



Estimation of observation errors for large-scale atmospheric inversion of CO₂ emissions from fossil fuel combustion

By YILONG WANG^{1*}, GRÉGOIRE BROQUET¹, PHILIPPE CIAIS¹, FRÉDÉRIC CHEVALLIER¹, FELIX VOGEL¹, NIKOLAY KADYGROV¹, LIN WU¹, YI YIN¹, RONG WANG¹ and SHU TAO², ¹*Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ- Université Paris Saclay, Gif-sur-Yvette Cedex, France;* ²*Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, Beijing, China*

(Manuscript received 22 December 2016; in final form 27 April 2017)

ABSTRACT

National annual inventories of CO₂ emitted during fossil fuel consumption (FFCO₂) bear 5–10% uncertainties for developed countries, and are likely higher at intra annual scales or for developing countries. Given the current international efforts of mitigating actions, there is a need for independent verifications of these inventories. Atmospheric inversion assimilating atmospheric gradients of CO₂ and radiocarbon measurements could provide an independent way of monitoring FFCO₂ emissions. A strategy would be to deploy such measurements over continental scale networks and to conduct continental to global scale atmospheric inversions targeting the national and one-month scale budgets of the emissions. Uncertainties in the high-resolution distribution of the emissions could limit the skill for such a large-scale inversion framework. This study assesses the impact of such uncertainties on the potential for monitoring the emissions at large scale. In practice, it is more specifically dedicated to the derivation, typical quantification and analysis of critical sources of errors that affect the inversion of FFCO₂ emissions when solving for them at a relatively coarse resolution with a coarse grid transport model. These errors include those due to the mismatch between the resolution of the transport model and the spatial variability of the actual fluxes and concentrations (i.e. the *representation errors*) and those due to the uncertainties in the spatial and temporal distribution of emissions at the transport model resolution when solving for the emissions at large scale (i.e. the *aggregation errors*). We show that the aggregation errors characterize the impact of the corresponding uncertainties on the potential for monitoring the emissions at large scale, even if solving for them at the transport model resolution. We propose a practical method to quantify these sources of errors, and compare them with the precision of FFCO₂ measurements (i.e. the *measurement errors*) and the errors in the modelling of atmospheric transport (i.e. the *transport errors*). The results show that both the representation and measurement errors can be much larger than the aggregation errors. The magnitude of representation and aggregation errors is sensitive to sampling heights and temporal sampling integration time. The combination of these errors can reach up to about 50% of the typical signals, i.e. the atmospheric large-scale mean afternoon FFCO₂ gradients between sites being assimilated by the inversion system. These errors have large temporal auto-correlation scales, but short spatial correlation scales. This indicates the need for accounting for these temporal auto-correlations in the atmospheric inversions and the need for dense networks to limit the impact of these errors on the inversion of FFCO₂ emissions at large scale. More generally, comparisons of the representation and aggregation errors to the errors in simulated FFCO₂ gradients due to uncertainties in current inventories suggest that the potential of inversions using global coarse-resolution models (with typical horizontal resolution of a couple of degrees) to retrieve FFCO₂ emissions at sub-continental scale could be limited, and that meso-scale models with smaller representation errors would effectively increase the potential of inversions to constrain FFCO₂ emission estimates.

Keywords: fossil fuel, radiocarbon, global atmospheric inversion, observation error, Europe

*Corresponding author. email: yilong.wang@lsce.ipsl.fr

1. Introduction

Emissions from combustion of fossil fuels are the primary driver of increasing atmospheric CO₂ (Ballantyne et al., 2015). Improved knowledge of FFCO₂ emissions and their trends is necessary to understand the drivers of their variations, as well as to measure the effectiveness of mitigation actions (Pacala et al., 2010). Accurate estimates of emissions for the baseline years and the years after help verifying agreed-upon emission reduction targets. Implicitly, this requires that the uncertainties in the estimates of the emissions are much smaller than the amount of emissions to be reduced over a certain period of time.

Currently, fossil fuel CO₂ emissions are established by inventories mainly at the scale of countries, based on energy or fuel use statistics. In these inventories, sectorial data concerning each activity that produces emissions are multiplied by combustion efficiencies and emission factors. Such inventories thus have uncertainties related to imperfect data of energy or fuel use statistics, combustion efficiencies and emission factors (Macknick, 2009; Andres et al., 2012; Liu et al., 2015). Emission inventories are self-reported by countries using non-comparable methodologies and different data-sets (Ciais et al., 2010), although the IPCC has published guidelines of good practice for emission reporting (IPCC, 2006). It is estimated that national annual FFCO₂ emissions have two-sigma uncertainties ranging from 5% in OECD countries (Marland, 2008), 15–20% for China (Gregg et al., 2008) to 50% or more for less-developed countries (Andres et al., 2014). Global FFCO₂ emission maps (e.g. EDGAR, <http://edgar.jrc.ec.europa.eu> (Olivier et al., 2005); PKU-CO₂, (Wang et al., 2013); CDIAC, (Andres et al., 1996); ODIAC, (Oda and Maksyutov, 2011)) are compiled based on these national inventories and on the disaggregation of national (regional) emissions, or by bottom-up modelling of emissions based on local to regional activity data (Gurney et al., 2009). These products are available at a relatively high spatial resolution, typically down to 0.1°, but often without considering detailed spatial variations in emission processes. Also, different downscaling assumptions result in disagreements between emission maps (Oda and Maksyutov, 2011; Wang et al., 2013). These products usually provide annual values without temporal profiles associated with emissions at the intra annual scale. Thus, these emission maps often have larger uncertainties at sub-national and monthly scale (Ciais et al., 2010; Gregg et al., 2008).

An appealing method to independently assess FFCO₂ emissions is to use an atmospheric inversion approach (Ray et al., 2014). The atmospheric inversion approach consists in adjusting the estimates of emissions to minimize the distance between modelled and observed mixing ratios, yielding an optimized posterior estimate. It uses a statistical method, which relies on statistics of the uncertainty in the prior estimate of the emissions and of the other sources of model-measurement misfits (transport errors, measurement errors, model-measurement

mismatch due to different spatial representativeness, etc., which are grouped under the generic term *observation errors*). Atmospheric inversions have been used so far for estimating natural CO₂ fluxes, with most studies being at the scale of large regions (Bousquet et al., 2000; Gurney et al., 2002), and few studies at the scale of small regions (Lauvaux et al., 2008; Broquet et al., 2011). These inversions have mainly used ground-based *in situ* atmospheric measurements while exploiting satellite measurements is presently challenging (Chevallier and O'Dell, 2013).

A first strategy to sample the atmosphere with *in situ* stations for the inversion of FFCO₂ emissions would be to place stations very close to the largest fossil fuel CO₂ sources (cities, power plants, etc.). This allows the detection of a clear signature of FFCO₂ emissions in the measured CO₂ gradients (Bréon et al., 2015). Very high-resolution inversion systems are required to exploit such data (Brioude et al., 2012; McKain et al., 2012; Newman et al., 2013; Bréon et al., 2015). A limitation of this sampling strategy is that it would necessitate dense networks and very high-resolution inversions around every large CO₂ emitting area, while smaller sources will not be captured.

The second strategy is to sample the atmosphere away from local FFCO₂ sources to monitor an atmospheric signal integrating their signature at the sub-continental scale. With this strategy, one may expect inversions to solve for fossil fuel emissions at the scale of sub-continental regions (e.g. middle-sized countries in EU, groups of States in the US, provinces in China) using a network of stations distributed across a large sub-continental domain (Pacala et al., 2010). This sampling strategy could benefit from the existing infrastructure of *in situ* networks already set-up for the monitoring of natural fluxes (e.g. the European Integrated Carbon Observing System, ICOS, <https://www.icos-ri.eu/>; NOAA-ESRL, <http://www.esrl.noaa.gov/research/themes/carbon/>).

A difficulty for inversions to solve for FFCO₂ emissions based on atmospheric observations on a continental scale network is to separate the signal from fossil fuels from that of natural (biogenic and oceanic) fluxes in the atmospheric measurements. The effect of natural fluxes on atmospheric CO₂ gradients is generally much larger than that of fossil fuel emissions at the sort of sites like ICOS or NOAA-ESRL, especially during the growing season (Shiga et al., 2014), and at least comparable to that of fossil fuel emissions during non-growing season (Levin and Karstens, 2008), if the stations are not immediately close to anthropogenic sources. A filtering of the FFCO₂ signature based on knowledge on the spatial distribution and temporal profiles of FFCO₂ emissions is presently challenging because of uncertainties in the spatial and temporal distribution of emissions and because large-scale transport models can hardly account for the potential of this information, which is concentrated at relatively high resolution. Shiga et al. (2014) analysed real measurements to study the potential of the surface observation networks to monitor anthropogenic emissions, and in particular to separate the signals of fossil fuel emissions from those of natural fluxes,

in North America. However, in practice, there has not been any attempt at conducting inversions of the emissions at large-scale using real CO₂ measurements alone from existing continental networks.

To circumvent the problem of separating natural fluxes and fossil fuel emissions in the atmospheric signals, it is possible to use proxies of the CO₂ mole fraction from fossil fuel emissions in large-scale inversions. Several proxies have been proposed for FFCO₂ (Gamnitzer et al., 2006; Rivier et al., 2006), but none of them is as close to a pure fossil fuel CO₂ tracer as radiocarbon in CO₂. Measurements of radiocarbon in CO₂ together with measurements of total CO₂ can be used to separate FFCO₂ (Levin et al., 2003) based on the principle that fossil fuel-emitted CO₂ comes from geological deposits, and is radiocarbon-free. In this context, our study gives insights on the potential of the inversion of fossil fuel emissions in Europe based on hypothetical networks of collocated measurements of radiocarbon in CO₂ and total CO₂ measurements.

Note that radiocarbon in CO₂ is only a proxy of FFCO₂ and that its atmospheric gradients are also partly influenced by the transport of fluxes from stratosphere, ocean, biosphere and nuclear facilities as well as by that of fossil fuel emissions (Randerson et al., 2002; Naegler and Levin, 2006; Graven and Gruber, 2011). Within industrialized continents, radiocarbon gradients are however dominated by the signal of FFCO₂ emissions (Graven and Gruber, 2011; Levin et al., 2011). In our studies, we postulate that atmospheric radiocarbon-CO₂ observations are exact measurements of the FFCO₂ component in atmospheric CO₂. We also postulate that numerous measurements of radiocarbon-CO₂ could be made at many sites of a continental atmospheric network. In practice, radiocarbon is expensive to measure (e.g. can only be performed in discrete air samples, not *in situ*), so that the implementation costs of dense radiocarbon sampling networks could be a limitation as well.

Nevertheless, these two assumptions are not a limitation to the scope of this study focusing on evaluating whether the signal of FFCO₂ gradients between continental sites that are not in the vicinity of high emission areas are large enough compared to modelling errors and radiocarbon measurement errors, and whether these gradients are representative enough of the emissions averaged at sub-national scales so that the use of a coarse-grid transport model remains valid for constraining sub-national FFCO₂ emissions.

The recent OSSE study of Ray et al. (2014) demonstrated that using a network of 35 towers sampling atmospheric FFCO₂ mixing ratios every 3 h across the U.S. with an uncertainty arbitrarily set to 0.1 ppm (which is very optimistic given the current precision of radiocarbon-CO₂ measurements), an atmospheric inversion at 1° × 1° resolution could reduce errors on eight-day-averaged country-level fossil-fuel emissions by a factor of two. In the context of the US Inter-academy report on emission verification, Pacala et al. (2010) presented

another OSSE experiment suggesting that, based on a hypothetical massive set of 10,000 atmospheric ¹⁴CO₂ measurements in one year and a perfect transport model of 5° horizontal resolution, an atmospheric inversion could reduce the uncertainty of the monthly mean fossil-fuel flux in the US from 100% to less than 10%. Moreover, Basu et al. (2016) developed a dual-tracer inversion framework assimilating both CO₂ and ¹⁴CO₂. They showed that given the actual coverage of ¹⁴CO₂ measurements available in 2010 over US, the dual-tracer inversion can recover the US national annual total FFCO₂ emission to better than 1%.

In this study, we attempt at analysing in detail the weight of potentially critical limitations when targeting large-scale budgets of fossil fuel emissions based on atmospheric inversion. In particular, we characterize how much such an approach relies on the knowledge of the spatial distribution of the emissions at high resolution, while continental scale observation networks and inversion systems can hardly solve for it. When dealing with a large-scale inversion system, one also needs to carefully account for the fact that the grid size of transport models (typically 100–300 km for global models, down to 5–10 km for regional models; Law et al., 2008) is larger than the scale of emissions, which have very fine scale patterns. This ensemble of misfits between the scales solved for or modelled within the inversion system and that of actual emissions and patterns in the mixing ratios generates so-called aggregation and representation errors in the inversion (Gerbig et al., 2003; Lin et al., 2006) to which this paper gives a special attention. Actually, it will be shown in our study that the aggregation errors still characterize the impact of uncertainties in the distribution of the emissions on the potential for monitoring the emission at large scale when solving for them at the transport model resolution.

This study specifically addresses the derivation and analysis of the statistics of representation and aggregation errors in comparison to the typical FFCO₂ signals at measurement sites. This work focuses on the derivation and analysis of these errors for an atmospheric inversion framework dedicated to the inference of national-scale monthly emissions over European countries using continental scale networks of measurement stations. This inversion framework uses a global atmospheric transport model and global maps of the emissions with spatial and temporal distributions within countries and one month. Having a global configuration ensures that uncertainties in fossil fuel CO₂ emitted over other regions of the globe outside a target continent are properly accounted for. This study assumes that daily to monthly mean FFCO₂ gradients can be estimated between numerous sites and a ‘reference’ site sampling the free tropospheric air over a continent by ¹⁴CO₂ measurements, with a precision of 1 ppm due to the typical measurement errors and to uncertainties in the conversion of ¹⁴CO₂ and CO₂ measurements into FFCO₂ (Levin et al., 2003).

The detailed objectives of this paper are:

- To develop a theoretical derivation of the different sources of observation errors arising from the estimation of fossil fuel emissions at regional scale by an atmospheric inversion using a coarse-grid transport model. We provide a theoretical definition of the aggregation and representation errors and to separate them from two other types of observation errors: the measurement errors and the model transport errors. This synthetic derivation of critical sources of errors, which have been analysed for inversion of natural fluxes in various studies (Kaminski et al., 2001; Engelen, 2002; Gerbig et al., 2003; Wu et al., 2011), is adapted to the inversion of FFCO₂ emissions.
- To derive practical estimates of the representation and aggregation errors based on the above theoretical definitions.
- To compare the representation and aggregation errors to simpler estimates of the model transport and measurements errors, to the signal of FFCO₂ simulated at the sites of continental scale networks, and to the corresponding statistics of errors due to the uncertainties in the prior estimates of the emissions at large scale (i.e. the errors that the inversion aims at filtering with the model–data comparisons). While the specific error values are function of the inversion configuration and while we compute them for FFCO₂ observations in Europe only, our analysis gives useful insights into typical sources of errors and signals intrinsically related to the monitoring of fossil fuel emissions at large scale.

Due to the link between the representation and aggregation errors with the configuration of the inversion, Section 2 first describes the large-scale fossil fuel emission inversion framework, and then develops the derivation for each term of the observation errors mentioned above. Special attention is given to the representation and aggregation errors when using a coarse-grid transport model and optimizing emissions at the scale of sub-continental regions. The section also discusses the significance of these two errors and the actual dependence of the underlying sources of uncertainties to the specific configuration of the inverse modelling framework. Practical ways to estimate these two errors are given in Section 3. Results for representation errors, aggregation errors and the errors due to the prior uncertainties in the simulation of observations in Europe are discussed in Sections 4 and 5. In Section 6, we compare these errors with the measurement error, model transport error, and typical signals of FFCO₂. We also discuss the effects of the spatial (temporal) resolution of the modelling (respectively observation) framework for the atmospheric inversion of FFCO₂ emissions. Conclusions are drawn in Section 7.

2. Methodology

The inversion framework considered here follows the Bayesian linear update (Enting et al., 1993; Tarantola, 2005) of a prior

statistical knowledge $p(\mathbf{x}^t | \mathbf{x}^b)$ on the actual value $\mathbf{x}^t$ for a set of control variables $\mathbf{x}$ (among which some variables underlie the target quantities i.e. budgets of FFCO₂ emissions at large scale), where $\mathbf{x}^b$ is a prior estimate of these variables. The update relies on some observations $\mathbf{y}^o$ (here FFCO₂ atmospheric measurements), on an affine observation operator $\mathbf{x} \mapsto \mathbf{H}\mathbf{x} + \mathbf{y}_{\text{fixed}}$ (including the global coarse-grid transport model and the distribution of the emissions at high resolution, and the signature of the influence of sources of FFCO₂ that is not solved for by the inversion) linking the control space $\mathbf{x}$ to the observation space $\mathbf{y}$ and on statistics $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{y}_{\text{fixed}} | \mathbf{x}^t)$ of the sources of observation errors (i.e. errors that are not due to the uncertainties in the estimate of $\mathbf{x}^t$ in the comparison between $\mathbf{H}\mathbf{x}^t + \mathbf{y}_{\text{fixed}}$ and the observations $\mathbf{y}^o$). It also follows the traditional assumption that the statistics of the prior and observation uncertainties are unbiased, Gaussian and independent of each other (Tarantola, 2005) so that $p(\mathbf{x}^t - \mathbf{x}^b | \mathbf{x}^b) \sim \mathcal{N}(\mathbf{0}, \mathbf{B})$ and $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{y}_{\text{fixed}} | \mathbf{x}^t) \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$ where $\mathbf{B}$ and $\mathbf{R}$ are the prior error and observation error covariance matrices, and so that the posterior statistical estimate of $\mathbf{x}^t$ from the optimal update given $\mathbf{x}^b$ and $\mathbf{y}^o$, is a Gaussian distribution that can be written $p(\mathbf{x}^t | \mathbf{x}^b, \mathbf{y}^o) \sim \mathcal{N}(\mathbf{x}^a, \mathbf{A})$, where

$$\mathbf{A} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \quad (1)$$

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{A} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b - \mathbf{y}_{\text{fixed}}) \quad (2)$$

We focus on the characterization of several critical terms of the observation error $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{y}_{\text{fixed}} | \mathbf{x}^t)$, on the relevance of the assumption that the observation error can be represented by a Gaussian and unbiased distribution $\mathcal{N}(\mathbf{0}, \mathbf{R})$, and on the derivation of a relevant $\mathbf{R}$ matrix for our configuration of a large-scale fossil-fuel emission inversion. The observation error plays a critical role in the estimate of the posterior uncertainty characterized by its covariance matrix $\mathbf{A}$. If its projection back to the flux space (i.e. the term $\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$ in Equation (1)) is far larger than the uncertainty in the fluxes that the inversion is expected to solve for (the $\mathbf{B}$ matrix), the assimilation of atmospheric observations will bring little and/or highly uncertain information about the fluxes and the potential of the inversion will be low. The observation error $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{y}_{\text{fixed}} | \mathbf{x}^t)$ will be compared to an estimate of the projection of the prior uncertainty in the observation space $p(\mathbf{H}(\mathbf{x}^t - \mathbf{x}^b) | \mathbf{x}^b)$ to give insights on this (indicating whether the signature of the prior uncertainty should be easy to filter in the prior model–data misfits $\mathbf{y}^o - \mathbf{H}\mathbf{x}^b - \mathbf{y}_{\text{fixed}}$) even though the full computation of Equation (1) is required to define whether the assimilation of atmospheric observation strongly decreases the uncertainty in the flux estimates.

