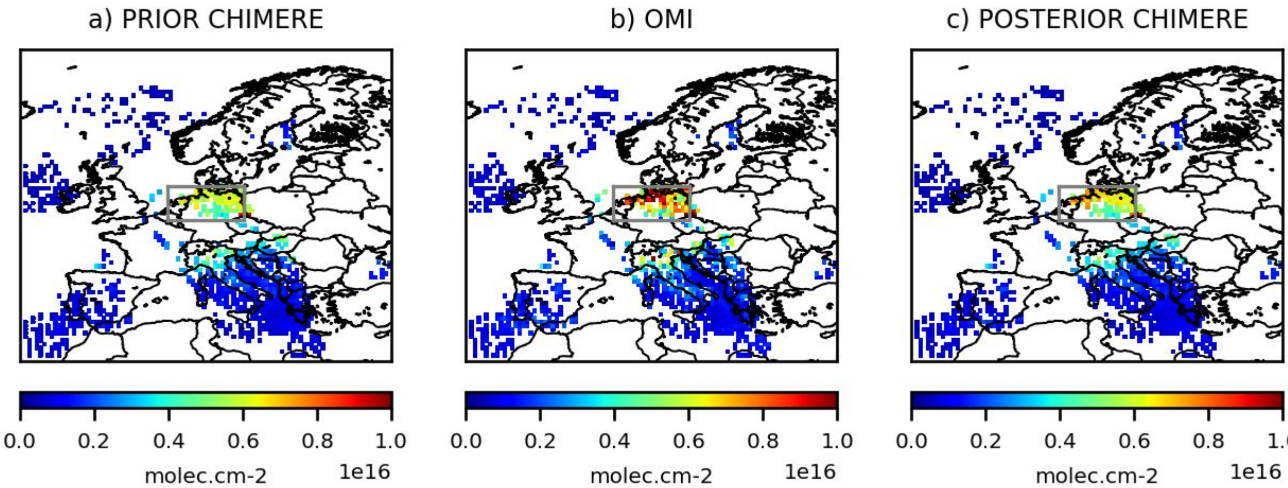

Figure 7. *NO_2 collocated surface concentrations a) simulated by CHIMERE using the prior TNO-GHGco-v1 emissions and the climatological values from the LMDZ-INCA global model for initial and boundary conditions, b) observed by OMI and c) simulated by CHIMERE using the posterior emissions, in $1e^{16}$ molec.cm-2, at the 0.5°x0.5° grid-cell resolution, over Europe the 19th, February 2015. Mean bias between simulations and observations are given in Section 4.2.4 for the area in the grey box.*

4.2. Inversions

4.2.1. Control vector x

For the CO inversion, the control vector **x** contains:

- the CO anthropogenic emissions for 7-day and at a 0.5° × 0.5° (longitude, latitude) resolution × 8 vertical levels, i.e. 101×85×8 grid cells,
- the CO 3D initial conditions at a 0.5° ×0.5° (longitude, latitude) resolution × 17 vertical levels,
- the CO lateral and top boundary conditions for 7-day at a 0.5° × 0.5° (longitude, latitude) resolution, i.e. (2x101 + 2x85) × 17 vertical levels.

Considering its short lifetime, there is no boundary conditions for NO_2. For the NO_x inversion, the control vector **x** contains:

- the NO and NO_2 anthropogenic emissions for 1-day and at a 0.5° ×0.5° (longitude, latitude) resolution× 8 vertical levels, i.e. 101×85×8 grid cells,
- the NO and NO_2 3D initial conditions at a 0.5° × 0.5° (longitude, latitude) resolution × 17 vertical levels.

4.2.2. Covariance matrices B and R

We hardly have sources of estimates ofthe uncertainties in bottom-up emission inventories at the 0.5° resolution. The characterization of their statistics in the inversion configuration is consequently often linked with crude assumptions from the inverse modelers. In the NO_x inversions, for both the prior NO and NO_2 emissions at 1-day and 0.5° resolution, the prior error standard deviations is assigned to 30% of the prior estimate of the emissions. As indicated in Section 3.3 and in Table 1, it is possible to use correlations in **B**, as in Broquet et al., [2011], in Broquet et al., [2013] and in Kadygrov et al., [2015]. For this NO_x illustration, spatial correlations are defined by an e-folding length of 50km over land and over the sea.

