

We thank both reviewers for their comments. Because of the time between the submission and the reception of the reviews, the LUMIA code had evolved significantly. In particular, the code used in the initial submission was written for python2.7, and was therefore not meant to evolve further (which in part justified not publishing the code on a public repository). We have now fully transitioned to python3, and have setup a website and git repository for the code (with a more complete documentation). This has also lead to some changes in the structure of the code, which are reflected in the paper. In response to the reviewer's comments, we have also modified some of the sensitivity tests (see list of changes below), and have removed some aspects of the results analysis.

Below are the reviewer's feedback, with our replies in red. Note that there are some differences from the initial Author comment that we posted after receiving the reviews (due to the time taken for the revision).

Further down in this document, we provide a summary of the changes, and finally a marked-up version of the manuscript

## Reviewer 1

This paper describes a regional flux inversion framework that is designed to have modular functionality. The performance of the system was demonstrated with a series of Observing System simulation experiments and real data experiments. They showed that the flux inversion has improved monthly mean fluxes and fitting to the observations irrespective of experimental setup in the OSSEs, but the fitting to the observations get worse at some sites with real observations. In spite of the improved monthly mean fluxes, the annual total fluxes get worse with almost all experiments. The paper did not explore ways to improve annual total flux estimate. Though this is a modeling development study, I would recommend more discussions about how to improve the system and the advantage of the regional flux inversion compared to existing flux inversion system.

Here are my detailed comments.

1. It is necessary to demonstrate that the regional flux inversion system developed here outperforms coarse global flux inversion. In this paper, TM5-4DVar is used as boundary conditions for regional flux inversions. I would recommend including discussions on the comparison between LUMIA and the TM5-4Dvar in both OSSE and real observation experiments.

We agree that the comparison between global and regional model is an interesting research question, but we feel that it would be difficult for us to do this in an appropriate way in this paper:

- At the continental scale, global inversions might in fact perform better than regional inversions, because there is no constraint on the CO<sub>2</sub> concentration of the outgoing air mass in regional inversions (and therefore on the net regional flux), so flux adjustments by the regional inversions do not have to be consistent with CO<sub>2</sub> observations outside Europe, in contrast this is implicitly taken care of in global inversions where there is by definition a constraint by the global atmospheric growth rate.
- The main benefit of regional inversions over global ones is the capacity to correctly assimilate observations from dense networks and/or more complex sites: the resolution of the global TM5 simulations used for producing our background concentrations is 6°×4° (450 km at 45°N). This is absolutely not adapted to the density of the ICOS network in Western Europe, and it is also not adapted for assimilating data at sites that are located close to strong CO<sub>2</sub> point sources such as urban centers. We expect the regional inversions to be more performant at such smaller sub-continental scales, but global inversions should remain more relevant at large scales.
- It would theoretically be possible to increase the resolution of the global model, but we would then encounter a performance limitation: a one-year LUMIA inversion with ICOS data typically takes 5 to 8 hours. A one-year TM5 inversion with a 1° zoom over Europe takes up to 6 days.

The comparison between global and regional inversions is therefore a rather complex topic, and we think that it should be done in a dedicated study. In particular, the TM5 inversion should use a higher resolution over Europe (it can run at 1° resolution), and the LUMIA inversion would benefit from additional developments, with in particular a higher temporal resolution of the optimization and more fine-tuning in the data selection.

2. The authors attributes the poor annual flux estimates to the larger adjustment to summer fluxes due to larger prior uncertainty. Since improving annual flux estimates is one of the major goals of regional flux inversions, I would recommend authors exploring ways to improve annual flux estimates, especially with OSSEs. In OSSEs, both true and prior fluxes are known, so specification of prior flux uncertainty can be based on the true prior flux errors. The percentage prior flux errors could be much larger during winter than during summer.

We have changed some of the sensitivity experiments to go in that direction. In particular, the new experiments SE.3H, SE.3Hcst and SE.x2 explore different distributions of the prior uncertainty throughout the year, which, unfortunately, do not really solve the problem. The fact that this metric is so sensitive to the inversion setup suggests that the surface observations used in our inversions do not appear to provide sufficient constraints to resolve the annual budget.

We discuss this aspect further in Section 6.1: one approach would be to add further constraints on the inversion (e.g. by adding more observations into the system, by adding constraints on the CO<sub>2</sub> concentration of the air exiting the regional domain, or by adding the annual budget as a control vector variable). We feel however that this question should be addressed via a specific, dedicated study (it is not specific to our model, but is rather a general issue with regional inversion systems).