The nature of the observation error strongly depends on the nature of the $\mathbf{x}$ and $\mathbf{y}$ space, and on the accuracy and precision of the observation operator $\mathbf{x} \mapsto \mathbf{H}\mathbf{x} + \mathbf{y}_{\text{fixed}}$. In the following we first present the practical configuration of these elements given our practical inversion framework. Then, we propose a theoretical decomposition of the observation errors with an emphasis

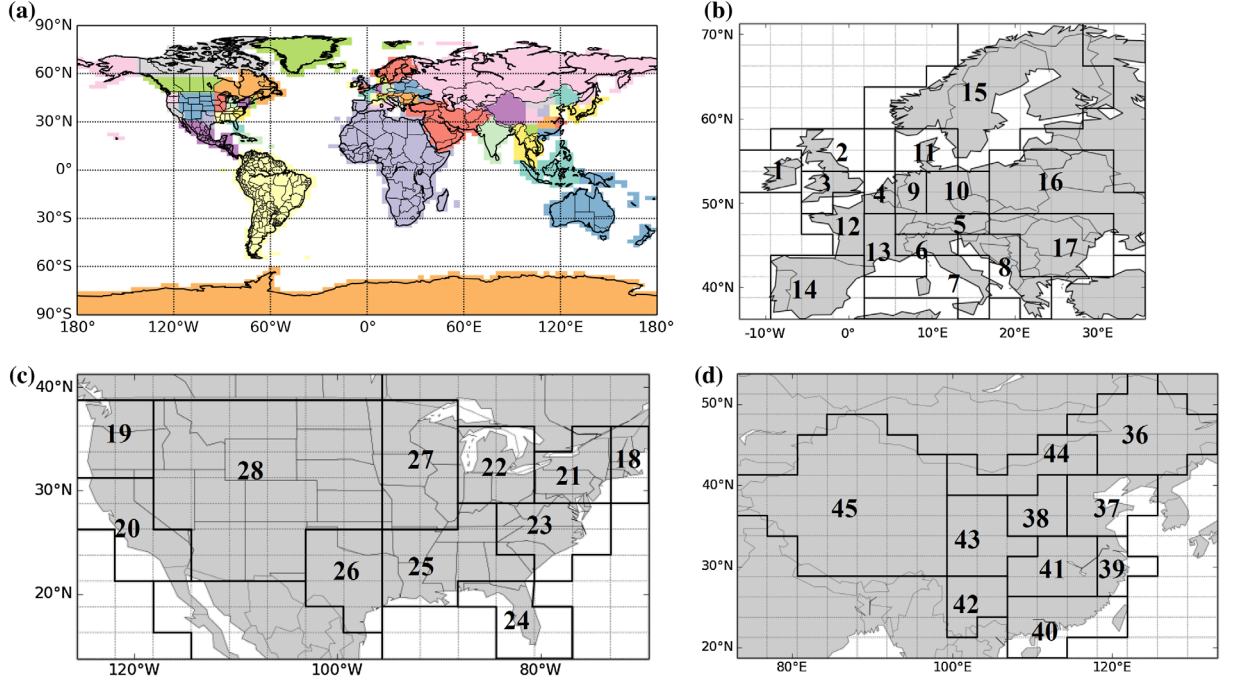


Fig. 1. (a) Map of the 56 regions whose monthly emission budgets are controlled by the inversion; (b) zoom on the 17 control regions in Europe; (c) zoom on the 11 control regions in the United States; (d) zoom on the 10 control regions in China.

on the terms that should be critical for our practical inversion framework and with a specific care at defining the representation and aggregation errors.

In practice the simulations, inversions and analysis are conducted for a 1-year period arbitrarily chosen to be a typical year 2007. This choice has consequences regarding the meteorological conditions and the level of emissions that are taken into account in our modelling framework but we expect that the conclusions from the analysis should not be strongly sensitive to this choice.

2.1. Configuration of the control and observation space of the inversion and of the observation operator

2.1.1. Control vector. We divide the globe, according to administrative boundaries, into a set of emitting regions whose monthly mean fossil fuel emission budgets are solved for during a whole year (Fig. 1a). The corresponding space discretization is higher in continents that have the largest emission densities (Europe, US and China, Fig. 1b–d). The spatial resolution in Europe is in agreement with the typical size of European countries. It is finer in western Europe where emissions can be high in specific regions such as northern Italy, southern England, eastern and western Germany. In the US and China, the spatial discretization is also increased in the most populated and industrialized areas (i.e. the east and west coasts in the US, and the south-eastern coast in China). In a given emitting

region, the inversion solves for the budget of FFCO₂ emissions (in Mg C/hour) for each of the 12 months during one year, but does not solve for the space and time distribution within each region or at sub-monthly intervals.

2.1.2. Observation vector. The observation vector consists in FFCO₂ gradients between sites of hypothetical ground-based European networks of atmospheric total CO₂ and ¹⁴CO₂ measurements (that are used together to compute FFCO₂) throughout one year, at typical heights for continuous measurement sites. More precisely, we consider gradients between simultaneous FFCO₂ observations at any site of these networks and a reference site sampling the free tropospheric air over Europe, as is traditionally done when analyzing ¹⁴CO₂ measurements (Levin et al., 2008). The actual sampling heights for the measurement sites are generally below 300 magl (Kadyrov et al., 2015). In this study, we first (in Section 4) consider a standard sampling height above the ground at all the sites except the reference site in each configuration of the observation vector: 100 magl. The choice of this standard height simplifies practical considerations for the analysis in this study but does not have the same physical impact at all sites due to the variations in the PBL and orography, and of the modelling skills depending on the locations and time. The sensitivity to other sampling heights will be discussed in Section 5.1. We select the High Alpine Research Station Jungfraujoch (JFJ, located at 3450 masl in Switzerland) as the reference site for European

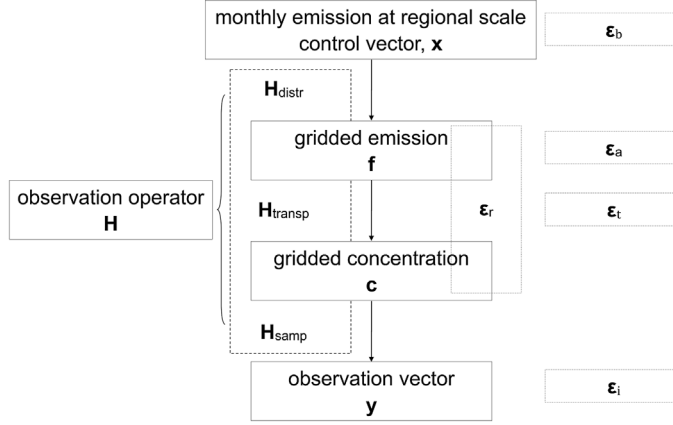


Fig. 2. Components of observation operator and its link with the control and observation vectors.

stations. Continuous measurement of total CO_2 has been made for years in Europe (within the CarboEurope-IP, GHG-Europe and ICOS programs) and US (within the NOAA-ESRL framework) at tens of sites. A given radiocarbon measurement can be applied to a sample with any temporal integration time from 1 h to 1 month since air samples could be filled at constant rates over long periods. However, the cost of the $^{14}\text{CO}_2$ analysis of one sample is presently high so that monitoring of $^{14}\text{CO}_2$ during a whole year favours the choice of integrated samples at the daily to monthly scale (Levin, 1980; Turnbull et al., 2009; Vogel et al., 2013). In this study, we only consider a single sampling frequency at all sites in each configuration of the observation vector. The two-week mean sampling is considered as a standard sampling strategy in Section 4, while the sensitivity to other sampling strategies will be discussed in Section 5.2. We have also accounted for the technical ability to have an intermittent filling (Levin et al., 2008; Turnbull et al., 2016). Indeed, state-of-art inversion systems generally make use of data during afternoon only due to limitations in modelling the vertical mixing during other periods of the day. We thus assume that mean afternoon FFCO_2 observations are sampled during 12:00–18:00 local time at the sites.

The locations of the stations where $^{14}\text{CO}_2$ measurements are made are assumed to be inland and distant from urban areas and other large sources, and aim to monitor the signature of the emissions at sub-continental scale. However, some sites will necessarily be closer to emitting areas (such as cities and power plants) than others, with consequences regarding the representativeness and amplitude of the measured FFCO_2 signal. We thus define two types of sites, both corresponding to land model grid cells: ‘urban’ and ‘rural’ sites, based on a threshold on the population density (ORNL, 2015) within the grid cells where the stations are located. This threshold is country-dependent and matches the World Bank urbanization data (available at http://data.worldbank.org/indicator/SP.URB.TOTL.IN.ZS?page=1&order=wbapi_data_value_2011%20wbapi_data_value%20wbapi_data_value-first&sort=asc) for each

country as done by Wang et al. (2013). In this study, the class ‘urban’ sites also include grid cells where large point sources exist, e.g. power plants (based on the CARMAv2 database, Ummel, 2012) and cement manufactures (Wang et al., 2013). The name ‘urban’ is used for convenience. One more straightforward approach would have been to select highest emitting grid cells according to EDG-IER with a threshold approximately consistent with the one on the population as discussed above. But we prefer to keep a level of independency to this inventory and to its uncertainties by taking an independent proxy of the high emitting locations.

2.1.3. Observation operator. The observations are only influenced by the initial condition and the emissions during the year. As indicated above, the emissions are solved for by the inversion. Through diffusion by atmospheric transport, the spatial gradients of FFCO_2 from a pulse of emissions at a given time appear to become negligible (with an amplitude smaller than 0.1 ppm) within about 2 weeks, so that the influence of the global FFCO_2 distribution on 1 January 2007 (i.e. the initial condition of the inversion experiments in our studies) is quite negligible for our simulations of gradients of FFCO_2 in Europe in 2007, even for the results in January 2007 (not shown here). In our modelling of framework and corresponding simulations, initial conditions for the FFCO_2 field in the atmosphere by the 1st day of the inversion year are thus ignored.

Consequently, the observation operator considered in this study is linear and does not bear an affine term $\mathbf{y}_{\text{fixed}}$ reflecting, in the observation gradients, the influence of a source or sink of FFCO_2 that is not rescaled by our control vector. Therefore, it can be denoted as $\mathbf{H}$. We decompose it into:

$$\mathbf{H} = \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}} \quad (3)$$

In this formulation, $\mathbf{H}$ is a chain of three operators denoting the distribution of emissions within each region-month corresponding to the control variables ($\mathbf{H}_{\text{distr}}$), the atmospheric transport ($\mathbf{H}_{\text{transp}}$), and the sampling of atmospheric gradients

corresponding to the observation vector from the transport model outputs ($\mathbf{H}_{\text{samp}}$), respectively. The spatial and temporal (sub-monthly) distribution operator $\mathbf{x} \rightarrow \mathbf{f} = \mathbf{H}_{\text{distr}} \mathbf{x}$ distributes the emission budgets for each region and month $\mathbf{x}$ into gridded emissions $\mathbf{f}$ at the spatial and temporal resolution expected as input of the atmospheric transport model. The atmospheric transport operator $\mathbf{f} \rightarrow \mathbf{c} = \mathbf{H}_{\text{transp}} \mathbf{f}$ simulates the FFCO₂ field $\mathbf{c}$ using an atmospheric transport model with prescribed emissions $\mathbf{f}$. The sampling operator $\mathbf{c} \rightarrow \mathbf{y} = \mathbf{H}_{\text{samp}} \mathbf{c}$ applies the atmospheric sampling procedure described above.

Each column of $\mathbf{H}$ represents the signature (the so-called response function) in the observation space of a unitary increment of the budget of the emissions in a given control region-month. Fig. 2 gives the frame of the observation operator and its link to control and observation vectors.

For the observation operator used in practice, we use a coarse-grid transport model and emission inventories which catch the typical spatial and temporal large-scale variations in the FFCO₂ emissions and concentrations and thus ensure the realism of the typical estimates of uncertainties in our study. The corresponding products bear the typical precision/accuracy of the products that are used by state-of-the-art inversion systems when assimilating real data to quantify CO₂ natural fluxes at large scale.

- (1) Inventory used for the mapping of the emissions at high resolution

We use the PKU-CO₂-2007 global emission inventory for 2007 (Wang et al., 2013) to model, by the $\mathbf{H}_{\text{distr}}$ operator, the spatial distribution of emissions within the regions of control. PKU-CO₂-2007 is a high-resolution (0.1°) annual emission map based on the disaggregation of national emission budgets using sub-national statistics. Regarding the sub-monthly temporal distribution of emissions within each month, we assume a flat temporal profile, as in many large-scale natural flux inversion systems (Peylin et al., 2013). We denote by $\mathbf{H}_{\text{distr}}^{\text{PKU}}$ the practical implementation of the distribution operator $\mathbf{H}_{\text{distr}}$.

- (2) Global transport model configuration

An off-line version of the atmospheric general circulation model of Laboratoire de Météorologie Dynamique (LMDZ) (version 4) (Hourdin et al., 2006) is used as our atmospheric transport operator. The corresponding LMDZ simulation was nudged to the reanalysed wind fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) Interim Reanalysis (ERA-Interim, (Berrisford et al., 2009)). LMDZ has participated to a series of intercomparison exercises for the simulation of CO₂ concentrations (Law et al., 2008) and is able to reproduce most of the daily variations of the large-scale transport of FFCO₂ (Peylin et al., 2011). The model configuration used here has a horizontal resolution of 3.75° × 2.5° (longitude × latitude) and 19 hybrid sigma-pressure layers to discretize the vertical profile between the surface and the top

of the atmosphere. We denote by $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ the resulting practical implementation of $\mathbf{H}_{\text{transp}}$.

- (3) Observation sampling of the transport model outputs

In the observation operator, the practical simulation of FFCO₂ gradients corresponding to the observation vector relies on the simple extraction of individual concentration data at the measurement locations and then on the computation of differences between these concentration at different sites. We extract a concentration for a given location by taking the value in the transport model grid cell within which the site locates rather than interpolating values from several transport model grid cells. Usually, the height of the first level of LMDZ is about 150 magl. All the observations being assumed at 100 magl, they are all extracted from the first level of this version of LMDZ, except that of the reference site, Jungfraujoch (JFJ). JFJ is located at 3450 m above sea level (masl) but close to the ground level, at the top of a mountain. Since the LMDZ model poorly solves the topography in mountain areas, its ground level in the grid cell corresponding to JFJ is located far lower than this height. In order to ensure that the modelled concentrations are representative of the free tropospheric air, JFJ observations are extracted from the sixth level of LMDZ, which is usually located between 2700 and 3800 masl. 1-day to 1-month mean afternoon FFCO₂ data are sampled in time by $\mathbf{H}_{\text{samp}}$ (depending on the corresponding single observation frequency, see Section 2.1.2). We denote by $\mathbf{H}_{\text{samp}}^{\text{coloc}}$ the resulting practical implementation of $\mathbf{H}_{\text{samp}}$.

To sum up, the observation operator that will be used in practice for inversions in the following can be written $\mathbf{H}^{\text{prac}} = \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{PKU}}$.

2.2. Theoretical derivation of the critical observation errors

In this section, we are interested in decomposing the observation error $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^i|\mathbf{x}^i)$ for a typical $\mathbf{H}$ in order to isolate some critical sources of errors in practice. The observation operator $\mathbf{H} = \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}}$ maps low-resolution budgets of the emissions into a coarse spatial grid. But each term of this operator is likely not perfectly represented in the following ways: (1) the products for the distribution of emissions within countries such as the one used to build $\mathbf{H}_{\text{distr}}^{\text{PKU}}$ are necessarily imperfect; (2) the-state-of-the-art transport model such as the one used in $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ are necessarily imperfect; (3) the spatial representativeness of the measurements close to the ground can be low with coarse-resolution transport models and it can be difficult to represent the measurements in the vertical grid of the coarse-resolution models (Broquet et al., 2011; Pillai et al., 2011) which impacts the precision/accuracy of practical models for $\mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}$. These add to the high measurement errors that have to be accounted for when monitoring FFCO₂.

Focusing on these sources of errors, the term $\mathbf{y}^o - \mathbf{H}\mathbf{x}^i$ can be decomposed as follows:

$$\begin{aligned}
\mathbf{y}^o - \mathbf{H}\mathbf{x}^t &= \underbrace{\left(\mathbf{y}^o - \mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t\right)}_{\epsilon_i} + \underbrace{\left(\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t\right)}_{\epsilon_r} \\
&+ \underbrace{\left(\mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}}^t \mathbf{x}^t\right)}_{\epsilon_t} \\
&+ \underbrace{\left(\mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}}^t \mathbf{x}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}} \mathbf{x}^t\right)}_{\epsilon_a}
\end{aligned} \tag{4}$$

where $\mathbf{H}_{\text{transpHR}}$ is a theoretical operator corresponding to the linear transport from emissions $\mathbf{f}_{\text{HR}}$ to $\mathbf{y}$, $\mathbf{f}_{\text{HR}}$ and this transport being represented using the ‘infinitely high’ resolution (i.e. continuously instead of using a discrete form) needed for catching all the patterns in the emissions and concentrations; Superscripts t denotes the true value of the emissions or observation operators at their corresponding space and time resolution (a ‘true observation operator’ meaning here a perfect operator without any model error). Even if this decomposition primarily aims at giving a physical characterization of each resulting term, it is also made such that these different terms can be assumed to be independent (see the justification for Equation (6) below).