Even though annual CO emissions in Western Europe may be well known, with uncertainties of 6% according to Super et al., [2020], larger uncertainties could affect Eastern Europe. Moreover, large uncertainties still affect bottom-up emission inventories at the 0.5° resolution: spatial disaggregation of the national scale estimates to provide such gridded estimates causes a significant increase in the uncertainty for CO [Super et al., 2020]. For the inversion of CO emissions, the error standard deviations assigned to the prior CO emissions at 7-day and 0.5° resolution are 100%. For this CO illustration, the covariance matrix B of the prior errors is defined as diagonal (i.e. only variances in the individual control variables listed in 4.2.1 are taken into account). With such a set-up, in theory, we could obtained negative posterior emissions since the inversion system does not impose a constraint of positivity in the results. Nevertheless, even 100% of uncertainty lead to a prior distribution mostly (>80%) on the positive side. The assimilation of data showing an increase above the background (at the edges of the domain; not shown) further drive the inversion towards positive emissions for both CO and NO_x inversions. In practice, our inversion does not lead to negative posterior emissions (Figure 7b). Spatial and temporal correlations in **B** would further limit the probability to get negative emissions locally by smoothing the posterior emissions at a spatial scale at which the "aggregated" prior uncertainty is smaller than 100%. However, a positivity constraint should be implemented in future versions of the system.

Based on the sensitivity test in Figure 5, the errors assigned to the CO lateral boundary conditions and to their initial conditions are set at 15%. As these relative errors are significantly lower than those for the emissions and as variations in the CO surface concentrations are mainly drivenby emissions (Figure 4), we assume a small relative influence of the correction of initial and boundary conditions on our results. The variance of the individual observation errors in **R** is defined as the quadratic sum of the measurement error reported in the MOPITT and the OMI data sets, and of the CTM errors (including chemistry and transport errors and representativity errors) set at 20% of the retrieval values. The representativity errors could have been reduced with the choice of a finer CTM resolution (e.g., with a resolution closer to the size of the satellite pixel). Error correlations between the super-observations are neglected, so that the covariance matrix **R** of the observation errors is diagonal.

4.2.3 Inversion of CO emissions

Ten iterations are needed to reduce the norm of the gradient of J by 90% with the minimization algorithm M1QN3 and obtain the increments, i.e. the corrections provided by the inversion. The prior CO emissions over Europe for the first week of March 2015 and their increments are shown in Figure 7. As expected from the large differences between the prior surface concentrations (Figure 6a) and the MOPITT observations (Figure 6b), local increments can reach more than +50% (Figure 8b). CO emissions are increased over Central and Eastern Europe, except in the south of Poland. On the contrary, CO emissions are decreased over Spain and Portugal.The analyzed concentrations are the concentrations simulated by CHIMERE with the posterior fluxes: as expected, the optimization of the fluxes improves the fit of the simulated concentrations to the observations (Figure 6c), particularly over Central and Eastern Europe. Over this area (see the grey box in Figure 6), the mean bias between the simulation and the observations has been reduced by about 27% when using the posterior emissions (mean bias of 11.6 ppbv) instead of the prior emissions (mean bias of 15.9 ppbv).

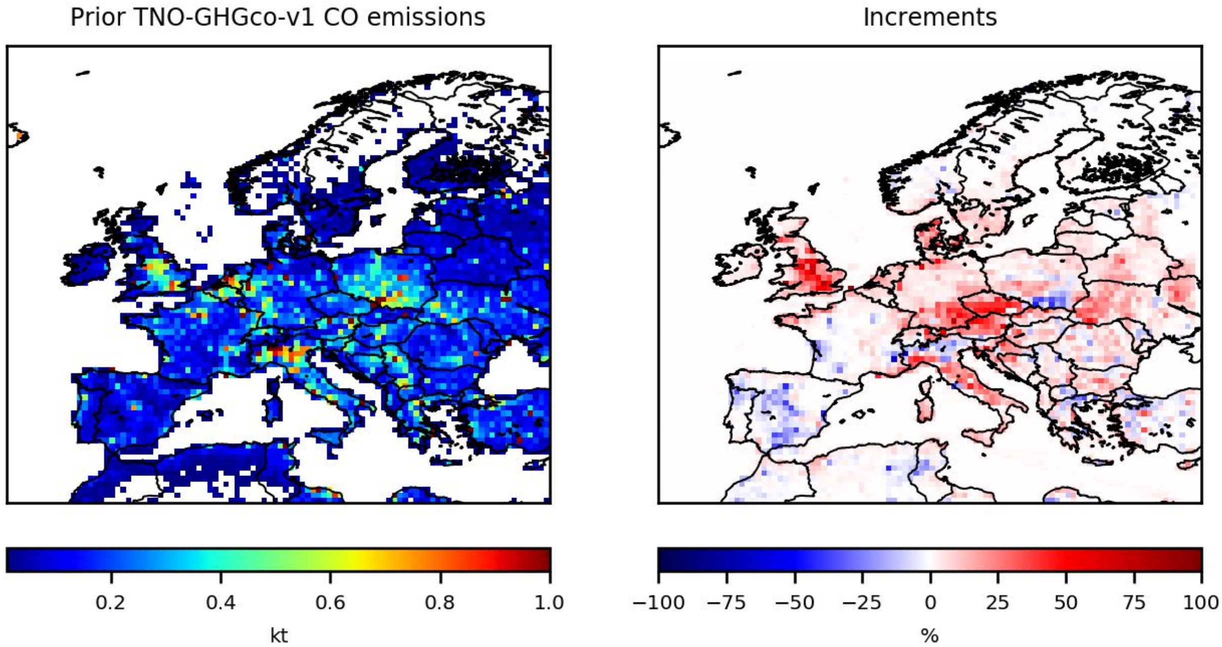

Figure 8. *a) TNO-GHGco-v1 CO anthropogenic prior emissions, in ktCO/grid-cell and b) increments provided by the inversion with constraints from MOPITTv8-NIR-TIR from the 1st to the 7th, March 2015, in %.*

4.2.4. Inversion of NO_x emissions

The prior NO_x emissions are shown in Figure 9a. Three iterations are needed to reduce the norm of the gradient of *J* by 90% with the minimization algorithm M1QN3 and obtain the increments shown in Figure 9b. As expected from the underestimation of the prior tropospheric columns in Figure 7, local increments may be large, for example over industrial areas (e.g., over the Po Valley) and over the Netherlands, with increments of more than +30% (Figure 9b).The analyzed NO_2 tropospheric columns in Figure 7c are the columns simulated by CHIMERE with the NO_2 posterior fluxes: as expected, the optimization of the fluxes improves the fit of the simulated concentrations to the observations, particularly over the Netherlands. Over this area (see the grey box in Figure 7), the mean bias between the simulation and the observations has been reduced by about 24% when using the posterior emissions (mean bias of $1.4e^{+15}$ molec.cm^{-2}) instead of the prior emissions (mean bias of $1.8e^{+15}$ molec.cm^{-2}). The posterior emissions and their uncertainties will have to be evaluated and may bring hints to the cause of the discrepancies.