We define the different terms of the observation error $\mathbf{y}^o - \mathbf{H}\mathbf{x}^t$ based on this decomposition:

- (1) $\mathbf{y}^o - \mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t$ corresponds to the ‘measurement error’ ϵ_i , which is associated with the precision of FFCO₂ gradients derived from measurements of ¹⁴C and CO₂. The assumption given in section 1 that this precision is 1 ppm is discussed in Section 2.3.
- (2) $\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t$ corresponds to the representation error ϵ_r which arises from the modelling of concentrations and emissions at the coarse resolution of the transport model in the observation operator. This error could be further split into errors due to missing high-resolution variations in the emissions at the model sub-grid scales, and errors due to comparing concentrations averaged at the model resolution to measurements with a far lower spatial representativeness. Appendix A1 discusses such a decomposition, which, in practice, artificially attributes most of the representation errors to the former or to the latter depending on the mathematical formulation. Therefore, even though this decomposition would have a physical meaning, it will be ignored hereafter.
- (3) $\mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}}^t \mathbf{x}^t$ corresponds to the transport errors ϵ_t due to the use of discretized and simplified equation for modelling the transport.
- (4) $\mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}}^t \mathbf{x}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}} \mathbf{H}_{\text{distr}} \mathbf{x}^t$ corresponds to the aggregation error ϵ_a due to the imperfect representation of the distribution of the monthly emissions within the region-months solved for by the inversion when using $\mathbf{H}_{\text{distr}}$.

The total observation error ϵ_o defined by $p(\mathbf{y}^o - \mathbf{H}\mathbf{x}^t - \mathbf{y}_{\text{fixed}} | \mathbf{x}^t)$ can be expressed as:

$$\epsilon_o = \epsilon_i + \epsilon_r + \epsilon_t + \epsilon_a \tag{5}$$

Several of these terms are proportional to the value of $\mathbf{x}^t$ while $\mathbf{x}^t$ can take any value in the statistical framework of our inversion problem. This prevents, theoretically, from computing a fixed covariance $\mathbf{R}$ of the observation error assuming that this error can be represented by a distribution $N(\mathbf{0}, \mathbf{R})$. The configuration of such an error in the inversion systems generally ignore such a dependence of the model errors (transport, representation and aggregation errors) on the possible values for the actual fluxes which is a strong limitation for the application of the traditional data assimilation framework to flux inversion problems. In practice, we will derive $\mathbf{R}$ based on assumptions regarding the typical value for $\mathbf{x}^t$ in our inversion cases.

The errors on the different components of the observation operator relates to strongly different underlying input datasets and different types of model (see Section 2.1.3) and are thus considered to be independent. Assuming that they are all Gaussian and unbiased, one can write that $\epsilon_i \sim N(\mathbf{0}, \mathbf{R}_i)$, $\epsilon_r \sim N(\mathbf{0}, \mathbf{R}_r)$, $\epsilon_t \sim N(\mathbf{0}, \mathbf{R}_t)$, $\epsilon_a \sim N(\mathbf{0}, \mathbf{R}_a)$, and compute $\mathbf{R}$ as the sum of the covariances of the different errors:

$$\mathbf{R} = \mathbf{R}_i + \mathbf{R}_r + \mathbf{R}_t + \mathbf{R}_a \tag{6}$$

Of note is that our formulations of the representation error and of the aggregation error are similar to the derivations of representation error by Gerbig et al. (2003) and of aggregation error by Engelen (2002), respectively. However, our formulation of the aggregation error slightly differs from that of Kaminski et al. (2001) and Bocquet et al. (2011). We use a sort of ‘bottom-up’ approach to derive it, starting from the decomposition of the observation errors once having defined it as the sum of all sources of model data misfits other than the prior uncertainties and that are independent from these prior uncertainties. Kaminski et al. (2001) and Bocquet et al. (2011) rather followed what we consider as a ‘top-down’ approach to derive this aggregation error. Indeed, their introduction of the covariance of the aggregation error in the observation error covariance matrix ensures that the computation of the statistics for $p(\mathbf{x}^t | \mathbf{y}^o, \mathbf{x}^b)$ is the same regardless of the control resolution. Due to the use of the usual assumption of the atmospheric inversion that the observation error is independent of the prior uncertainty, our ‘bottom-up’

method ignores potential correlations between the aggregation errors and the prior uncertainties, which is not the case of the ‘top-down’ approaches. Therefore, the formulations of the covariance of the aggregation error in Kaminski et al. (2001) and Bocquet et al. (2011) include a component related to this correlation, which is ignored in our formulation. As discussed in Appendix A2, we have nevertheless computed the corresponding component and concluded that its weight is relatively small and negligible for our study. The mathematical details and a discussion regarding the potential correlations between the aggregation errors and the prior uncertainties are given in Appendix A2.

2.3. *Insights on the specificity or generality of the observation errors investigated in this study*

In theory, results for regions and months targeted by the inversion do not vary with the resolution of the control vector if the aggregation error ϵ_a is perfectly accounted for by $\mathbf{R}$ in the inversion configuration (see the demonstration in the Appendix based on the notations given above in Section 2.2). This assumes that the uncertainty in the emissions at the scale of interest is independent of the uncertainty at higher resolution (which is approximately verified with our modelling set-up, see the Appendix). In other words, an inversion at coarse resolution that accounts for aggregation errors ϵ_a should give the same results for monthly fluxes over large regions as the same inversion applied to solve for hourly fluxes at the highest resolution (transport model grid). This is due to the equivalence between accounting for the uncertainties of fluxes within regions/month through their projection in the observation error or through their assigned prior uncertainty (given the assumptions underlying the inversion framework). In this sense, even though they are formally a function of the control vector, the aggregation errors at a scale larger than the transport model resolution are not specific to a given inverse modelling set-up. Considering that the choice of the control resolution reflects a targeted resolution for the fluxes, aggregation errors rather reflect the impact for the monitoring of the fluxes at this targeted resolution of the uncertainties in the distribution of the fluxes at higher spatial or temporal resolutions. Increasing the control resolution would thus not, in theory, help solving for fluxes at the targeted resolution.

On the opposite, representation error is strongly linked to a specific inversion configuration. Increasing the resolution of the transport model used for the inversion necessarily decreases them without a full compensation of this decrease by the rise of prior uncertainties. The transport errors should also depend on the transport modelling configuration. For example, synoptic patterns and the influence of the surface topography on the transport are better simulated at higher resolution. However, different transport models are also based on different parameterizations and computational approach, etc., which makes the quantification and evaluation of the transport errors

as a function of the model complicated and efforts have rather focused on the derivation of typical transport errors based on the spread of different transport models (Law et al., 2008; Peylin et al., 2011).

Finally, the errors in the measurements in our study should be fully independent of the inverse modelling framework. The 1 ppm measurement error for FFCO₂ gradients between sites corresponds to typical values based on the analysis of air samples by accelerator mass spectrometry (AMS) for ¹⁴CO₂ (2–3‰, (Vogel et al., 2010; Turnbull et al., 2014)) and by typical analyzers for continuous CO₂ samples (Chen et al., 2010; Turnbull et al., 2011). Apart from these errors, various fluxes that influence the atmospheric ¹⁴CO₂, such as those from cosmogenic production, ocean, biosphere and nuclear facilities, make the direct conversion into FFCO₂ gradients bear complex uncertainties whose typical values may exceed 1 ppm for some locations and periods of times (Hsueh et al., 2007; Bozhinova et al., 2013; Vogel et al., 2013). These additional sources of uncertainties are not included in this study. In addition, we assume that all ¹⁴CO₂ and CO₂ samples will be analysed in the same laboratory such as the present ICOS Central Radiocarbon Laboratory, or at least if the samples are measured by different instruments and laboratories, they will follow the official target of compatibility made by the World Meteorological Organization (WMO) for ¹⁴CO₂ measurements (GGMT-2013). The consequence is that there should not be significant biases associated with instrumental errors impacting the gradients between sites analysed in this study.

3. Practical calculation of observation errors

In the inversion system, we use $\mathbf{H}^{\text{prac}} = \mathbf{H}_{\text{sam}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{PKU}}$ as the observation operator. But here, we use a relatively independent representation of the ‘actual’ and higher resolution operators involved in the theoretical formulation of the observation errors in Section 2.2 in order to derive an estimate of these errors. These actual and higher resolution operators should bear patterns of the emissions, transport and concentration variability which should be realistic enough so that this estimate of the observation errors can provide a realistic characterization of the representation and aggregation errors when using real measurements.

A European configuration of the meso-scale transport model CHIMERE (Schmidt et al., 2001) run with a 0.5° horizontal resolution, with 25 hybrid sigma-pressure vertical layers from the surface to the pressure altitude of 450 hPa, and with hourly concentration outputs (to be aggregated into one-day to one-month mean afternoon data) is used to simulate $\mathbf{H}_{\text{transpHR}}^{\text{t}}$ and $\mathbf{H}_{\text{transp}}^{\text{t}}$. However, the LMDZ model is still used to model the practical $\mathbf{H}_{\text{transp}}^{\text{t}}$ when calculating the aggregation error. The CHIMERE simulations are initialized at 50 ppm at 1 January 2007.

We model $\mathbf{H}_{\text{transpHR}}^t$ by feeding CHIMERE with 0.5° resolution maps of the emissions and using the one-day to one-month mean afternoon 0.5° -resolution and 25 vertical-level concentration fields to extract simulated gradients of FFCO_2 . Since the resolution of this CHIMERE configuration is not infinitely high, in practice, a sampling operator is still needed to model $\mathbf{H}_{\text{transpHR}}^t$. The vertical resolution of CHIMERE being about 35–45 m for the first three levels, the 100 magl observations involved in the computation of the FFCO_2 gradients at high resolution are extracted in the third level of this model (and in the 0.5° horizontal model grid cell containing the horizontal position of the stations, using a sampling option similar to that used in $\mathbf{H}_{\text{samp}}^{\text{coloc}}$). The FFCO_2 concentrations at the reference site are extracted from the 23rd vertical level of CHIMERE corresponding to the altitude of 3450 masl in the 0.5° grid cell where the reference site located. This transport (and sampling) configuration that is used to model $\mathbf{H}_{\text{transpHR}}^t$ is denoted $\mathbf{H}_{\text{transpHR}}^{\text{CHIM}}$.

By degrading the horizontal and temporal resolution of the emissions in input of the CHIMERE model and by averaging (horizontally and vertically) the mole fractions in output of the CHIMERE simulations we model $\mathbf{H}_{\text{transp}}^t$. The spatial aggregation of the CHIMERE outputs consists first in a vertical aggregation, and then on a horizontal aggregation. The horizontal aggregation does not fully correspond to an aggregation within the LMDZ grid cells (i.e. to the interpolation of the 0.5° resolution fields for CHIMERE into the $3.75^\circ \times 2.5^\circ$ resolution grid of LMDZ). For simplicity, the 0.5° CHIMERE grid cells are rather aggregated from blocks of 6×6 grid cells to yield coarse grid at the 3° resolution which is close to that of the LMDZ grid. $\mathbf{H}_{\text{transp}}^{\text{CHIM}}$ denotes the configuration where CHIMERE is fed with emissions maps aggregated at 3° resolution (close to that of the LMDZ model) and over 3-h time windows, and where CHIMERE one-day to one-month mean afternoon output concentrations are, again, aggregated at 3° resolution. For the modelling of $\mathbf{H}_{\text{samp}}$ in the computation of the error due to aggregation at the transport model resolution and in the computation of the representation error we apply an operator which follows the principle of $\mathbf{H}_{\text{samp}}^{\text{coloc}}$ (and which we will thus also denote $\mathbf{H}_{\text{samp}}^{\text{coloc}}$) i.e. FFCO_2 observations are extracted in the first aggregated vertical levels for all the sites but the reference site, which is extracted in the sixth aggregated vertical level, and in the co-located aggregated 3° horizontal grid cells of the $\mathbf{H}_{\text{transp}}^{\text{CHIM}}$ outputs. When calculating the error due to the aggregation at region-month scale, we use $\mathbf{H}_{\text{samp}}^{\text{coloc}}$ to model $\mathbf{H}_{\text{samp}}$ and apply it to $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$.

Associating CHIMERE at 0.5° resolution with $\mathbf{H}_{\text{transpHR}}^t$ (and consequently 0.5° resolution maps of the emissions with $\mathbf{f}_{\text{HR}}^t$) assumes that the main variations (i.e. those which have the largest impact for data at 100 magl) of emissions or concentrations within 3° resolution grid cells occur at scales larger than 0.5° . Furthermore, simulating $\mathbf{H}_{\text{transpHR}}^t$ and $\mathbf{H}_{\text{transp}}^t$ with CHIMERE which is a regional model (over Europe) assumes that the aggregation and representation errors in Europe due to the

coarse representation of these emissions or of their signature outside Europe is negligible.

$\mathbf{H}_{\text{distr}}^t$ (with outputs at 3° and 3-h resolution), the distribution of $\mathbf{f}_{\text{HR}}^t$ at 0.5° and 1-h resolution and $\mathbf{x}^t$ are modelled using the $0.1^\circ \times 0.1^\circ$ EDGARv4.2 2007 emission map (<http://edgar.jrc.ec.europa.eu>) convoluted with temporal profiles (at 1-h resolution) from IER (available at <http://carbones.ier.uni-stuttgart.de/wms/index.html>). We denote this emission inventory EDG-IER afterwards. Aggregating this inventory at 1-h/ 0.5° resolution or at the scale of the inversion control region-month provides respectively $\mathbf{f}_{\text{HR}}^{\text{EDG-IER}}$ and $\mathbf{x}^{\text{EDG-IER}}$ that are used to model $\mathbf{f}_{\text{HR}}^t$ and $\mathbf{x}^t$. Aggregating this inventory at 3-h/ 3° resolution (when computing representation errors at the coarse transport resolution using CHIMERE) or 3-h/ $3.75^\circ \times 2.5^\circ$ (when computing aggregation errors using LMDZ) and then rescaling it homogeneously within each region/month of control for the inversions to get unitary budget of emissions provides $\mathbf{H}_{\text{distr}}^{\text{EDG-IER}}$ (using the same notation for the operator when the output ‘emission’ space is at 3° or $3.75^\circ \times 2.5^\circ$ resolution) which is used to model $\mathbf{H}_{\text{distr}}^t$.

With these practical choices for modelling, the operators involved in the different types of observation errors defined in Section 2.2, the representation error writes:

$$\mathbf{H}_{\text{transpHR}}^{\text{CHIM}} \mathbf{f}_{\text{HR}}^{\text{EDG-IER}} - \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{CHIM}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}} \quad (7)$$

and the aggregation error writes:

$$\mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}} - \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{PKU}} \mathbf{x}^{\text{EDG-IER}} \quad (8)$$

We only have one practical realization for each of these terms and thus of the corresponding errors, therefore, in order to derive their standard deviation and to investigate whether they bear potential temporal or spatial correlation, we make the strong assumption that the errors at different time and locations have relatively similar statistical distributions. However, this assumption of spatial and temporal homogeneity will be applied for adequate subset of observation time, locations and type, which will require a categorization of the observations.

Based on this assumption, we analyse the typical statistics of the representation and aggregation errors by using distributions of occurrences of these errors for different subsets (categories) of observations. Since observation sites of continental networks could locate in any grid cell, all the spatial grid cells and all one-day to one-month afternoon time windows are used and categorized among different subsets for this analysis. The different spatial and temporal categories will be defined based on the analysis of the spatial and temporal variations of the errors. The potential temporal auto-correlations and spatial correlations within/across categories are also analysed.

Using a similar approach, transport errors could have been evaluated using $\mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{CHIM}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}}$ and $\mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}}$. However, the spatial resolution and orography of $\mathbf{H}_{\text{transp}}^{\text{CHIM}}$ and $\mathbf{H}_{\text{transp}}^{\text{LMDZ}}$ are not exactly the same and this could have artificially increased the transport error with representation error. Therefore, we make a simpler estimation

of the transport errors for simulated FFCO₂ gradients based on that of transport errors for simulated FFCO₂ at individual sites.

For this estimation, we make several assumptions. First, we assume that there is no temporal auto-correlation of the transport error in simulated daily mean afternoon concentrations between different days at a given location. This assumption, however, may be violated if there are significant biases in the simulation of the meteorological conditions, i.e. the planetary boundary layer height and the vertical mixing strength (Miller et al., 2015; Basu et al., 2016). Nevertheless, the estimate of the structure of the temporal auto-correlations of the transport error is challenging (Lin and Gerbig, 2005; Lauvaux et al., 2009; Miller et al., 2015) and there is no clear evidence that such autocorrelations is significant at daily scale (Lin and Gerbig, 2005; Lauvaux et al., 2009; Broquet et al., 2011). In this study, we follow the assumption made by the majority of existing inversion studies (Peters et al., 2007; Chevallier et al., 2010; Niwa et al., 2012; Peylin et al., 2013), in which the temporal auto-correlations in the transport error are usually ignored. Second, we assume that the standard deviation of the transport error in simulated daily mean afternoon concentrations is constant in time at a given location. Finally, we assume that the ratio between this standard deviation and the temporal standard deviations of the 1-year long time series of the high-frequency variability of the detrended and deseasonalized simulated daily mean afternoon concentrations in the corresponding grid cell of the transport model is constant in space (i.e. that this ratio is identical for all grid cells of the transport model). The high-frequency variability is calculated by the method of Thoning et al. (1989). The underlying assumption is that the transport models should be less reliable at sites where the concentrations have a larger variability (Peylin et al., 2005; Geels et al., 2007).

These assumptions allow us to use the station of Saclay (SAC) near Paris for deriving a generalized estimate of the ratio between the transport errors and the simulated FFCO₂ temporal variability. According to Peylin et al. (2011), the annual average of the standard deviations between simulated hourly mean FFCO₂ concentrations at this site from a set of state-of-the-art transport models is 2.34 ppm. We use this value to define the standard deviation of the transport errors associated with the simulated daily afternoon mean concentrations. The standard deviation of the one-year long-time series of the daily afternoon mean concentrations simulated within one year with our practical implementation of the simulation of 3-hourly concentrations $\mathbf{H}_{\text{transp}}^{\text{LMDZ}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}}$ at SAC is 1.74 ppm. So the ratio between the standard deviation of the transport error for daily afternoon mean concentrations and the standard deviation of simulated time series for the daily afternoon mean concentrations within one year for any site is assumed to be $2.34/1.74 = 1.34$. For any potential sites (in any grid cells in LMDZ), we thus multiply this ratio by the standard deviation of the simulated daily afternoon mean concentrations within one year to get an estimate

of the transport error statistics for the daily afternoon mean FFCO₂. Transport errors for daily to monthly mean afternoon FFCO₂ are then derived based on the value obtained for daily afternoon mean FFCO₂ and on the above mentioned assumption that there is no temporal auto-correlation of the transport errors between afternoon mean concentrations in different days. For example, our estimate of the transport error for two-week mean afternoon concentrations (mean of 14 days) is equal to $1.74 \times 1.34 / \sqrt{14} = 0.62$ ppm at SAC site.

Following this estimation, the transport error in the two-week mean afternoon FFCO₂ concentration at JFJ site is 0.21 ppm. The transport error in the one-day to one-month mean afternoon FFCO₂ gradients between any site and JFJ is calculated assuming no spatial correlation of the transport errors between sites, i.e. as $\sqrt{(\epsilon_{i,i}^2 + \epsilon_{i,\text{JFJ}}^2)}$ where $\epsilon_{i,i}$ is the transport error for concentrations at site i and $\epsilon_{i,\text{JFJ}}$ is the transport error in concentrations at site JFJ at the corresponding one-day to one-month scale. As a result, the transport errors in the two-week mean afternoon FFCO₂ gradients from 100 magl sites to the JFJ reference site range from 0.42 to 1.07 ppm. The transport errors in the one-day mean afternoon FFCO₂ gradients range from 1.58 to 3.99 ppm (and from 0.29 to 0.74 ppm in the case of errors on 1-month mean afternoon FFCO₂ gradients, respectively).

As indicated in Section 2, we also want to compare the observation errors to the projection of the prior uncertainty in the observation space $p(\mathbf{H}(\mathbf{x}^i - \mathbf{x}^b)|\mathbf{x}^b)$ denoted $\mathbf{H}\epsilon_b$ (and called ‘prior FFCO₂ errors’ hereafter, ϵ_b corresponding to the prior uncertainties). Following the same approach as for the estimation of the representation and aggregation errors, and setting $\mathbf{x}^b$, as in the companion inversion studies, with emission budgets from PKU-CO₂-2007 (hereafter $\mathbf{x}^{\text{PKU}}$), we derive estimates of $\mathbf{H}\epsilon_b$ based on statistics on $\mathbf{H}^{\text{prac}}(\mathbf{x}^{\text{EDG-IER}} - \mathbf{x}^{\text{PKU}})$.

4. Results: estimates of the representation and aggregation errors

This Section characterizes the representation, aggregation and prior FFCO₂ errors, derived from the method described in Section 3. This characterization consists in providing their typical values (estimates of their standard deviations), investigating whether they bear temporal or spatial correlations while such correlations of the observation errors are traditionally ignored by atmospheric inversions (Rödenbeck et al., 2003; Chevallier et al., 2005; Peylin et al., 2013), and in investigating the validity of the assumptions that these observation errors have Gaussian and unbiased distributions. Section 4 focuses on the errors for a standard sampling strategy i.e. two-week mean afternoon sampling at 100 magl. Section 5 will explore the sensitivity of the results to the temporal sampling strategy (from one-day to one-month mean afternoon sampling) and give insights on the errors that would have been obtained if considering measurements sites with a different measurement height.

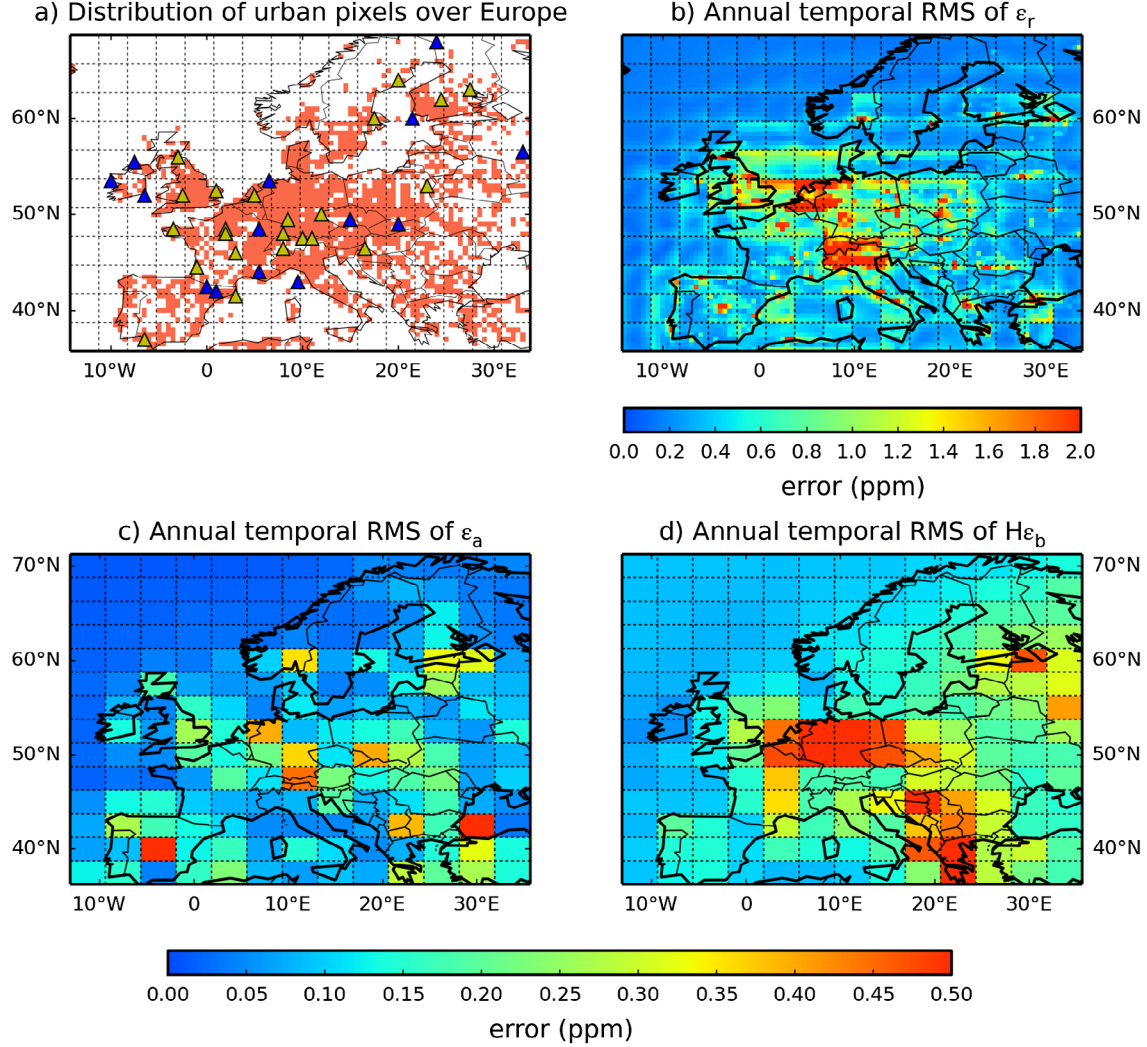


Fig. 3. Distribution of urban pixels (defined by population density, section 2.1.2) over Europe at 0.5° resolution (a) and maps of the RMS of the 1-year long time series of the representation errors ϵ_r (at 0.5° resolution) (b) ϵ_a (at $3.75^\circ \times 2.5^\circ$ resolution) (c) and the prior FFCO_2 errors $H\epsilon_b$ (at $3.75^\circ \times 2.5^\circ$ resolution) (d) for 2-week mean afternoon FFCO_2 gradients (from 100 magl sites to the JFJ reference site) (unit: ppm). In (a), the triangles give the location of the sites of a typical continental observation network similar to ICOS (ICOS, 2008; 2013); blue triangles means that the stations are in ‘rural’ pixels, while yellow triangles means the stations fall in ‘urban’ pixels.

Finally (in Section 6), we compare the typical values of the representation, aggregation and prior FFCO_2 errors to the model transport errors (derived in Section 3), to the measurements errors (given in the introduction), and to the typical signal of FFCO_2 modelled at the sites considered in this study.

4.1. Spatial distribution of the errors and spatial categorization

The root mean square (RMS) of the representation, aggregation and prior FFCO_2 errors for the one-year long time-series of two-week mean FFCO_2 gradients at each of the 0.5° to $3.5^\circ \times 2.75^\circ$ horizontal grid cells (depending on the error and thus on the scale at which it can be computed) are given in

Fig. 3b–d. In general, the RMS of the representation error for land grid cells ranges between 0.4 and 4.0 ppm across Europe, while the RMS of the aggregation and prior FFCO_2 errors are much smaller.

From Fig. 3b, the representation error at 100 magl shows higher values in the grid-cells classified as urbanized and large cities such as London, Paris, industrialized areas in Germany, etc. More generally, the spatial distribution of representation error shows a good consistency with the mask of the urban grid cells defined based on the population density and large point sources (Fig. 3a, see its definition in Section 2.1.2), indicating higher representation error in urban grid cells. We conclude that different statistics of the representation error need to be derived for the ‘urban’ 0.5° resolution land grid cells of the mask in

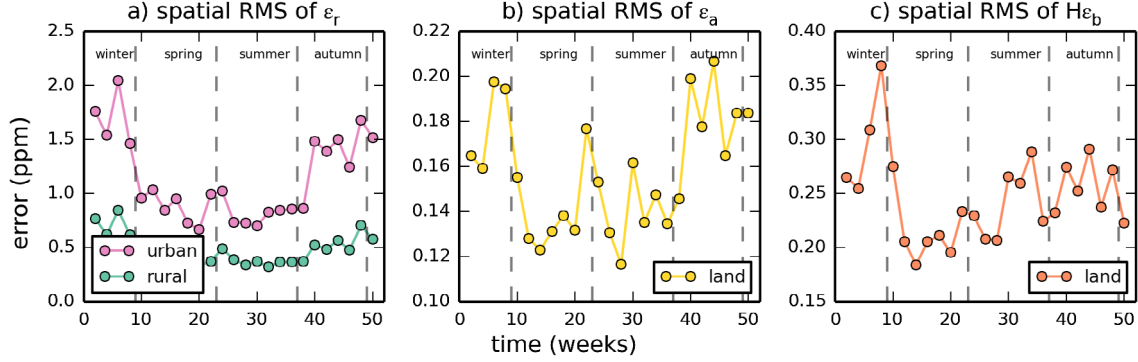


Fig. 4. Time series of the spatial RMS of the urban and rural representation errors (a), of the aggregation errors (b) and of the prior FFCO₂ errors (c) for 2-week mean afternoon FFCO₂ gradients (from 100 magl sites to the JFJ reference site) (unit: ppm).

Fig. 3a in one hand, and ‘rural’ 0.5° land grid cells of this mask on the other hand.

On the opposite, the aggregation error (Fig. 3c) being sampled at the coarse (~3° horizontally) grid resolution while urban area have generally smaller horizontal scales, their sub-sampling and thus the derivation of statistics for ‘urban’ and ‘rural’ grid cells (defined in Section 2.1.2) does not seem to be adapted.

The prior FFCO₂ errors (Fig. 3d) have a magnitude similar to the aggregation error. Their spatial distribution is strongly linked with that of the differences between the emission budgets for the control regions/months from the two inventories EDG-IER and PKU-CO₂-2007. As a consequence, it is not systematically consistent with that of the most urbanized areas in Europe. As an example, mean prior FFCO₂ errors reach 0.5 ppm in the Balkans but do not exceed 0.3 ppm in Northern Italy or in England. As for the aggregation error, the derivation of statistics of the prior FFCO₂ error for ‘urban’ and ‘rural’ grid cells is not adapted.

4.2. Temporal evolution of the errors and temporal categorization

Assuming that the statistics of the observation errors are independent of the location within the spatial categories defined above, we analyse the temporal variations of these statistics through that of the spatial RMS (over all corresponding 0.5° to 3.5° × 2.75° horizontal land grid cells) of the urban and rural representation errors, of the aggregation errors and of the prior FFCO₂ errors for the two-week mean afternoon FFCO₂ gradients to JFJ. The corresponding time series are given in Fig. 4.

There is a clear seasonal variation in the urban and rural representation errors (Fig. 4a). In spring and summer, when the vertical mixing of the lower atmosphere is stronger, the representation error drops to about 0.7 ppm for urban areas and 0.4 ppm for rural areas, while in winter, it can peak at about 2.0 ppm over urban grid cells and 0.8 ppm over rural grid cells, which

is about twice the values in summer. Levene’s test shows that the variances of the representation errors are distinct between the four seasons, except when comparing values for urban representation errors in spring vs. summer ($p < 0.05$ between rural values of all other pairs of seasons, or between urban values of any pair of season). Therefore, different statistics of the representation errors need to be derived for the different seasons i.e. spring (March to May), summer (June to August.), autumn (September to November) and winter (December, January and February). The prior FFCO₂ errors also show significant differences between the different seasons ($p < 0.05$ between all pairs of seasons). So the same seasonal categorization will also apply to them.

By contrast, there is only a small seasonal variation in the aggregation error (Fig. 4b). ϵ_a has lower values in spring and summer than in autumn and winter (Levene’s test, $p < 0.05$). Consequently, we use different statistics for ϵ_a in spring/summer and in autumn/winter.

4.3. Statistics of the errors

The distributions of most of the different categories of representation, aggregation and prior FFCO₂ errors defined by the Sections 4.1 and 4.2 are shown in Fig. 5. These categories used to build these distributions have a high number of values (at least 715 samples within one category, this minimum number applying to $H\epsilon_b$ in winter) so that the statistics from these distributions should be robust.

Two theoretical distributions are superimposed to each of the practical sampling of the errors in Fig. 5: a Gaussian distribution whose mean and standard deviation correspond to that of the practical sampling, and a Cauchy distribution whose location and scale parameters correspond to that of the practical sampling. For all categories of representation, aggregation and prior FFCO₂ errors, the Gaussian distribution is a poor approximation of the practical distribution, whereas the Cauchy distribution, with a relatively narrow peak, generally fits better with the practical distribution.

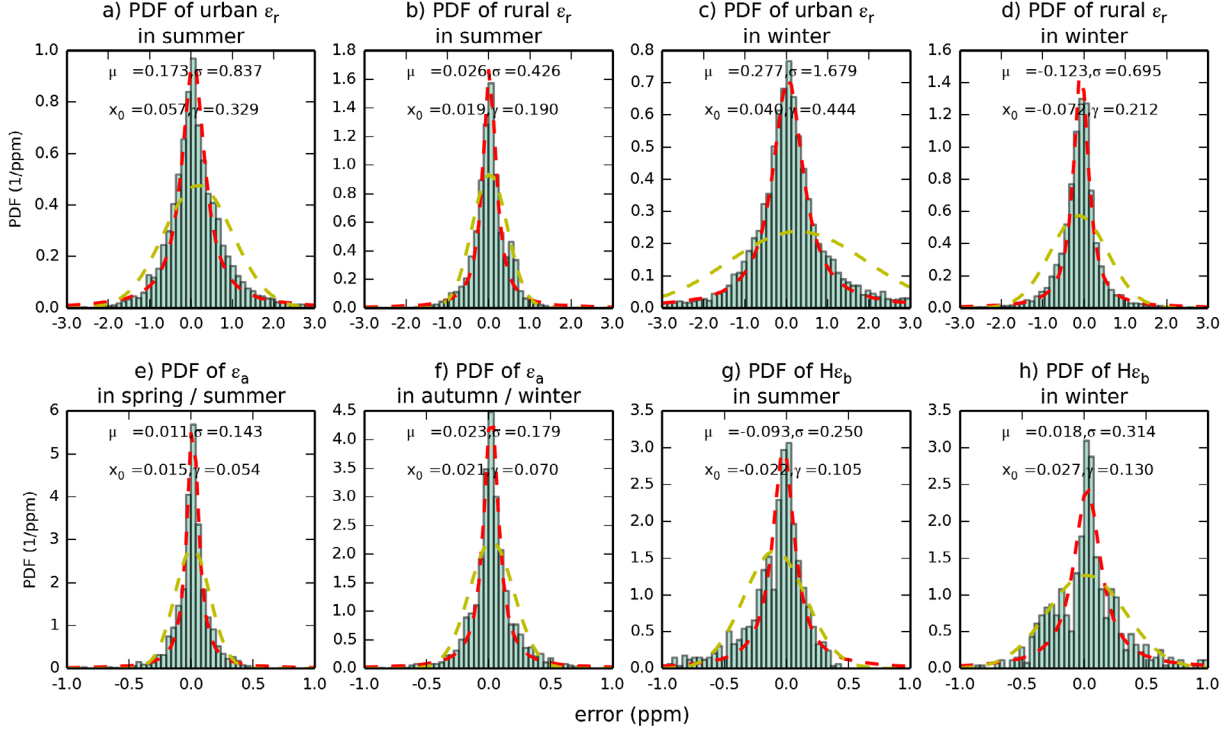


Fig. 5. Probability density functions (PDFs) of the representation, aggregation and prior FFCO₂ errors for 2-week mean afternoon gradients (from 100 magl sites to the JFJ reference site) for nearly all of the categories defined by sections 4.1 and 4.2 (only PDFs in spring and fall for urban and rural representation errors and for the prior FFCO₂ errors are not shown). The theoretical fit of these PDFs with Gaussian distributions (yellow dash lines) in terms of mean (μ) and standard deviations (σ), and the theoretical fit of these PDFs with Cauchy distributions (red dash lines) in terms of location parameter (x_0) and scale factor (γ) are also reported on the graphs.