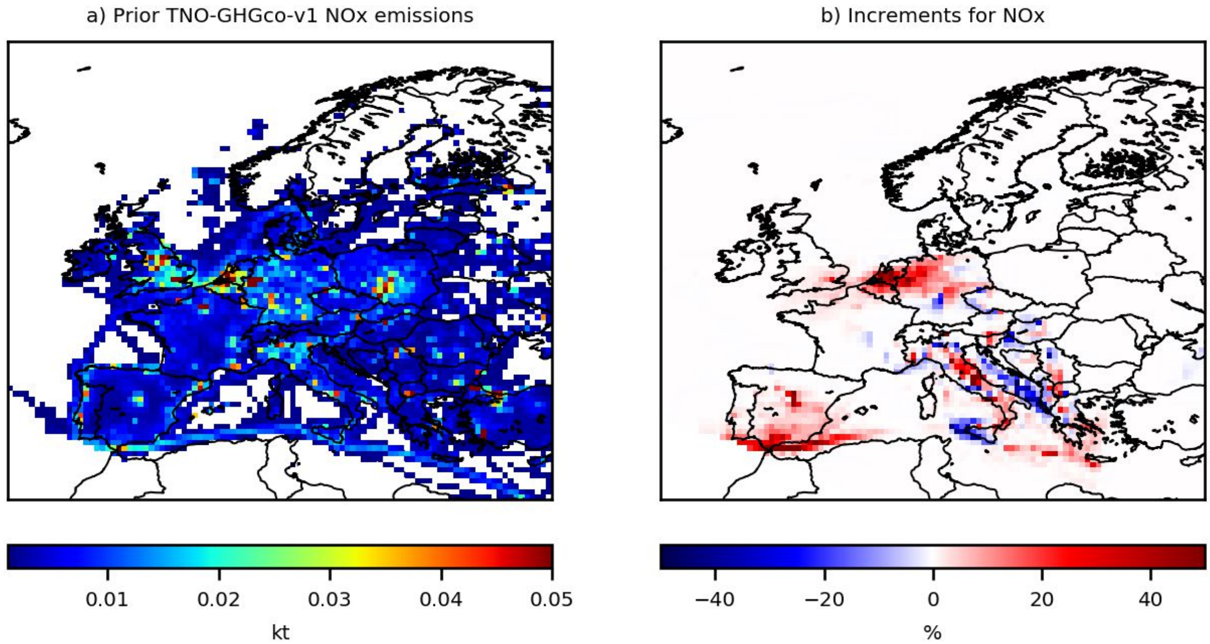

Figure 9. *a) TNO-GHGco-v1 NOx anthropogenic prior emissions, in ktNO_2/grid-cell and b) increments provided by the inversion with constraints from OMI the 19th, February 2015, in %.*

5. Conclusion/Discussion

This paper presents the Bayesian variational inverse system PYVAR-CHIMERE, which has been adapted to the inversion of reactive species such as CO and NO_x, taking advantage of the previous developments for long-lived species such as CO_2 [Broquet et al., 2011] and CH_4 [Pison et al., 2018]. We show the potential of PYVAR-CHIMERE, with inversions for CO and NO_x illustrated over Europe. PYVAR-CHIMERE will now be used to infer CO and NO_x emissions over long periods, e.g. first for a whole season or year and then for the recent decade 2005-2015 in the framework of the H2020 VERIFY project over Europe, and in the framework of the ANR PolEASIA over China, to quantify their trend and their spatio-temporal variability.

The PYVAR-CHIMERE system can handle any large number of both control parameters and observations. It will be able to cope with the dramatic increase in the number of data in the near future with, for example, the high-resolution imaging (pixel of 7x3.5 km^2) of the new Sentinel-5P/TROPOMI program, launched in October 2017. These new space missions with high-resolution imaging have indeed the ambition to monitor atmospheric chemical composition for the quantification of anthropogenic emissions. Moreover, a step forward in the joint assimilation of co-emitted pollutants will soon be possible with the PYVAR-CHIMERE system and the availability of TROPOMI co-localized images of CO and NO_2. This should improve the consistency of the inversion results and can be used to inform inventory compilers, and subsequently improve emission inventories. Moreover, this development will help in further understanding air quality problems and addressing air quality related emissions at the national to subnational scales.

Author Contribution

All authors have contributed to the manuscript writing (main authors: AFC, GB, IP and GD) and to the development of the present version of the PYVAR-CHIMERE system (main developer: IP). IP and GD have parallelized the adjoint version from Menut et al., [2000], Menut et al., [2003] and Pison et al., [2007]. IP has complemented the adjoint of new parameterizations since the CHIMERE release in 2011 and the tangent-linear model.

Code and Data Availability

OMI QA4ECV NO_2 product can be found here: http://temis.nl/qa4ecv/no2.html.
MOPITTv8-NIR-TIR CO product can be found here: ftp://l5ftl01.larc.nasa.gov/MOPITT/
The CHIMERE code is available here: www.lmd.polytechnique.fr/chimere/.

The associated documentation of PYVAR-CHIMERE is available on the website https://pyvar.lsce.ipsl.fr/doku.php/3chimere:headpage. The documentation includes a whole description of PYVAR-CHIMERE and several tutorials on how to run a first PYVAR-CHIMERE simulation or how to run an inversion.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

We acknowledge L. Menut and C. Schmechtig for their contributions to the development work on the adjoint code of CHIMERE and its parallelization. We acknowledge the TNO team (H.A. Denier van der Gon, J. Kuenen, S. Dellaert, S.Jonkers, A. Visschedijk, et al.) for providing NO_x and CO emissions over Europe.We also acknowledge the free use of tropospheric NO_2 column data from the OMI sensor from http://temis.nl/qa4ecv/no2.htmland the free use of CO surface concentrations from the MOPITT sensor from ftp://l5ftl01.larc.nasa.gov/MOPITT/. For this study, A. Fortems-Cheiney was funded by the French Space Agency-Centre National d'EtudesSpatiales CNES and by the H2020 VERIFY project, funded by the European Commission Horizon 2020 research and innovation programme, under agreement number 776810. L. Costantino was funded by the PolEASIA ANR project under the allocation ANR-15-CE04-0005.This work was granted access to the HPC resources of TGCC under the allocation A0050107232made by GENCI. Finally, we wish to thank F. Marabelle (LSCE) and his team for computer support.