Table 1. Standard deviations (in ppm) of the different categories of representation, aggregation and prior FFCO₂ errors for the 2-week mean afternoon FFCO₂ gradients and seasonal RMS (in ppm) of the FFCO₂ gradients between all potential rural or urban locations of 100 magl continental sites and JFJ and over all time periods during each season as simulated at 0.5° resolution when using CHIMERE and the EDG-IER inventory (i.e. our practical representation of the true gradients $\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t$).

	Spring		Summer		Autumn		Winter	
	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural
ϵ_r	0.89	0.40	0.84	0.43	1.33	0.51	1.68	0.70
ϵ_a	0.14				0.18			
$\mathbf{H}\epsilon_b$	0.22		0.25		0.27		0.31	
‘True’ gradient	3.4	2.2	3.2	2.1	4.6	2.8	4.8	2.9

All these distributions, which are based on all sites within a category, have near-zero means, which supports the assumption that observation errors are unbiased. Of note is that the potential consistency of the error in time at a given site or between neighbouring sites is not reflected in this distribution but is characterized through the analysis of the temporal and spatial correlations of these errors (see below). The standard deviations of these distributions are indicated in Table 1. The representation errors are much larger than the aggregation errors, and reach as high as 1.68 ppm for ‘urban’ grid cells in winter (Fig. 5c). The aggregation errors (Fig. 5e and f) are about one order of

magnitude smaller than representation errors. The prior FFCO₂ errors (Fig. 5g and h) are slightly larger than the aggregation errors but are still lower than the representation error by a factor of 3 to 6.

The temporal correlations of the different spatial categories of representation, aggregation and prior FFCO₂ errors for two-week mean afternoon gradients are illustrated in Fig. 6a using the temporal auto-correlations for errors of all the occurrences of the gradients from all the potential 100 magl sites to the JFJ reference site. The autocorrelation for a given time lag is derived from the ensemble across all times and sites of all couples

Table 2. The parameters optimized by the regressions of the temporal auto-correlations of the representation, aggregation and prior FFCO₂ errors for 2-week and 1-day mean afternoon FFCO₂ gradients (from all the potential 100 magl sites to JFJ), using e-folding functions $r(\Delta t) = a \times e^{-\Delta t/b} + (1 - a) \times e^{-\Delta t/c}$ and ignoring the different temporal categories.

	a	b	c
<i>2-week mean afternoon FFCO₂ gradients</i>			
Urban ε_r	0.31	16.6	>365
Rural ε_r	0.35	9.3	328
ε_a	0.16	11.2	>365
$H\varepsilon_b$	0.21	10.5	332
<i>1-day mean afternoon FFCO₂ gradients</i>			
Urban ε_r	0.69	0.87	>365
Rural ε_r	0.79	0.74	>365
ε_a	0.54	1.1	>365
$H\varepsilon_b$	0.49	2.1	325

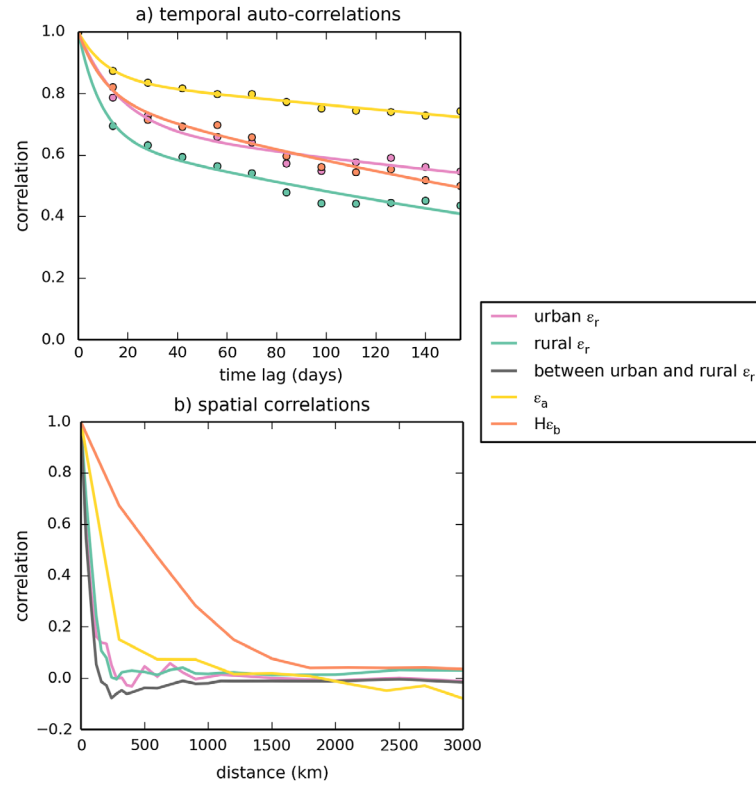


Fig. 6. Estimates of the correlations for the different spatial categories of representation errors (urban ε_r in purple and rural ε_r in green), aggregation errors (in yellow) and prior FFCO₂ errors (in red) for 2-week mean afternoon FFCO₂ gradients (from all the potential 100 magl sites to JFJ), ignoring the different temporal categories (i.e. mixing errors from all seasons and thus computing temporal auto-correlations between errors across different seasons or using a temporal sampling across different seasons to compute spatial correlations). (a) Temporal auto-correlations. Dots correspond to the estimates of the temporal auto-correlations. Lines correspond to regression curves with e-folding functions $r(\Delta t) = a \times e^{-\Delta t/b} + (1 - a) \times e^{-\Delta t/c}$ where Δt is the timelag (in days), and where a , b and c are the parameters optimized by the regression. (b) Spatial correlations. The spatial correlations between urban and rural representation errors (in brown) are given along with the correlations within the different categories of errors.

of errors which both apply to the same site and to two times separated by the given time-lag. Initial estimates accounting for the temporal categories (not shown) indicated that the temporal auto-correlations for different seasons are quite close to each other, so the temporal categorization is ignored here.

There are strong temporal auto-correlations in all types of errors for two-week mean afternoon gradients. The temporal auto-correlations of the representation errors, of the aggregation errors and of the prior FFCO₂ errors are above 0.4 even when the time lag exceeds 3 months. The estimates of the

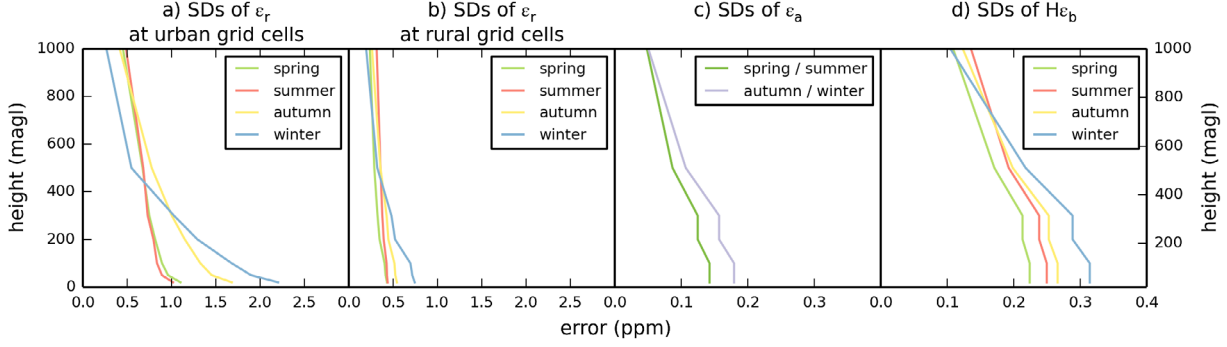


Fig. 7. Standard deviations (SDs) of all the occurrences of the representation, aggregation and prior FFCO₂ errors for specific categories of 2-week mean afternoon FFCO₂ gradients, as a function of the sampling height above ground (unit: ppm).

autocorrelations of the errors computed separately for each potential site used in the gradient FFCO₂ computations (not shown), are nearly null for time lag larger than 1 month for most of the potential sites. This indicates that the errors combine a sort of long term error component that is specific to each site (acting as a bias it does not show up in the site correlation), and a short-term error component whose typical correlation timescale is smaller than 1 month. A sum of two exponentially decaying functions $r(\Delta t) = a \times e^{-\Delta t/b} + (1 - a) \times e^{-\Delta t/c}$ is thus fitted to the estimate of the temporal auto-correlations (when using a sampling of the errors across all the times and sites) of each type of error, where Δt is the time lag (in days) and a , b , c are the parameters that are optimized by the regressions. The short timescale of correlation b arising from these regressions ranges from 9.3 days for the rural representation error to 16.6 days for the urban representation error. These values are close to the sampling integration time of 2 weeks. The long timescale of correlation c is larger than 1 year except for the rural representation error and prior FFCO₂ errors. The relative weight of the short term component of the errors (a) for the two-week mean afternoon FFCO₂ gradients is systematically below 40%. It is more important for the representation errors than for the aggregation and prior FFCO₂ errors.

The spatial correlations within the different categories of error on two-week mean afternoon gradients or between the urban and rural representation errors on two-week mean afternoon gradients are shown in Fig. 6b. Their estimates for a given distance are based on the ensemble across all times and sites of all couples of errors which both apply to the same time and to two sites separated by the given distance (using intervals for this distance of ± 20 km for the representation errors at the 0.5 horizontal resolution, and of ± 150 km for the aggregation errors and the prior FFCO₂ errors at the $3.75^\circ \times 2.5^\circ$ resolution). Again, initial estimates accounting for the temporal categories (not shown) indicated that the spatial correlations for the different seasons are quite close, so the temporal categorization is also ignored here.

The spatial correlations of the representation and aggregation errors drop very fast with increasing distance which is not the case for the prior FFCO₂ errors. The urban and rural rep-

resentation errors even have negative spatial correlations when the distance is within the range of 100–300 km. This is driven by the fact that the representation and aggregation errors when using the average concentration and emissions (respectively) within a given area (a grid cell or a region) are necessarily balanced and have thus opposite signs over areas smaller than that of this area. An exponentially decaying function $r(\Delta d) = e^{-\Delta d/a}$ is fitted to these estimates of spatial correlations, where Δd is the distance (in kilometers) and where a is the parameter that the regressions derive. The e-folding correlation lengths a are 75 and 89 km for the urban and rural representation errors respectively. The spatial correlations between urban and rural representation error has a similar e-folding correlation length of 55 km. In general, even though those correlations between rural and urban representation errors are smaller than that within a given category of representation error, they are very close to them, and the spatial correlations of the representation errors are weakly impacted by the categories which we have defined for these errors. The e-folding correlation length a is 171 km for the aggregation error. All of the correlation lengths derived for the representation and aggregation errors are thus smaller than the length of the LMDZ transport model grid cells. The spatial correlations of the representation and aggregation errors are thus negligible at this transport model resolution. However, the correlation length scale of the prior FFCO₂ errors is approximately 700 km, which is larger than transport model resolution.

5. Sensitivity of the results to the sampling heights and to the temporal sampling

5.1. Sensitivity to the sampling heights

All the results above are derived for two-week mean afternoon gradients between sites at 100 magl and JFJ. Here, we investigate the variations of the different categories of representation, aggregation and prior FFCO₂ errors for two-week mean afternoon gradients as a function of the sampling heights for all sites whose difference to JFJ correspond to these gradients,

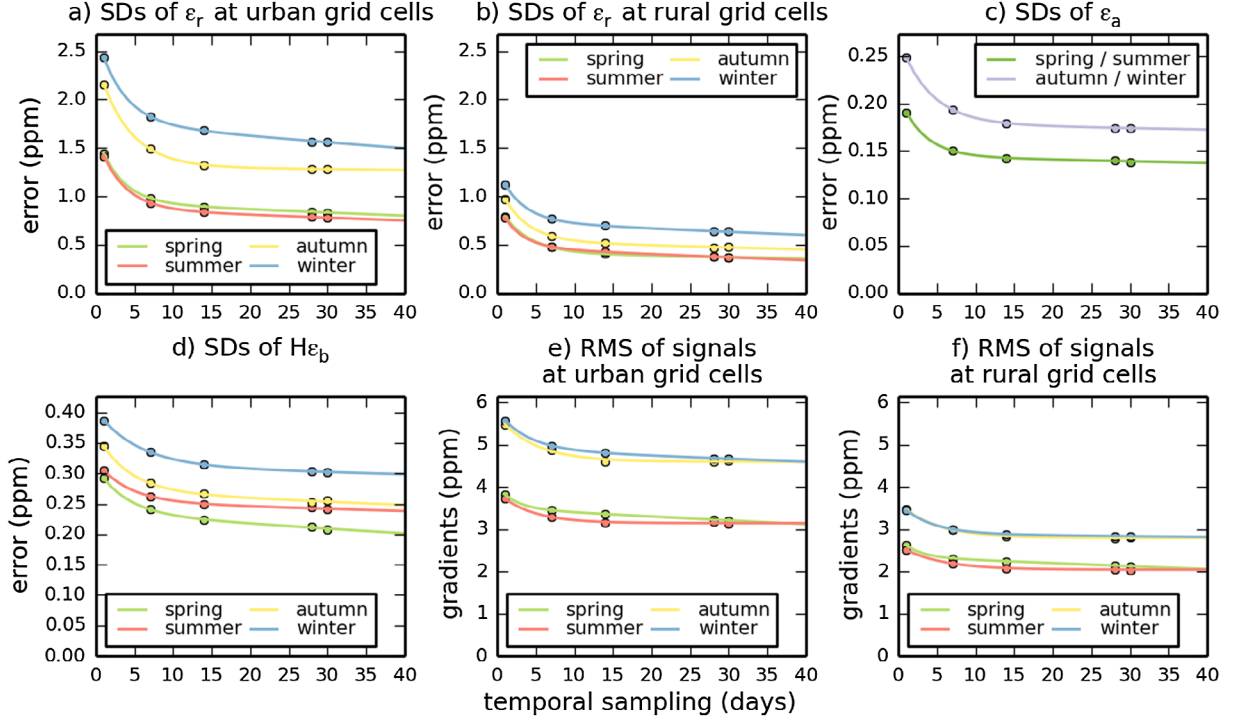


Fig. 8. Standard deviations (SDs) of all the occurrences of the representation, aggregation and prior FFCO₂ errors for 1-day to 1-month mean afternoon FFCO₂ gradients between 100 magl sites and the JFJ reference site (unit: ppm), and RMS of the simulated gradients at 0.5° resolution when using CHIMERE and the EDG-IER inventory (i.e. our practical representation of $\mathbf{H}_{\text{HR}} \rightarrow \mathbf{f}_{\text{HR}}^{\text{f}}(\mathbf{H}_{\text{HR}})$). The sampling durations are expressed in days. Dots correspond to the estimates of the standard deviations of the errors and of the quadratic mean of the simulated gradients. Lines correspond to regression curves with e-folding functions $\varepsilon(l) = \varepsilon(1) \times [a \times e^{-(l-1)^b} + (1-a) \times e^{-(l-1)^c}]$, where l is the duration (in days) of the mean afternoon sampling, $\varepsilon(1)$ is the standard deviations of the errors (or simulated gradients) for 1-day sampling, and where a , b and c are the parameters optimized by the regressions. Results for (a) urban ε_r ; (b) rural ε_r ; (c) ε_a ; (d) $\mathbf{H}\varepsilon_b$; (e) simulated gradients for urban grid cells; (f) simulated gradients for rural grid cells.

from near ground (20 magl) to top of planetary boundary layer (1000 magl, roughly). The height of the measurements at the JFJ reference site is not modified hereafter. The $\mathbf{H}_{\text{coloc}}^{\text{coloc}}$ and $\mathbf{H}_{\text{transpHR}}^{\text{CHIM}}$ operators (their selection of the LMDZ and CHIMERE vertical levels corresponding to the measurement locations; see Sections 2.1.3 and 3) are adapted for such a derivation of the vertical profiles of the errors. The sampling heights tested are 20, 50, 100, 200, 300, 500 and 1000 magl. These heights correspond to seven different vertical levels in CHIMERE (the bottom sampling height corresponds to the 1st CHIMERE level while the top sampling height corresponds to the 12nd to 15th CHIMERE level depending on the horizontal grid cells). The representation errors being computed at the spatial resolution of CHIMERE, we thus obtain different value of the different categories of representation errors per sampling height. However, due to its coarse vertical discretization, these sampling heights correspond to only the first five levels of the LMDZ model. This explains why only five values of the different type of categories of aggregation and prior FFCO₂ errors (that are computed at the spatial resolution of LMDZ) are derived for these seven sampling heights.

Fig. 7 shows the corresponding vertical variations of the standard deviations of all the occurrences of each category of representation, aggregation and prior FFCO₂ errors for two-week mean afternoon gradients. All categories of the errors decrease significantly with increasing sampling height. However, the different categories of errors have different vertical profiles. The standard deviation of the representation error for gradients between 300 magl sites in ‘urban’ grid cells and JFJ is equal to ~70% (in summer) or to ~50% (in winter) of the values for gradients between 20 magl sites in urban grid cells and JFJ. This seasonal variation in the decrease with height is likely due to the higher emissions but shallower depth of the vertical mixing in fall–winter than in spring–summer. The representation errors of FFCO₂ gradients between rural grid cells and JFJ have relatively smaller vertical variations. This is likely due to the fact that the atmospheric signature of the emitting urban grid cells and thus the corresponding representation errors have been highly diffused in the vertical during the transport from such urban grid cells to the rural grid cells. The standard deviations of the aggregation errors for gradients between 300 magl sites in ‘urban’ grid cells and JFJ are equal to about 75%

Table 3. The parameters optimized by the regressions of the standard deviations of the representation, aggregation and prior FFCO₂ errors as functions of the temporal sampling of the observations, for 1-day to 1-month mean afternoon FFCO₂ gradients between 100 magl sites and the JFJ reference site (unit: ppm), and RMS of the simulated gradients at 0.5° resolution when using CHIMERE and the EDG-IER inventory (i.e. our practical representation of $\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t$).