References

Belikov, D. A., Maksyutov, S., Yaremchuk, A., Ganshin, A., Kaminski, T., Blessing, S., Sasakawa, M., Gomez-Pelaez, A. J., and Starchenko, A.: Adjoint of the global Eulerian–Lagrangian coupled atmospheric transport model (A-GELCA v1.0): development and validation, Geosci. Model Dev., 9, 749-764, https://doi.org/10.5194/gmd-9-749-2016, 2016.

Boersma, K. F., Vinken, G. C. M., and Eskes, H. J.: Representativeness errors in comparing chemistry transport and chemistry climate models with satellite UV–Vis tropospheric column retrievals, Geosci. Model Dev., 9, 875-898, https://doi.org/10.5194/gmd-9-875-2016, 2016.

Boersma, K. F., Eskes, H., Richter, A., De Smedt, I., Lorente, A., Beirle, S., Van Geffen, J., Peters, E., Van Roozendael, M. and Wagner, T.: QA4ECV NO_2 tropospheric and stratospheric vertical column data from OMI (Version 1.1) [Data set], Royal Netherlands Meteorological Institute (KNMI), http://doi.org/10.21944/qa4ecv-no2-omi-v1.1, 2017.

Bousquet, P., P. Ciais, P. Peylin, M. Ramonet, and P. Monfray: Inverse modeling of annual atmospheric CO_2 sources and sinks: 1. Method and control inversion, *J. Geophys. Res*., 104(D21), 26,161 – 26,178, doi:10.1029/1999JD900342, 1999.

Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P.: A European summertime CO_2 biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, *J. Geophys. Res*., 116, D23303, doi: 10.1029/2011JD016202, 2011.

Broquet, G., Chevallier, F., Bréon, F.-M., Kadygrov, N., Alemanno, M., Apadula, F., Hammer, S., Haszpra, L., Meinhardt, F., Morguí, J. A., Necki, J., Piacentino, S., Ramonet, M., Schmidt, M., Thompson, R. L., Vermeulen, A. T., Yver, C., and Ciais, P.: Regional inversion of CO_2 ecosystem fluxes from atmospheric measurements: reliability of the uncertainty estimates, Atmos. Chem. Phys., 13, 9039–9056, https://doi.org/10.5194/acp-13-9039-2013, 2013.

Chevallier, F., M. Fisher, P. Peylin, S. Serrar, P. Bousquet, F.-M. Bréon, A. Chédin, and P. Ciais: Inferring CO2 sources and sinks from satellite observations: method and application to TOVS data, *J. Geophys. Res*., 110, D24309, doi:10.1029/2005JD006390, 2005.

Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E., Bousquet, P., Brunke, E. G., Ciattaglia, L., Esaki, Y., Fröhlich, M., Gomez, A., Gomez-Pelaez, A. J., Haszpra, L., Krummel, P. B., Langenfelds, R. L., Leuenberger, M., Machida, T., Maignan, F., Matsueda, H., Morguí, J. A., Mukai, H., Nakazawa, T., Peylin, P., Ramonet, M., Rivier, L., Sawa, Y., Schmidt, M., Steele, L. P., Vay, S. A., Vermeulen, A. T., Wofsy, S., and Worthy, D.: CO_2 surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements, J. Geophys. Res., 115, 1–17, https://doi.org/10.1029/2010jd013887, 2010.

CHIMERE documentation, https://www.lmd.polytechnique.fr/chimere/docs/CHIMEREdoc2017.pdf, Last update of this documentation: June 8, 2017,2017.

Deeter, M. N., Edwards, D. P., Francis, G. L., Gille, J. C., Mao, D., Martinez-Alonso, S., Worden, H. M., Ziskin, D., and Andreae, M. O.: Radiance-based Retrieval Bias Mitigation for the MOPITT Instrument: The Version 8 Product, Atmos. Meas. Tech. Discuss., https://doi.org/10.5194/amt-2019-41, in review, 2019.

Ding, J., Miyazaki, K., van der A, R. J., Mijling, B., Kurokawa, J.-I., Cho, S., Janssens-Maenhout, G., Zhang, Q., Liu, F., and Levelt, P. F.: Intercomparison of NO_x emission inventories over East Asia, *Atmos. Chem. Phys*., 17, 10125-10141, https://doi.org/10.5194/acp-17-10125-2017, 2017.

EEA, Air quality in Europe - 2018 report, 12/2018, https://www.eea.europa.eu/publications/air-quality-in-europe-2018.