	Spring			Summer			Autumn			Winter		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
Urban ϵ_r	0.35	3.0	252	0.39	3.2	264	0.40	4.0	>365	0.28	3.3	244
Rural ϵ_r	0.48	3.5	279	0.39	2.5	116	0.45	3.3	228	0.34	3.0	186
ϵ_a	0.24	3.2	>365	0.24	3.2	>365	0.28	4.1	>365	0.28	4.1	>365
He_b	0.20	3.9	263	0.17	4.1	>365	0.22	4.1	>365	0.19	5.4	>365
Simulated gradients (urban)	0.09	2.4	350	0.16	4.5	>365	0.17	4.7	>365	0.13	3.8	>365
Simulated gradients (rural)	0.12	1.9	264	0.21	5.6	245	0.21	4.8	>365	0.19	3.8	>365

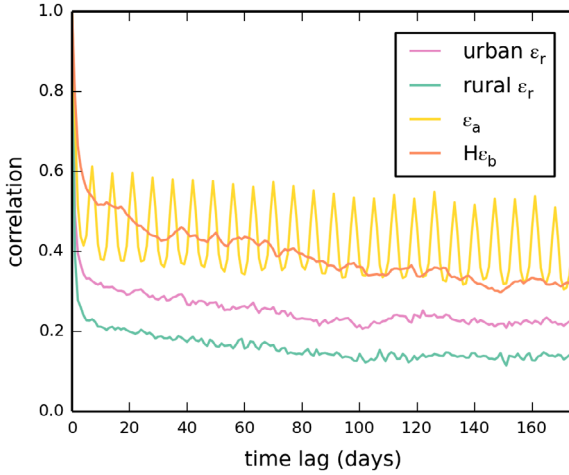


Fig. 9. Temporal auto-correlations of the representation (urban ϵ_r in purple and rural ϵ_r in green), aggregation (in yellow) and prior FFCO₂ errors (in red) for 1-day mean afternoon FFCO₂ gradients (from all the potential 100 magl sites to JFJ), ignoring the different temporal categories (i.e. mixing errors from all seasons and thus computing temporal auto-correlations between errors across different seasons).

of those for gradients between 20 magl sites and JFJ. Of note is also that the aggregation errors drop significantly as the heights exceed 400 magl unlike the representation errors. The vertical distribution of the prior FFCO₂ errors is similar to that of the aggregation errors. As a consequence, from the surface to 100–300 magl, the ratio of the prior FFCO₂ errors over the sum of all observation errors increases. However, mainly due to the fact that measurement errors do not decrease with altitude, this ratio decreases with increasing heights above 300 magl.

5.2. Sensitivity to the temporal sampling

All the results above about the representation and aggregation errors are derived for 2-week mean FFCO₂ afternoon gradients. Here, we investigate the representation and aggregation errors for one-day, one-week, two-week and one-month mean afternoon gradients between 100 magl sites and the JFJ reference site. The $\mathbf{H}_{\text{samp}}^{\text{coloc}}$ and $\mathbf{H}_{\text{transp}}^{\text{CHIM}}$ operators (their temporal averaging

of the afternoon FFCO₂ gradients; see Sections 2.1.3 and 3) are adapted accordingly.

The standard deviations of all the occurrences of FFCO₂ gradients within each category of the representation, aggregation and prior FFCO₂ errors, and of the gradients generated using our practical simulation of the actual gradients $\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t$ are shown in Fig. 8 as a function of the sampling integration time. All the errors and simulated gradients decrease significantly from one-day mean afternoon samplings to two-week mean afternoon samplings, while the decrease of the values from two-week to one-month mean afternoon samplings is relatively small. As analysed earlier when studying the temporal auto-correlations of the errors, this highlights the fact that these errors combine a long-term component specific to each site and a short term component. Fig. 8e and f show that this also applies to the simulation of the FFCO₂ gradients. As for the analysis of the temporal autocorrelations of the errors, a sum of two exponentially decaying functions $\epsilon(l) = \epsilon(1) \times [a \times e^{-(l-1)/b} + (1-a) \times e^{-(l-1)/c}]$ is thus fitted to the values of the errors and simulated gradients as functions of the sampling integration time, where l is the integration time (in days) of the mean afternoon sampling, $\epsilon(1)$ is the standard deviations of the errors (or simulated gradients) with one-day sampling, and where a , b and c are the parameters that are optimized by the regressions. The values obtained for the b range between 2 and 5 days, and those of c often exceed 1 year (Table 3), reflecting, as when fitting the temporal auto correlations, the synoptic timescales and a long-term site specific error respectively. While for the representation and aggregation errors, a (the weight of the short term component, Table 3) ranges between 24 and 48%, it is lower for the prior FFCO₂ errors (17 to 22% depending on the season) and for the simulated gradients (9 to 21% depending on the season).

Further analysis of the results when using different sampling integration time for the observations indicates that the spatial correlations of the errors do not evolve significantly as a function of this integration time. However, Fig. 9 shows that the temporal correlations of the representation errors associated with one-day mean afternoon gradients decrease a lot with in-

creasing timelag until the timelag reaches 1 week, which could not be characterized when analyzing errors for two-week mean gradients in Section 4.3. Fitting the temporal correlations of the errors for one-day mean samplings with the sum of two exponentially decaying functions (as when analyzing the temporal correlations of the errors for two-week mean gradients, Table 2) indicate that the timescale of correlation for the short term components of the representation and aggregation errors is about 1 day and that for the prior FFCO₂ error is 2.1 days. On the other hand, the timescale of the long-term component of the errors for one-day mean gradients still exceed 1 year except for the prior FFCO₂ error, as when analyzing errors for two-week mean gradients. However, the relative weight of the short-term component (a) are one to two times higher than that for two-week mean afternoon gradients when analyzing the 1-day mean gradients.

Of note is the fact that the aggregation error for one-day gradients due to the coarse resolution of the control vector has a component (in addition to the short term and long term components already characterized) that has a weekly cycle which is shown by the cycle of the temporal autocorrelations of the error at this frequency, and which reflects the quite artificial differences between the flat temporal profiles of the emissions in the PKU-CO₂-2007 inventory and the hourly variations of the emissions in the EDG-IER inventory. The existence, for all types of errors, of a third component at the daily scale in addition to a short term component at the synoptic scale and to the long term component, which could not be detected by the analysis of the autocorrelations at the two-week mean scale, could explain the differences between the results obtained when analysing the results at the one-day vs. two-week scale.

6. Discussion

6.1. *Validity of the assumption that the observation errors of FFCO₂ gradients have an unbiased and Gaussian distribution*

In Section 2.2, we justified our estimation of the representation, aggregation and prior FFCO₂ errors based on the assumption that the distributions of these errors are Gaussian and unbiased (as required for the application of the atmospheric inversion framework; (Lorenc, 1986)). Fig. 5 shows that the means of the representation, aggregation and prior FFCO₂ errors are much smaller than the standard deviations of these errors, indicating that the assumption that their distribution is unbiased is relevant. A Cauchy distribution generally shows a better fit with the practical sampling of the errors than the Gaussian distribution, but the Gaussian distribution is still a good approximation of the former. However, the dependence of the long-term component of the errors to the sites used for the computation of the gradients could be interpreted as a sort of local bias in the

representation and aggregation errors. While such a behaviour of the prior errors is generally well anticipated by inversion systems assuming long temporal autocorrelations of the prior uncertainties, these representation and aggregation errors can hardly be considered as a random noise for the observations as is done traditionally. However, accounting for both the short term and the long term error temporal correlations in the configuration of the inversion systems is feasible if using the type of regressions used in this study, and, if done, should enable a good characterization of these errors. And these analyses strengthen the posterior justification for our practical derivation of the representation and aggregation errors.

6.2. *The different temporal components of the errors and simulated gradients*

The analysis of the temporal autocorrelations of the errors for two-week to one-day mean errors and of the decrease in the standard deviation of the errors as a function of the temporal sampling integration time reveals three rather than two dominating component of the representation, aggregation and prior FFCO₂ errors. The first component, at the daily scale, may be driven by the day to day variations of the emissions which are high compared to the long-term variations in these emissions and which are highly uncertain in present inventories. This component can be highlighted when analyzing errors for one-day mean gradients only. The second component, at the synoptic scale, is likely related to the transport of the error from the emission areas to the sites through synoptic events which are a critical component of the transport in Europe (Parazoo et al., 2008; García et al., 2010). The third component is related to the slowly varying (at the ‘long-term’ seasonal to inter-annual scales), continuous and direct influence of the emission in the vicinity (in the same transport model grid cell) of the different sites, which thus significantly varies from site to site and depending whether it applies to urban or rural sites. The urban sites receive a strong signature of the emissions in their vicinity while the rural sites are affected by these emissions in their corresponding grid cells through a more indirect process of aggregation, and with a weaker signal. This explains why the relative weight of the synoptic component compared to the long-term one is generally smaller for gradients between urban sites and JFJ than for gradients between rural sites and JFJ.

The analysis of these different components reveals that the daily components dominate in the errors and simulated signal. This component is cancelled when sampling the observations at the one-week to the one-month scale, in which case the long-term component dominates the errors and simulated signal. The differences between the numbers (Table 2) characterizing the temporal scales of these components and their relative weight depending on the specific analysis lead in this study can be explained by the uncertainties associated with these statistical analyses, by the difficulty to make a regression with a sum of

three exponentially decaying functions when analyzing the errors and signal for one-day mean gradients (which could, in principle, help reconcile it with the results when analyzing the errors and the signal for two-week mean gradients), and by the difference between the analysis of the temporal correlations of errors for a given temporal sampling of the observations and that of the amplitude of the error for different temporal sampling of the observations. However, their general consistency gives insights into the typical correlation length scales to be used for accounting for such correlations when conducting atmospheric inversions, which will be critical given the amplitude of these correlations.

6.3. Comparison of the different errors: potential for filtering the signature of the uncertainties in the average emissions over large region?

The analysis of the standard deviation and of the temporal scales of autocorrelations of the errors indicate that the representation error and the transport errors are the largest observation errors for 100 magl afternoon observations for any temporal scale of sampling when using the modelling framework of our study. They are larger than the measurement error (about 1 ppm) which is the third dominant type of observation errors in our modelling framework. The aggregation errors have a relatively small standard deviation compared to these errors (see their smaller standard deviation in Table 1 and their shorter temporal scale of auto-correlation in Table 2).

In total, the weight of the observation errors can reach up to 50% of the typical amplitude of the simulated 1-week to 2-week mean afternoon FFCO₂ gradients, either for urban or rural sites, or for any season and temporal sampling. It can even reach up to 90% of this signal for one-day mean afternoon FFCO₂ gradients. This questions the precision of the signature of the large-scale budget of FFCO₂ emissions that could be filtered from the observations. Furthermore, the actual signal that the inversion aims at filtering from the assimilated prior-model data misfits in order to correct for the prior knowledge on the emissions at the control/one-month scale is that of the prior FFCO₂ errors whose amplitude is generally smaller than that of the representation, transport and measurements errors. From our analysis, the signals at a given site from uncertainties in the distribution of the local emissions/concentrations (characterized by the representation error) exceed those of the uncertainties in the emissions at the regional scale (characterized by the prior FFCO₂ error) and the signal from the uncertainties in the distribution of the emissions at the sub-regional scale (characterized by the aggregation error) at any temporal scale.

The temporal autocorrelations of the prior FFCO₂ error have a structure that is similar to that of the representation errors. Its long-term component has a shorter temporal scale than that of the representation error but this can hardly be viewed as a basis for a practical separation of the prior FFCO₂ and

representation errors given that both these temporal scales exceed 300 days. The temporal length scale of the short term components for the prior FFCO₂ and representation errors are different. However, the combination of representation, aggregation, measurement and transport errors which all have their own temporal scale of correlations will likely make it difficult to exploit the structures of the short term variations to filter the prior FFCO₂ errors.

The most promising result for the potential filtering of the prior FFCO₂ errors lie in the analysis of the spatial correlations of the errors. The prior FFCO₂ errors are connected to uncertainties in emissions at large spatial scales which are not necessarily compensated between neighbouring control regions, while representation and aggregation errors should be compensated at the resolutions of the transport model to the control vector (and thus cancelled through atmospheric mixing faster than the prior FFCO₂ errors). At the same time, transport errors and measurement errors should not be correlated in space. In consequence, the spatial correlations of the prior FFCO₂ errors are larger than that of the observation errors. This could be exploited by the inversion to filter it if it can rely on a spatially dense network of measurement sites.

An analysis of the correlations between different types of errors and the transport conditions (e.g. wind direction and speed) could bring additional insights on the capability for isolating the prior FFCO₂ errors from the observations. However, such an analysis could hardly be based on the framework of this study for which the different types of error (in particular the representation and prior FFCO₂ errors) are derived using different modelling frameworks. A full assessment of this capability requires atmospheric inversions using the characterization of the errors from this study. However, this study already give insights into the challenges underlying the monitoring the emissions at large scale.

Of note is that the quite simple analysis of the sensitivity of the ratio between the prior FFCO₂ errors and the observation errors supports the monitoring of FFCO₂ at heights ranging between 100 and 300 magl when targeting the large-scale budgets of emissions. It is also the traditional heights for ICOS network (Kadyrov et al., 2015).

The results from this study do not strongly encourage to sample FFCO₂ observations at high temporal resolution when targeting the regional/one-month scale budgets of emissions with atmospheric inversion built on a coarse-grid transport model. We have shown the relative weight of the short term components of the observation errors are larger than that of the prior FFCO₂ errors and, as said previously, making it difficult to exploit the difference between the temporal correlation of these short term components to filter the prior FFCO₂ errors. However, again, atmospheric inversion experiments would be required to check how much the correlations of the errors with the transport could be exploited, even at high temporal resolution. Still, having one-day mean sampling would dramatically decrease the weight of the measurement errors on longer time

scales that can be considered as a random noise on each individual measurement of FFCO₂.

6.4. *Increasing the spatial resolution of the atmospheric transport model and of the control variables to increase the potential of the atmospheric inversion?*

As explained in Section 2.3 and in the Appendix, in theory, solving for the emissions at high resolution or solving for emissions at large scale but perfectly accounting for the aggregation error would yield similar estimates of the emissions at large scale. In practice, as illustrated by Fig. 6, the complex structure of the correlations in the aggregation error may prevent such a perfect account for this error in the inversion set-up. Therefore, in principle, it is better to solve for the emissions at the highest resolution as possible.

On the other hand, the representation and transport errors are also highly dependent on the transport model resolution. Since the representation errors are the largest component of the observation error for our modelling framework, increasing the transport model resolution should be viewed as the most critical mean for improving (if needed) the results from atmospheric inversion for the large-scale monitoring of the emissions. Cancelling the representation errors would dramatically increase the ratio between the prior FFCO₂ errors and the observation errors. However, this would require using a regional inverse modelling framework focusing on a specific area (such as Europe, the US or China) since the horizontal resolution of the atmospheric transport model used for global inversion hardly exceed 3° resolution.

7. Conclusion

This paper analyses the critical sources of errors that influence the estimate of FFCO₂ emissions at sub-continental/monthly scale from atmospheric inversion based on continental networks of daily to monthly mean afternoon atmospheric FFCO₂ observations. We provide a theoretical derivation of the representation and aggregation errors affecting daily to monthly mean afternoon FFCO₂ gradients between possible measurement sites and a background station. This theoretical derivation is adapted to the practical estimation of these errors in Europe for our specific inverse modelling framework that is based on a global coarse-resolution transport model. Our analysis focuses on the derivation of the standard deviations, temporal and spatial correlations of the representation and aggregation errors, the standard deviation of the transport model and measurement errors, along with the standard deviation of the atmospheric signature of the prior uncertainty in the regional/1-month budgets of the emissions. These statistical parameters will be primarily used to set up a realistic configuration of the observation errors in inversion experiments described in future papers. In particular, the modelling of the spatial and temporal correlations using

one or two exponentially decaying functions and the optimized parameters from the regressions in this study will be used to model the observation error correlations in these experiments.

Our comparison between these statistical parameters for the different types of errors also aims at assessing the ability to filter the signature of the prior uncertainty in the large-scale budgets of the emissions from the total observation errors when assimilating one-day to one-month mean afternoon FFCO₂ gradients (which underlies the potential for estimating the emissions at large scale). It highlights that the representation, transport and measurement errors dominate the observations errors, while the weight of aggregation error is relatively small. In total, observation errors can reach up to 50% (90%) of the typical 2-week (1-day) mean FFCO₂ gradients, and are larger than the signature of the prior uncertainty in the large-scale budgets of the emissions. Moderating the representation and transport errors by using a regional transport models at higher resolution could thus be a requirement for the monitoring of FFCO₂, even when targeting their large-scale budgets. The analysis also highlight the fact the critical weight of the temporal correlation of the representation and aggregation errors, and in particular that they have a long-term component, make it difficult to separate these errors from the signature of the prior uncertainty in the emissions at large scale when assimilating the one-day to one-month mean afternoon FFCO₂ gradients. Filtering of the signature of the prior uncertainties could potentially rely on its spatial correlations scales which are significantly longer than that of the observation errors. This would require a network dense enough to capture the spatial coherence of this signature (at scales shorter than ~700 km), which could represent a larger number of sites than that of the present ICOS network. Finally, from this study, we do not recommend sampling FFCO₂ data at high (one-day to one-week) rather than low (two-week to one-month) temporal resolution for the global atmospheric inversion based on coarse-grid transport model, since it demonstrates the difficulties associated with filtering the signal from prior uncertainties at such temporal scales.

More generally, while the statistics of representation and aggregation errors derived in this study primarily relate to the specific atmospheric inversion framework we use in this study, this study brings insights regarding these errors for a wide range of atmospheric applications, and more specifically to that dedicated to the inversion of the FFCO₂ emissions. The practical derivation of their statistics can be easily generalized based on our theoretical framework and used for other studies. The structure and typical amplitude of the representation error derived for the transport of uncertainties in the fossil fuel emissions in Europe at ~3° resolution with LMDZ should be similar for other transport models with similar spatial resolution. The aggregation errors are shown to characterize the weight of uncertainties in the spatial and temporal resolution of the emissions even when solving for them at the transport model resolution. And the general conclusions raised above regarding the potential for

filtering the signature of the uncertainty in the large-scale budget of the emissions using ‘remote’ measurement stations should be a general outreach of this study. While the correlations of the observation errors are generally ignored in atmospheric inversion, this study demonstrates how critical it should be to account for them. By conducting inversion experiments accounting for these correlations, the companion papers of this study should now indicate how much the potential for filtering the signature of the uncertainty in the large-scale budget of the emissions based on its specific spatial structure can be effectively exploited by atmospheric inversion to solve for the regional/monthly scale budgets of the emissions.