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749-3769, https://doi.org/10.5194/acp-7-3749-2007, 2007.

EMEP/EEA air pollutant emission inventory guidebook, 2016.

EMEP/CEIP, https://ceip.at/ms/ceip_home1/ceip_home/webdab_emepdatabase/emissions_emepmodels/

de Foy, B., Lu. Z. and Streets, D.G.: Satellite NO_2 retrievals suggest China has exceeded its NOx reduction goals from the twelfth Five-Year Plan, *Nature Scientific Reports*, 6:35912, 2016.

Fisher, M., and P. Courtier:Estimating the covariance matrices of analysis and forecast error in variational data assimilation, Tech. Mem. 220, 26 pp., Eur. Cent. for Medium-Range Weather Forecasting, Reading, U. K., 1995.

Gilbert, J., and C. Lemaréchal (1989), Some numerical experiments with variable storage quasi Newton algorithms,Math. Program., 45, 407–435.

Hein, R., et coll.: An inverse modeling approach to investigate the global atmospheric methane cycle, *Global. Biogeochem.Cycles*, 11, 43-76, 1997.

Henze, D. K., Hakami, A., and Seinfeld, J. H.: Development of the adjoint of GEOS-Chem, Atmos. Chem. Phys., 7, 2413–2433, https://doi.org/10.5194/acp-7-2413-2007, 2007.

Hooghiemstra, P. B., Krol, M. C., Bergamaschi, P., de Laat, A. T. J., van der Werf, G. R., Novelli, P.C., Deeter, M. N., Aben, I., and Rockmann, T.: Comparing optimized CO emission estimates usingMOPITT or NOAA surface network observations, *J. Geophys. Res.*, 117, D06309,doi:10.1029/2011JD017043, 2012.

Kadygrov, N., Broquet, G., Chevallier, F., Rivier, L., Gerbig, C., and Ciais, P.: On the potential of the ICOS atmospheric CO_2 measurement network for estimating the biogenic CO_2 budget of Europe, Atmos. Chem. Phys., 15, 12765–12787, https://doi.org/10.5194/acp-15-12765-2015, 2015.

Konovalov, I. B. et coll.: Inverse modelling of the spatial distribution of NO emissions on a continental scale using satellite data, *Atmos. Chem. Phys.*, 6, 1747-1770, doi:10.5194/acp-6-1747-2006, 2006.

Konovalov, I. B., Beekmann, M., Burrows, J. P., and Richter, A.: Satellite measurement based estimates of decadal changes in European nitrogen oxides emissions, *Atmos. Chem. Phys.*, 8, 2623-2641, doi:10.5194/acp-8-2623-2008, 2008.

Konovalov, I. B., Beekmann, M., Richter, A., Burrows, J. P., and Hilboll, A.: Multi-annual changes of NO_x emissions in megacity regions: nonlinear trend analysis of satellite measurement based estimates, *Atmos. Chem. Phys.*, 10, 8481-8498, doi:10.5194/acp-10-8481-2010, 2010.

Koohkan, M. R., Bocquet, M., Roustan, Y., Kim, Y., and Seigneur, C.: Estimation of volatile organic compound emissions for Europe using data assimilation, *Atmos. Chem. Phys*., 13, 5887-5905, https://doi.org/10.5194/acp-13-5887-2013, 2013.

Kuenen, J. J. P., Visschedijk, A. J. H., Jozwicka, M., and Denier van der Gon, H. A. C.: TNO-MACC_II emission inventory; a multi-year (2003–2009) consistent high-resolution European emission inventory for air quality modelling, Atmos. Chem. Phys., 14, 10963-10976, https://doi.org/10.5194/acp-14-10963-2014, 2014.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, *Atmos. Chem. Phys*., 13, 11019-11058, doi:10.5194/acp-13-11019-2013, 2013.

Lattuati, M., Impact des émissions européennes sur le bilan de l'ozone troposphérique a l'interface de l'europe et de l'atlantique nord : apport de la modélisation lagrangienne et des mesures en altitude, Ph.D. thesis, Université Paris VI, 1997.

Lelieveld, J., Klingmüller, K., Pozzer, A., Pöschl, U., Fnais, M., Daiber, A., Münzel, T.; Cardiovascular disease burden from ambient air pollution in Europe reassessed using novel hazard ratio functions, *European Heart Journal*, , ehz135, https://doi.org/10.1093/eurheartj/ehz135, 2019.

Lin, J.-T., McElroy, M. B., and Boersma, K. F.: Constraint of anthropogenic NO_x emissions in China from different sectors: a new methodology using multiple satellite retrievals, *Atmos. Chem. Phys.*, 10, 63-78, doi:10.5194/acp-10-63-2010, 2010.

Liu, F., Beirle, S., Zhang, Q., Dörner, S., He, K., and Wagner, T.: NO_x lifetimes and emissions of cities and power plants in polluted background estimated by satellite observations, Atmos. Chem. Phys., 16, 5283–5298, https://doi.org/10.5194/acp-16-5283-2016, 2016.

Locatelli, R., Bousquet, P., Saunois, M., Chevallier, F., and Cressot, C.: Sensitivity of the recent methane budget to LMDz sub-grid-scale physical parameterizations, Atmos. Chem. Phys., 15, 9765-9780, https://doi.org/10.5194/acp-15-9765-2015, 2015.

Mailler S., L. Menut, D. Khvorostyanov, M. Valari, F. Couvidat, G. Siour, S. Turquety, R. Briant, P. Tuccella, B. Bessagnet, A. Colette, L. Letinois, and F. Meleux, CHIMERE-2017: from urban to hemispheric chemistry-transport modeling ,*Geosci. Model Dev*., 10, 2397-2423, https://doi.org/10.5194/gmd-10-2397-2017, 2017.

Menut, L., R. Vautard, M. Beekmann, and C. Honoré: Sensitivity of photochemical pollution using the adjoint of a simplified chemistry-transport model, *J. Geophys. Res.*, 105, 15,379–15,402, 2000.

Menut L.: Adjoint modelling for atmospheric pollution processes sensitivity at regional scale during the ESQUIF IOP2, Journal of Geophysical Research - Atmospheres, 108, D17, https://doi.org/10.1029/2002JD002549, 2003.

Menut, L., Goussebaile, A., Bessagnet, B., Khvorostiyanov, D., and Ung, A.: Impact of realistic hourly emissions profiles on air pollutants concentrations modelled with CHIMERE, Atmospheric Envionment, 49, 233–244, doi:10.1016/j.atmosenv.2011.11.057, 2012.

Menut, L., Bessagnet, B., Khvorostyanov, D., Beekmann, M., Blond, N., Colette, A., Coll, I., Curci, G., Foret, G., Hodzic, A., Mailler, S., Meleux, F., Monge, J.-L., Pison, I., Siour, G.,Turquety, S., Valari, M., Vautard, R., and Vivanco, M. G.: CHIMERE 2013: a model for regional atmospheric composition modelling, *Geosci. Model Dev.*, 6, 981–1028, doi:10.5194/gmd-6-981-2013, 2013.

Mijling, B., and R. J. van der A: Using daily satellite observations to estimate emissions of short-lived air pollutants on a mesoscopic scale, *J. Geophys. Res.*, 117, D17302, doi:10.1029/2012JD017817, 2012.

Mijling, B., et al., Regional nitrogen oxides emission trends in East Asia observed from space, *Atmos. Chem. Phys.*, 3, 12003, 2013.

Muller, J.-P., Kharbouche, S., Gobron, N., Scanlon, T., Govaerts, Y., Danne, O., Schultz, J., Lattanzio, A., Peters, E., De Smedt, I., Beirle, S., Lorente, A., Coheur, P. F., George, M., Wagner, T., Hilboll, A., Richter, A., Van Roozendael, M., and Boersma, K. F.: Recommendations (scientific) on best practices for retrievals for Land and Atmosphere ECVs (QA4ECV Deliverable 4.2 version 1.0), 186 pp., available at: http://www.qa4ecv.eu/sites/default/ files/D4.2.pdf (last access: 12 April 2018), 2016.

Owens, R. G. and Hewson, T.: ECMWF Forecast User Guide, Reading, https://doi.org/10.21957/m1cs7h,https://software.ecmwf.int/wiki/display/FUG/Forecast+User+Guide, 2018.

Pison, I., Menut, L., and Bergametti, G.: Inverse modeling of surface NOx anthropogenic emission fluxes in the Paris area during the ESQUIF campaign, *J. Geophys. Res. Atmos.*, 112, D24302, doi:10,1029/2007JD008871, 2007.

Pison, I.,Bousquet, P., Chevallier, F., Szopa, S., and Hauglustaine, D.: Multi-species inversion of CH_4, CO and H_2 emissions from surface measurements, *Atmospheric Chemistry and Physics*, 9, 5281-5297, 2009.