Acknowledgement

The authors acknowledge the support of the French Commissariat à l’énergie atomique et aux énergies alternatives (CEA). This study is co-funded by the European Commission under the EU Seventh Research Framework Programme (grant agreement No. 283080, geocarbon project). G. Broquet and F. Vogel acknowledge funding from the industrial chair BridGES (supported by the Université de Versailles Saint-Quentin-en-Yvelines, the Commissariat à l’Energie Atomique et aux Energies Renouvelables, the Centre National de la Recherche Scientifique, Thales Alenia Space and Veolia). We are also grateful to Ingeborg Levin for the useful discussions on this topic. We also would like to thank the partners of the ICOS infrastructure for details of radiocarbon samplings and FFCO₂ monitoring.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the European Commission under the EU Seventh Research Framework Programme [grant agreement number 283080], geocarbon project).

References

- Andres, R. J., Marland, G., Fung, I. and Matthews, E. 1996. A 1°×1° distribution of carbon dioxide emissions from fossil fuel consumption and cement manufacture, 1950–1990. *Global Biogeochem. Cycles* **10**, 419–429.
- Andres, R. J., Boden, T. A., Bréon, F. M., Ciais, P., Davis, S. and co-authors. 2012. A synthesis of carbon dioxide emissions from fossil-fuel combustion. *Biogeosciences* **9**, 1845–1871.
- Andres, R. J., Boden, T. A. and Higon, D. 2014. A new evaluation of the uncertainty associated with CDIAC estimates of fossil fuel carbon dioxide emission. *Tellus B* **66**, 23616.
- Ballantyne, A. P., Andres, R., Houghton, R., Stocker, B. D., Wanninkhof, R. and co-authors. 2015. Audit of the global carbon budget: estimate errors and their impact on uptake uncertainty. *Biogeosciences* **12**, 2565–2584.
- Basu, S., Miller, J. B. and Lehman, S. 2016. Separation of biospheric and fossil fuel fluxes of CO₂ by atmospheric inversion of CO₂ and 14CO₂ measurements: observation system simulations. *Atmos. Chem. Phys. Discuss.* **2016**, 1–34.
- Berrisford, P., Dee, D., Fielding, K., Fuentes, M., Kallberg, P. and co-authors. 2009. *The ERA-Interim Archive*. ERA Report Series, Reading, pp. 1–16.
- Bocquet, M., Wu, L. and Chevallier, F. 2011. Bayesian design of control space for optimal assimilation of observations. Part I: consistent multiscale formalism. *Q. J. Roy. Meteor. Soc.* **137**, 1340–1356.
- Bousquet, P., Peylin, P., Ciais, P., Le Quééré, C., Friedlingstein, P. and co-authors. 2000. Regional changes in carbon dioxide fluxes of land and oceans since 1980. *Science* **290**, 1342–1346.
- Bozhinova, D., van der Molen, M. K., Krol, M. C., van der Laan, S., Meijer, H. A. J. and co-authors. 2013. Simulating the integrated Δ14CO₂ signature from anthropogenic emissions over Western Europe. *Atmos. Chem. Phys. Discuss.* **13**, 30611–30652.
- Bréon, F. M., Broquet, G., Puygrenier, V., Chevallier, F., Xueref-Remy, I. and co-authors. 2015. An attempt at estimating Paris area CO₂ emissions from atmospheric concentration measurements. *Atmos. Chem. Phys.* **15**, 1707–1724.
- Brioude, J., Petron, G., Frost, G. J., Ahmadov, R., Angevine, W. M. and co-authors. 2012. A new inversion method to calculate emission inventories without a prior at mesoscale: application to the anthropogenic CO₂ emission from Houston, Texas. *J. Geophys. Res.* **117**, D05312.
- Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I. and co-authors. 2011. A European summertime CO₂ biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements. *J. Geophys. Res.* **116**, D23303.
- Chen, H., Winderlich, J., Gerbig, C., Hoefer, A., Rella, C. W. and co-authors. 2010. High-accuracy continuous airborne measurements of greenhouse gases (CO₂ and CH₄) using the cavity ring-down spectroscopy (CRDS) technique. *Atmos. Meas. Tech.* **3**, 375–386.
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P. and co-authors. 2005. Inferring CO₂ sources and sinks from satellite observations: method and application to TOVS data. *J. Geophys. Res.* **110**, D24309.
- Chevallier, F., Ciais, P., Conway, T. J., Aalto, T., Anderson, B. E. and co-authors. 2010. CO₂ surface fluxes at grid point scale estimated from a global 21 year reanalysis of atmospheric measurements. *J. Geophys. Res.* **115**, D21307.
- Chevallier, F. and O’Dell, C. W. 2013. Error statistics of Bayesian CO₂ flux inversion schemes as seen from GOSAT. *Geophys. Res. Lett.* **40**, 1252–1256.
- Ciais, P., Paris, J. D., Marland, G., Peylin, P., Piao, S. L. and co-authors. 2010. The European carbon balance. Part 1: fossil fuel emissions. *Global Change Biol.* **16**, 1395–1408.
- Engelen, R. J. 2002. On error estimation in atmospheric CO₂ inversions. *J. Geophys. Res.* **107**, 4635.
- Enting, I. G., Trudinger, C. M., Francey, R. J. and Granek, H. 1993. Synthesis inversion of atmospheric CO₂ using the GISS tracer transport model. *CSIRO Aust. Div. Atmos. Res. Tech. Pap.* **29**, 1–44.
- Gamitzner, U., Karstens, U., Kromer, B., Neubert, R. E. M., Meijer, H. A. J. and co-authors. 2006. Carbon monoxide: a quantitative tracer for fossil fuel CO₂? *J. Geophys. Res.* **111**, D22302.

- García, M. Á., Sánchez, M. L. and Pérez, I. A. **2010**. Synoptic weather patterns associated with carbon dioxide levels in Northern Spain. *Sci. Total Environ.* **408**, 3411–3417.
- Geels, C., Gloor, M., Ciais, P., Bousquet, P., Peylin, P. and co-authors. **2007**. Comparing atmospheric transport models for future regional inversions over Europe – Part 1: mapping the atmospheric CO₂ signals. *Atmos. Chem. Phys.* **7**, 3461–3479.
- Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E. and co-authors. **2003**. Toward constraining regional-scale fluxes of CO₂ with atmospheric observations over a continent: 1. Observed spatial variability from airborne platforms. *J. Geophys. Res.: Atmos.* **108**, n/a–n/a.
- Graven, H. D. and Gruber, N. **2011**. Continental-scale enrichment of atmospheric ¹⁴CO₂ from the nuclear power industry: potential impact on the estimation of fossil fuel-derived CO₂. *Atmos. Chem. Phys.* **11**, 12339–12349.
- Gregg, J. S., Andres, R. J. and Marland, G. **2008**. China: emissions pattern of the world leader in CO₂ emissions from fossil fuel consumption and cement production. *Geophys. Res. Lett.* **35**, L08806.
- Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D. and co-authors. **2002**. Towards robust regional estimates of CO₂ sources and sinks using atmospheric transport models. *Nature* **415**, 626–630.
- Gurney, K. R., Mendoza, D. L., Zhou, Y., Fischer, M. L., Miller, C. C. and co-authors. **2009**. High resolution fossil fuel combustion CO₂ emission fluxes for the United States. *Environ. Sci. Technol.* **43**, 5535–5541.
- Hourdin, F., Musat, I., Bony, S., Braconnot, P., Codron, F. and co-authors. **2006**. The LMDZ4 general circulation model: climate performance and sensitivity to parametrized physics with emphasis on tropical convection. *Clim. Dynam.* **27**, 787–813.
- Hsueh, D. Y., Krakauer, N. Y., Randerson, J. T., Xu, X., Trumbore, S. E. and co-authors. **2007**. Regional patterns of radiocarbon and fossil fuel-derived CO₂ in surface air across North America. *Geophys. Res. Lett.* **34**, L02816.
- IPCC. **2006**. *2006 IPCC Guidelines for National Greenhouse Gas Inventories* Institute for Global Environmental Strategies, Hayama.
- Kadyrov, N., Broquet, G., Chevallier, F., Rivier, L., Gerbig, C. and co-authors. **2015**. On the potential of ICOS atmospheric CO₂ measurement network for the estimation of the biogenic CO₂ budget of Europe. *Atmos. Chem. Phys. Discuss.* **15**, 14221–14273.
- Kaminski, T., Rayner, P. J., Heimann, M. and Enting, I. G. **2001**. On aggregation errors in atmospheric transport inversion. *J. Geophys. Res.* **106**, 4703–4715.
- Lauvaux, T., Pannekoucke, O., Sarrat, C., Chevallier, F., Ciais, P. and co-authors. **2009**. Structure of the transport uncertainty in mesoscale inversions of CO₂ sources and sinks using ensemble model simulations. *Biogeosciences* **6**, 1089–1102.
- Lauvaux, T., Ullrich, M., Sarrat, C., Chevallier, F., Bousquet, P. and co-authors. **2008**. Mesoscale inversion: first results from the CERES campaign with synthetic data. *Atmos. Chem. Phys.* **8**, 3459–3471.
- Law, R., Peters, W., Rödenbeck, C., Aulagnier, C., Baker, I. and co-authors. **2008**. TransCom model simulations of hourly atmospheric CO₂: experimental overview and diurnal cycle results for 2002. *Global Biogeochem. Cycles* **22**, n/a–n/a.
- Levin, I. and Karstens, U. **2008**. Quantifying Fossil Fuel CO₂ over Europe. In: *The Continental-Scale Greenhouse Gas Balance of Europe* (eds. A. J. Dolman, A. Freibauer, and R. Valentini), Springer, New York, pp. 53–72.
- Levin, I., Kromer, B., Schmidt, M. and Sartorius, H. **2003**. A novel approach for independent budgeting of fossil fuel CO₂ over Europe by ¹⁴CO₂ observations. *Geophys. Res. Lett.* **30**, 2194.
- Levin, I., Hammer, S., Kromer, B. and Meinhardt, F. **2008**. Radiocarbon observations in atmospheric CO₂: determining fossil fuel CO₂ over Europe using Jungfraujoch observations as background. *Sci. Total Environ.* **391**, 211–216.
- Levin, I., Hammer, S., Eichelmann, E. and Vogel, F. R. **2011**. Verification of greenhouse gas emission reductions: the prospect of atmospheric monitoring in polluted areas. *Philos. Trans. Series A, Math. Phys. Eng. Sci.* **369**, 1906–1924.
- Levin, I. M. K. O. W. W. **1980**. The effect of anthropogenic CO₂ and ¹⁴C sources on the distribution of ¹⁴C in the atmosphere. *Radiocarbon* **22**, 379–391.
- Lin, J. C. and Gerbig, C. **2005**. Accounting for the effect of transport errors on tracer inversions. *Geophys. Res. Lett.* **32**, L01802.
- Lin, J. C., Gerbig, C., Wofsy, S. C., Daube, B. C., Matross, D. M. and co-authors. **2006**. What have we learned from intensive atmospheric sampling field programmes of CO₂? *Tellus B* **58**, 331–343.
- Liu, Z., Guan, D., Wei, W., Davis, S. J., Ciais, P. and co-authors. **2015**. Reduced carbon emission estimates from fossil fuel combustion and cement production in China. *Nature* **524**, 335–338.
- Lorenc, A. C. **1986**. Analysis methods for numerical weather prediction. *Q. J. Roy. Meteor. Soc.* **112**, 1177–1194.
- Macknick, J. **2009**. *Energy and Carbon Dioxide Emission Data Uncertainties*. IIASA Interim Report IR-09-32. International Institute for Applied Systems Analysis, Laxenburg.
- Marland, G. **2008**. Uncertainties in accounting for CO₂ from fossil fuels. *J. Ind. Ecol.* **12**, 136–139.
- McKain, K., Wofsy, S. C., Nehrkorn, T., Eluszkiewicz, J., Ehleringer, J. R. and co-authors. **2012**. Assessment of ground-based atmospheric observations for verification of greenhouse gas emissions from an urban region. In: *Proc. Nat. Acad. Sci. USA* **109**, 8423–8428.
- Miller, S. M., Hayek, M. N., Andrews, A. E., Fung, I. and Liu, J. **2015**. Biases in atmospheric CO₂ estimates from correlated meteorology modeling errors. *Atmos. Chem. Phys.* **15**, 2903–2914.
- Naegler, T. and Levin, I. **2006**. Closing the global radiocarbon budget 1945–2005. *J. Geophys. Res.* **111**, D12311.
- Newman, S., Jeong, S., Fischer, M. L., Xu, X., Haman, C. L. and co-authors. **2013**. Diurnal tracking of anthropogenic CO₂ emissions in the Los Angeles basin megacity during spring 2010. *Atmos. Chem. Phys.* **13**, 4359–4372.
- Niwa, Y., Machida, T., Sawa, Y., Matsueda, H., Schuck, T. J. and co-authors. **2012**. Imposing strong constraints on tropical terrestrial CO₂ fluxes using passenger aircraft based measurements. *J. Geophys. Res.* **117**, D11303.
- Oak Ridge National Laboratory (ORNL): LandScan Global Population 2007 Database. **2015**. Online at: <http://www.ornl.gov/sci/landscan/>.
- Oda, T. and Maksyutov, S. **2011**. A very high-resolution (1 km × 1 km) global fossil fuel CO₂ emission inventory derived using a point source database and satellite observations of nighttime lights. *Atmos. Chem. Phys.* **11**, 543–556.
- Olivier, J. G. J., Van Aardenne, J. A., Dentener, F. J., Pagliari, V., Ganzeveld, L. N. and co-authors. **2005**. Recent trends in global

- greenhouse gas emissions: regional trends 1970–2000 and spatial distribution of key sources in 2000. *Environ. Sci.* **2**, 81–99.
- Pacala, S. W., Breidenich, C., Brewer, P. G., Fung, I. Y., Gunson, M. R. and co-authors. 2010. Verifying greenhouse gas emissions: methods to support international climate agreements. National Academies Press, Washington, DC.
- Parazoo, N. C., Denning, A. S., Kawa, S. R., Corbin, K. D., Lokupitiya, R. S. and co-authors. 2008. Mechanisms for synoptic variations of atmospheric CO₂ in North America, South America and Europe. *Atmos. Chem. Phys.* **8**, 7239–7254.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J. and co-authors. 2007. An atmospheric perspective on North American carbon dioxide exchange: carbon tracker. *Proc. Nat. Acad. Sci. USA* **104**, 18925–18930.
- Peylin, P., Rayner, P. J., Bousquet, P., Carouge, C., Hourdin, F. and co-authors. 2005. Daily CO₂ flux estimates over Europe from continuous atmospheric measurements: 1, inverse methodology. *Atmos. Chem. Phys.* **5**, 3173–3186.
- Peylin, P., Houweling, S., Krol, M. C., Karstens, U., Rödenbeck, C. and co-authors. 2011. Importance of fossil fuel emission uncertainties over Europe for CO₂ modeling: model intercomparison. *Atmos. Chem. Phys.* **11**, 6607–6622.
- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R. and co-authors. 2013. Global atmospheric carbon budget: results from an ensemble of atmospheric CO₂ inversions. *Biogeosci. Discuss.* **10**, 5301–5360.
- Pillai, D., Gerbig, C., Ahmadov, R., Rödenbeck, C., Kretschmer, R. and co-authors. 2011. High-resolution simulations of atmospheric CO₂ over complex terrain – representing the Ochsenkopf mountain tall tower. *Atmos. Chem. Phys.* **11**, 7445–7464.
- Rödenbeck, C., Houweling, S., Gloor, M. and Heimann, M. 2003. CO₂ flux history 1982–2001 inferred from atmospheric data using a global inversion of atmospheric transport. *Atmos. Chem. Phys.* **3**, 1919–1964.
- Randerson, J. T., Enting, I. G., Schuur, E. A. G., Caldeira, K. and Fung, I. Y. 2002. Seasonal and latitudinal variability of troposphere Δ14CO₂: Post bomb contributions from fossil fuels, oceans, the stratosphere, and the terrestrial biosphere. *Global Biogeochem. Cycles* **16**, 59–51–59–19.
- Ray, J., Yadav, V., Michalak, A. M., van Bloemen Waanders, B. and McKenna, S. A. 2014. A multiresolution spatial parameterization for the estimation of fossil-fuel carbon dioxide emissions via atmospheric inversions. *Geosci. Model Dev.* **7**, 1901–1918.
- Rivier, L., Ciais, P., Hauglustaine, D. A., Bakwin, P., Bousquet, P. and co-authors. 2006. Evaluation of SF₆, C₂Cl₄, and CO to approximate fossil fuel CO₂ in the Northern Hemisphere using a chemistry transport model. *J. Geophys. Res.* **111**, D16311.
- Schmidt, H., Derognat, C., Vautard, R. and Beekmann, M. 2001. A comparison of simulated and observed ozone mixing ratios for the summer of 1998 in Western Europe. *Atmos. Environ.* **35**, 6277–6297.
- Shiga, Y. P., Michalak, A. M., Gourdji, S. M., Mueller, K. L. and Yadav, V. 2014. Detecting fossil fuel emissions patterns from subcontinental regions using North American in situ CO₂ measurements. *Geophys. Res. Lett.* **41**, 4381–4388.
- Tarantola, A. 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*. Siam, Philadelphia.
- Thoning, K. W., Tans, P. P. and Komhyr, W. D. 1989. Atmospheric carbon dioxide at Mauna Loa Observatory: 2. Analysis of the NOAA GMCC data, 1974–1985. *J. Geophys. Res.* **94**, 8549–8565.
- Turnbull, J. C., Keller, E. D., Norris, M. W. and Wiltshire, R. M. 2016. Independent evaluation of point source fossil fuel CO₂ emissions to better than 10%. *Proc. Nat. Acad. Sci.* **113**, 10287–10291.
- Turnbull, J., Miller, J., Lehman, S., Hurst, D., Peters, W. and co-authors. 2009. Spatial distribution of Δ14CO₂ across Eurasia: measurements from the TROICA-8 expedition. *Atmos. Chem. Phys.* **9**, 175–187.
- Turnbull, J. C., Tans, P. P., Lehman, S. J., Baker, D., Conway, T. J. and co-authors. 2011. Atmospheric observations of carbon monoxide and fossil fuel CO₂ emissions from East Asia. *J. Geophys. Res.: Atmos.* **116**, D24306.
- Turnbull, J. C., Keller, E. D., Baisden, T., Brailsford, G., Bromley, T. and co-authors. 2014. Atmospheric measurement of point source fossil CO₂ emissions. *Atmos. Chem. Phys.* **14**, 5001–5014.
- Ummel, K. 2012. CARMA revisited: an updated database of carbon dioxide emissions from power plants worldwide. *Center for Global Development Working Paper*, 304, Washington, DC.
- Vogel, F. R., Hammer, S., Steinhof, A., Kromer, B. and Levin, I. 2010. Implication of weekly and diurnal 14C calibration on hourly estimates of CO-based fossil fuel CO₂ at a moderately polluted site in southwestern Germany. *Tellus B* **62**, 512–520.
- Vogel, F. R., Levin, I. and Worthy, D. 2013. Implications for deriving regional fossil fuel CO₂ estimates from atmospheric observations in a hot spot of nuclear power plant 14CO₂ emissions. *Radiocarbon* **55**, 1556–1572.
- Wang, R., Tao, S., Ciais, P., Shen, H. Z., Huang, Y. and co-authors. 2013. High-resolution mapping of combustion processes and implications for CO₂ emissions. *Atmos. Chem. Phys.* **13**, 5189–5203.
- WMO: Report of the 17th WMO/IAEA Meeting of Experts on Carbon Dioxide, Other Greenhouse Gases and Related Tracers Measurement Techniques (GGMT-2013). 2013. Beijing, China, 10–13 June 2013, GAW Report No. 213. Online at: <http://www.wmo.int/pages/prog/arep/gaw/gaw-reports.html>.
- Wu, L., Bocquet, M., Lauvaux, T., Chevallier, F., P., Y. and co-authors. 2011. Optimal representation of source-sink fluxes for mesoscale carbon dioxide inversion with synthetic data. *J. Geophys. Res.* **116**, D21304.

Appendix 1

A1 Further decomposition of the representation error

In Equation (4), we derived the representation error as $\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t$. As indicated in Section 2.2, it can be further decomposed to (1) errors arises from the modelling of concentrations at coarse resolution in the observation operator while it is compared to local measurements; (2) errors due to representing (averaging) the emissions at the coarse resolution of the transport model (into $\mathbf{f} = \mathbf{H}_{\text{distr}}^t \mathbf{x}^t$) and thus missing the true very-high resolution patterns in the emissions:

$$\begin{aligned} & \mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t \\ &= \left(\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transpHR} \rightarrow \text{LR}}^t \mathbf{f}_{\text{HR}}^t \right) \\ &+ \left(\mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transpHR} \rightarrow \text{LR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t \right) \end{aligned} \quad (\text{A1-1})$$

or in the other way round, accounting for the variability in the fluxes first:

$$\begin{aligned} & \mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}}^t \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t \\ &= \left(\mathbf{H}_{\text{transpHR}}^t \mathbf{f}_{\text{HR}}^t - \mathbf{H}_{\text{samp}}^t \mathbf{H}_{\text{transpLR} \rightarrow \text{HR}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t \right) \\ &+ \left(\mathbf{H}_{\text{samp}}^t \mathbf{H}_{\text{transpLR} \rightarrow \text{HR}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t - \mathbf{H}_{\text{samp}}^t \mathbf{H}_{\text{transp}}^t \mathbf{H}_{\text{distr}}^t \mathbf{x}^t \right) \end{aligned} \quad (\text{A1-2})$$

where $\mathbf{H}_{\text{transpHR} \rightarrow \text{LR}}$ is a theoretical operator corresponding to the transport from emissions $\mathbf{f}_{\text{HR}}$ but whose output concentrations are projected into the coarse resolution space of the outputs of $\mathbf{H}_{\text{transp}}$, while $\mathbf{H}_{\text{transpLR} \rightarrow \text{HR}}$ is a theoretical operator corresponding to the transport from emissions at low resolution but whose output concentrations are at high resolution.

Following the notations as in Section 3, we can derive a practical estimate corresponding to Equation (A1-1):

$$\mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{CHIM}} \mathbf{f}_{\text{HR}}^{\text{EDG-IER}} - \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transpHR} \rightarrow \text{LR}}^{\text{CHIM}} \mathbf{f}_{\text{HR}}^{\text{EDG-IER}} \quad (\text{A1-3})$$

and

$$\mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transpHR LR}}^{\text{CHIM}} \mathbf{f}_{\text{HR}}^{\text{EDG-IER}} - \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{CHIM}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}} \quad (\text{A1-4})$$

or the practical estimate corresponding to Equation (A1-2):

$$\mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{CHIM}} \mathbf{f}_{\text{HR}}^{\text{EDG-IER}} - \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transpLR}} \rightarrow \text{HR}^{\text{CHIM}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}} \quad (\text{A1-5})$$

and

$$\begin{aligned} & \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transpLR}} \rightarrow \text{HR}^{\text{CHIM}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}} \\ & - \mathbf{H}_{\text{samp}}^{\text{coloc}} \mathbf{H}_{\text{transp}}^{\text{CHIM}} \mathbf{H}_{\text{distr}}^{\text{EDG-IER}} \mathbf{x}^{\text{EDG-IER}} \end{aligned} \quad (\text{A1-6})$$

We have made those two practical computations for these two decomposition (not shown here). The dominant component is dependent on the order of the decomposition (A1-1 vs. A1-2), the first term being larger than the second term for each decomposition. Therefore, there is no robust characterization of each sub-component of the representation error, even though they correspond to different physical processes. We thus avoid further analysis of these decomposition of the representation error.

A2 Equivalence of the results for large region-months from an inversion at transport model resolution and from an inversion at large region-months resolution

In this section, we demonstrate that using the same atmospheric transport model but solving for emissions at the transport model resolution (called hereafter ‘pixel level’) or for large regions/time windows (accounting for aggregation error) should lead to the same results for large regions/time windows. The relationship between the control vectors at pixel level and at large region level is $\mathbf{x}_{\text{PIX}} = \mathbf{H}_{\text{distr}} \mathbf{x}$. We define the aggregation operator $\mathbf{\Gamma}$ which aggregates the emissions at pixel level into large regions and time windows. The product between $\mathbf{\Gamma}$ and $\mathbf{H}_{\text{distr}}$ yields the identity matrix ($\mathbf{\Gamma} \mathbf{H}_{\text{distr}} = \mathbf{I}$) and $\mathbf{\Gamma} \mathbf{x}_{\text{PIX}} = \mathbf{x}$. The observation operator when using a pixel level control vector writes $\mathbf{H}_{\text{PIX}} = \mathbf{H}_{\text{samp}} \mathbf{H}_{\text{transp}}$ and $\mathbf{H} = \mathbf{H}_{\text{PIX}} \mathbf{H}_{\text{distr}}$.

Assuming that errors have a Gaussian and unbiased distribution, at the pixel level, the prior uncertainty covariance is denoted $\mathbf{B}_{\text{PIX}}$. The pixel level uncertainty $\mathbf{\epsilon}_{\text{PIX}}$ is composed by an uncertainty at a resolution higher than that of the large region $\mathbf{\epsilon}_{\text{SUBR}}$ ($\mathbf{\Gamma} \mathbf{\epsilon}_{\text{SUBR}} = 0$) whose covari-

ance is denoted $\mathbf{B}_{\text{SUBR}}$ and verifies $\mathbf{\Gamma} \mathbf{B}_{\text{SUBR}} = 0$ (see below), and by the projection $\mathbf{H}_{\text{distr}} \mathbf{\epsilon}$ of the uncertainty at the level of the large region $\mathbf{\epsilon}$, whose covariance can be written $\mathbf{H}_{\text{distr}} \mathbf{B} \mathbf{H}_{\text{distr}}^T$. These 2 components are not necessarily independent, so that $\mathbf{B}_{\text{PIX}} = \mathbf{B}_{\text{SUBR}} + \mathbf{H}_{\text{distr}} \mathbf{B} \mathbf{H}_{\text{distr}}^T + 2\mathbf{B}_{\text{cov}}$ where $\mathbf{B}_{\text{cov}}$ is the matrix of covariance between $\mathbf{H}_{\text{distr}} \mathbf{\epsilon}$ and $\mathbf{\epsilon}_{\text{SUBR}}$ which verifies $\mathbf{\Gamma} \mathbf{B}_{\text{cov}} = 0$ (see below). Our bottom-up definition of the aggregation error is based on the assumption that $\mathbf{B}_{\text{cov}} = 0$ and corresponds to $\mathbf{R}_a = \mathbf{H}_{\text{PIX}} \mathbf{B}_{\text{SUBR}} \mathbf{H}_{\text{PIX}}^T$ (i.e. on the projection of $\mathbf{\epsilon}_{\text{SUBR}}$). On the other hand, the top down approach of (Kaminski et al. (2001)) fully account for $\mathbf{B}_{\text{cov}}$ by defining $\mathbf{R}_a = \mathbf{H}_{\text{PIX}} (\mathbf{B}_{\text{PIX}} - \mathbf{H}_{\text{distr}} \mathbf{B} \mathbf{H}_{\text{distr}}^T) \mathbf{H}_{\text{PIX}}^T$ i.e. $\mathbf{R}_a = \mathbf{H}_{\text{PIX}} (\mathbf{B}_{\text{SUBR}} + 2\mathbf{B}_{\text{cov}}) \mathbf{H}_{\text{PIX}}^T$. A common formulation of the covariance of the aggregation error is thus $\mathbf{R}_a = \mathbf{H}_{\text{PIX}} (\mathbf{B}_{\text{SUBR}} + 2\mathbf{B}_{\text{cov}}) \mathbf{H}_{\text{PIX}}^T$ with the assumption that $\mathbf{B}_{\text{cov}} = 0$ in our bottom-up framework (see the discussion regarding this assumption at end of this section).

The observation error covariance when using a pixel level control vector is denoted $\mathbf{R}_{\text{PIX}}$. Applying the alternative formulation of Equation (2) (ignoring $\mathbf{y}_{\text{fixed}}$ which is null in our framework):

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R})^{-1} (\mathbf{y}^o - \mathbf{H} \mathbf{x}^b) \quad (\text{A2-1})$$

at the pixel level and at the large region levels isolating aggregation error yields:

$$\mathbf{x}_{\text{PIX}}^a = \mathbf{x}_{\text{PIX}}^b + \mathbf{B}_{\text{PIX}} \mathbf{H}_{\text{PIX}}^T (\mathbf{H}_{\text{PIX}} \mathbf{B}_{\text{PIX}} \mathbf{H}_{\text{PIX}}^T + \mathbf{R}_{\text{PIX}})^{-1} (\mathbf{y}^o - \mathbf{H}_{\text{PIX}} \mathbf{x}_{\text{PIX}}^b) \quad (\text{A2-2})$$

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{B} \mathbf{H}^T (\mathbf{H} \mathbf{B} \mathbf{H}^T + \mathbf{R}_a + \mathbf{R}_{\text{PIX}})^{-1} (\mathbf{y}^o - \mathbf{H} \mathbf{x}^b) \quad (\text{A2-3})$$

Aggregating the results from Equation (A2-2) to large regions, we get

$$\begin{aligned} \mathbf{\Gamma} \mathbf{x}_{\text{PIX}}^a &= \mathbf{\Gamma} \mathbf{x}_{\text{PIX}}^b + \mathbf{\Gamma} \mathbf{B}_{\text{PIX}} \mathbf{H}_{\text{PIX}}^T \\ & (\mathbf{H}_{\text{PIX}} \mathbf{B}_{\text{PIX}} \mathbf{H}_{\text{PIX}}^T + \mathbf{R}_{\text{PIX}})^{-1} (\mathbf{y}^o - \mathbf{H}_{\text{PIX}} \mathbf{x}_{\text{PIX}}^b) \end{aligned} \quad (\text{A2-4})$$

$$\begin{aligned} \text{i.e. } \mathbf{x}^a &= \mathbf{x}^b + (\mathbf{\Gamma} \mathbf{B}_{\text{SUBR}} + 2\mathbf{\Gamma} \mathbf{B}_{\text{cov}} + \mathbf{\Gamma} \mathbf{H}_{\text{distr}} \mathbf{B} \mathbf{H}_{\text{distr}}^T) \mathbf{H}_{\text{PIX}}^T \\ & (\mathbf{H}_{\text{PIX}} \mathbf{H}_{\text{distr}} \mathbf{B} \mathbf{H}_{\text{distr}}^T \mathbf{H}_{\text{PIX}}^T + \mathbf{H}_{\text{PIX}} (\mathbf{B}_{\text{SUBR}} + 2\mathbf{B}_{\text{cov}}) \mathbf{H}_{\text{PIX}}^T + \mathbf{R}_{\text{PIX}})^{-1} \\ & (\mathbf{y}^o - \mathbf{H}_{\text{PIX}} \mathbf{H}_{\text{distr}} \mathbf{x}^b) \end{aligned} \quad (\text{A2-5})$$

which simplifies into Equation (A2-3) and thus it means that the inverted emissions at large scale are the same when using the two types of control vectors.

Regarding the demonstration that $\mathbf{\Gamma} \mathbf{B}_{\text{SUBR}} = 0$ and $\mathbf{\Gamma} \mathbf{B}_{\text{cov}} = 0$: detailing $\mathbf{B}_{\text{SUBR}}$ (and $\mathbf{B}_{\text{cov}}$) as the covariance of errors $\mathbf{\epsilon}_{\text{SUBR}}$ (between $\mathbf{H}_{\text{distr}} \mathbf{\epsilon}$ and $\mathbf{\epsilon}_{\text{SUBR}}$) with $\mathbf{\Gamma} \mathbf{\epsilon}_{\text{SUBR}} = 0$, we get $\mathbf{B}_{\text{SUBR}} = \mathbf{E}[\mathbf{\epsilon}_{\text{SUBR}} \mathbf{\epsilon}_{\text{SUBR}}^T]$ and $\mathbf{B}_{\text{cov}} = \mathbf{E}[\mathbf{\epsilon}_{\text{SUBR}} \mathbf{\epsilon}_{\text{SUBR}}^T]$ so that $\mathbf{\Gamma} \mathbf{B}_{\text{SUBR}} = \mathbf{\Gamma} \mathbf{E}[\mathbf{\epsilon}_{\text{SUBR}} \mathbf{\epsilon}_{\text{SUBR}}^T] = \mathbf{E}[\mathbf{\Gamma} \mathbf{\epsilon}_{\text{SUBR}} \mathbf{\epsilon}_{\text{SUBR}}^T] = 0$ and $\mathbf{\Gamma} \mathbf{B}_{\text{cov}} = \mathbf{\Gamma} \mathbf{E}[\mathbf{\epsilon}_{\text{SUBR}} \mathbf{\epsilon}_{\text{SUBR}}^T] = \mathbf{E}[\mathbf{\Gamma} \mathbf{\epsilon}_{\text{SUBR}} \mathbf{\epsilon}_{\text{SUBR}}^T] = 0$.

The assumption that $\mathbf{B}_{\text{cov}} = 0$ i.e. of the independence between the prior uncertainty and the uncertainty in the distribution of the emissions in the control regions $\mathbf{\epsilon}_{\text{SUBR}}$ (and consequently between the prior uncertainty and the aggregation error) is likely strong.

The typical inventories used to produce both the prior estimate of the emissions and $\mathbf{H}_{\text{distr}}$ are built on a mix of ‘top-down’ computations relying on a disaggregation of large-scale budgets of the emissions from statistics on the fossil fuel consumption, and on ‘bottom-up’ computations relying on emissions factors and local activity data. In the first type of computations, the data used for the disaggregation can be highly independent of the large-scale statistics which

would qualitatively support the assumption that $\mathbf{B}_{\text{cov}} = 0$. However, the statistics on the consumption of different fossil fuels can be used to derive the separation of the emissions between different types of anthropogenic activities (e.g. liquid fuel can be mainly related to traffic, gas and coal mainly to power generation ...) and thus to derive information on the spatial and temporal distribution of the emissions. Through such a process, uncertainties at large scale can thus be down-scaled into uncertainties at higher resolution. Furthermore, errors in the different emission factors used for the bottom-up computations generate uncertainties that are highly correlated between the high and low resolution.

The assumption that the prior uncertainty and the aggregation error are independent (e.g. $\mathbf{B}_{\text{cov}} = 0$) is thus unlikely, and we have tried to analyse the relative weight of the term $2\mathbf{H}_{\text{PIX}}\mathbf{B}_{\text{cov}}\mathbf{H}_{\text{PIX}}^T$ that we have implicitly ignored in the derivation of the aggregation error $\mathbf{R}_a = \mathbf{H}_{\text{PIX}}$

instead of $\mathbf{R}_a = \mathbf{H}_{\text{PIX}}(\mathbf{B}_{\text{SUBR}} + 2\mathbf{B}_{\text{cov}})\mathbf{H}_{\text{PIX}}^T$. Following a similar method and using the same notation as in Section 3, we can derive an estimate of the diagonal of $\mathbf{H}_{\text{PIX}}(\mathbf{B}_{\text{SUBR}} + 2\mathbf{B}_{\text{cov}})\mathbf{H}_{\text{PIX}}^T = \mathbf{H}_{\text{PIX}}$. $E[\mathbf{\epsilon}_{\text{SUBR}}(\mathbf{\epsilon}_{\text{SUBR}} + 2\mathbf{H}_{\text{distr}}\mathbf{\epsilon})^T]\mathbf{H}_{\text{PIX}}^T$ from statistics on the diagonal elements of $\left[\mathbf{H}_{\text{samp}^{\text{coloc}}}\mathbf{H}_{\text{transp}^{\text{LMDZ}}}(\mathbf{H}_{\text{distr}^{\text{EDG-IER}}} - \mathbf{H}_{\text{distr}^{\text{PKU}}})\mathbf{x}^{\text{EDG-IER}} \right]$.

Following the categorization discussed in Section 4.2, the difference between the corresponding estimate of the aggregation error and the one we have described in this study is about -0.03 ppm for both spring/summer and fall/winter. This would decrease the aggregation error by less than 20%. Considering the comparison with other sources of observation error (see Table 1), especially the representation error and the measurement error, such a modification would not impact the conclusions of this study.