Pison, I., A. Berchet, M. Saunois, .How a European network may constrain methane emissions at the French national scale. *Atmospheric Chemistry and Physics*, 2018.

Stavrakou, T. and J.-F. Müller: Grid-based versus big region approach for inverting CO emissions using Measurement of Pollution in the Troposphere (MOPITT) data, *Journal of Geophysical Research: Atmospheres*,111, D15, 2006.

Stavrakou, T., Muller, J.-F., Boersma, K. F., De Smedt, I., and van der A, R. J.: Assessing the distribution and growthrates of NOx emission sources by inverting a 10-year record of NO_2 satellite columns, *Geophys. Res. Lett.*, 35, 1–5, doi:10.1029/2008GL033521, 2008.

Super, I., Dellaert, S. N. C., Visschedijk, A. J. H., and Denier van der Gon, H. A. C.: Uncertainty analysis of a European high-resolution emission inventory of CO_2 and CO to support inverse modelling and network design, Atmos. Chem. Phys., 20, 1795–1816, https://doi.org/10.5194/acp-20-1795-2020, 2020.

Szopa, S., Foret, G., Menut, L., and Cozic, A.: Impact of large scale circulation on European summer surface ozone: consequences for modeling, *Atmospheric Environment*, 43, 1189–1195,doi:10.1016/j.atmosenv.2008.10.039, 2008.

Talagrand, O. : Assimilation of observations : an introduction, J. Met. Soc., Japan, 75, 191-209, 1997.

Valin, L. C., Russell, A. R., and Cohen, R. C.: Variations of OH rad-ical in an urban plume inferred from NO2 column measurements, Geophys. Res. Lett., 40, 1856–1860, doi:10.1002/grl.50267, 2013.

van der A, R. J., Mijling, B., Ding, J., Koukouli, M. E., Liu, F., Li, Q., Mao, H., and Theys, N.: Cleaning up the air: Effectiveness of air quality policy for SO_2 and NO_x emissions in China, *Atmos. Chem. Phys.*, 17, 1775-1789, 2017.

Vestreng, V., Breivik, K., Adams, M., Wagner, A., Goodwin, J., Rozovskaya, O., andOacyna, J. . Inventory Review 2005 - Emission Data reported to CLRTAP and under the NEC Direc-tive - Initial review for HMs and POPs .EMEP Status report, Norwegian Meteorological Institute, Oslo, 2005.

Yin, Y., Chevallier, F., Ciais, P., Broquet, G., A. Fortems-Cheiney, Pison, I. and Saunois, M: Decadal trends in global CO emissions as seen by MOPITT, *Atmos. Chem. Phys*., 15, 13433-13451, 2015.

Yumimoto, K. and Uno, I.: Adjoint inverse modeling of CO emissions over Eastern Asia using four-dimensional variational data assimilation, Atmospheric Environment, 40, 35, 6836-6845, DOI: 10.1016/j.atmosenv.2006.05.042, 2006.

Wang, Y., G. Broquet, P. Ciais, F. Chevallier, F. Vogel, N. Kadygrov, L. Wu, Y. Yin, R. Wang and S. Tao: Estimation of observation errors for large-scale atmospheric inversion of CO_2emissions from fossil fuel combustion, Tellus B: Chemical and Physical Meteorology, 69:1, DOI: 10.1080/16000889.2017.1325723, 2017.

Wang, Y., Broquet, G., Ciais, P., Chevallier, F., Vogel, F., Wu, L., Yin, Y., Wang, R., and Tao, S.: Potential of European $^{14}CO_2$ observation network to estimate the fossil fuel CO_2 emissions via atmospheric inversions, Atmos. Chem. Phys., 18, 4229–4250, https://doi.org/10.5194/acp-18-4229-2018, 2018.

WHO World Health Organization: Ambient Air Pollution: a global assessment of exposure and burden of disease, 2016.

Zheng, T., French, N. H. F., and Baxter, M.: Development of the WRF-CO2 4D-Var assimilation system v1.0, Geosci. Model Dev., 11, 1725–1752, https://doi.org/10.5194/gmd-11-1725-2018, 2018.

Zheng, B., Chevallier, F., Yin, Y., Ciais, P., Fortems-Cheiney, A., Deeter, M. N., Parker, R. J., Wang, Y., Worden, H. M., and Zhao, Y.: Global atmospheric carbon monoxide budget 2000–2017 inferred from multi-species atmospheric inversions, Earth Syst. Sci. Data, 11, 1411–1436, https://doi.org/10.5194/essd-11-1411-2019, 2019.