3. Validation of flux estimates from top-down flux inversion is a necessary step to assess the quality of the system. The framework described in this paper is lacking the flux validation component. A common method is to compare the posterior concentrations against independent CO<sub>2</sub> concentrations. With high resolution regional fluxes, is it possible to use other independent observations?

It can technically be done, the same way than with a global inversion, but there are limitations that make it impractical: To our knowledge, there is no comprehensive observation dataset that could be used for validation purpose (for example, we are not aware of a European equivalent of the HIPPO campaigns that were made in the US). Also the interpretation of such comparison could be risky: the observation network is not homogeneous, and the performance of the inversion can vary widely from one region to the other, which would make a comparison with independent data difficult to interpret. A potentially more promising approach would be the use of satellite data, but this would also open many other questions, regarding how to best model them in a coupled system.

Instead, the validation of the results relies on three aspects: Firstly, the OSSEs provide a first demonstration that the inversion is functioning as expected. Secondly, performing an (ideally large) ensemble of sensitivity runs (including the use of alternative transport model, which hasn't been shown here) helps assessing the robustness of the results (and the limitations of the inversion system). Finally, the participation of inter-comparison exercises is a key aspect of the evaluation of the results.

4. Please add computational cost of each component of the inversion system

The computational cost of the LUMIA inversions is briefly mentioned in Section 6.2. That cost (5-8 hours on 24 CPUs) is dependent on the number of observations, and does not include the cost of computing the footprints and background concentrations:

- the cost of computing the footprints depends on the number of observations (it takes ~1 CPU hour to compute 50 footprints with FLEXPART, for our regional domain), however, each footprint needs to be computed only once.
- likewise, the TM5-4DVAR inversion from which the backgrounds are extracted takes approximately 24 hours to compute, for one year, but the background extraction itself (one TM5 forward run) takes only a few minutes (regardless the domain size and the number of observation sites). The inversion needs to be computed only once, and even that could be skipped in the future, by using the fluxes from a "routine" TM5 inversion, such as CarbonTracker (for CO<sub>2</sub>) or CAMS (for CH<sub>4</sub>).

The 5-8 hours therefore correspond only to the marginal cost of each additional inversion performed. By comparison, a one year TM5 inversion with a 1x1 zoom over Europe takes 5 to 7 days to complete. Our background TM5 inversion is not a good base for comparison, as it is really optimized for speed, at the detriment

of accuracy of the posterior fluxes (it needs to fit well the observed spatial gradients of CO<sub>2</sub>, but the accuracy of the optimized fluxes is secondary as those are not kept, therefore we can reduce the number of iterations).

5. Page 24, line 516, replace “than” with “as”

The whole sentence has been modified

6. Page 26, Line 531, replace “im” with “in”.

The whole sentence has been modified

## Reviewer 2

It is a good effort by the authors but I had a hard time understanding whether it is an algorithm/software oriented paper or paper completely devoted to scientific results.

As explained in the paper, the inversion algorithm is not new, neither are the transport models used (TM5 and FLEXPART), however this particular combination has never been implemented before, and the software used to implement that coupling (LUMIA) is new. The experiments presented are designed to test the performances and limitations of the system, rather than to derive actual scientific conclusions on the European NEE (we do not discuss those). We think that it is a necessary stage before the use of the system for production of scientific results.

The authors themselves say quote:

"The inversion technique used in this study is by design not innovative (the definition of the control vector, the specification of the uncertainties, etc. replicate what has been done in previous studies (e.g. Kountouris et al. (2018)), as the aim is to have a reference setup. The scientific results are therefore at this stage limited (as it also wasn't the aim of the paper), but the analysis of the OSSEs results show that the inversions are working as expected" [So nothing new here].

If it is not the aim of the paper then why devote half the manuscript to it !!!

We agree with the reviewer that some parts of the results section were leaning too much towards scientific results, which was not in the scope of this paper. We have therefore largely revised the results sections. We have slightly reduced the number of figures, and have removed the analysis by regions, which makes Section 5 shorter (Inversions with real observations). However, Section 4 has increased a bit in length, in response to reviewer's 1 comments, so the overall length of the results analysis did not change.

With respect to code the authors say quote: "

"The LUMIA code is not meant to be a "key in hand" system, it target users having or willing to acquire robust understanding of inverse modelling (it is perfectly usable as a toy model for learning). We therefore do not publish the code in a public repository, but we are very open to collaborations and distribute the code on-demand. An archive of the code in it's current shape is nonetheless included as SI of this document"

Maybe there was a confusion here: the code was not made available on a public (git/svn/ftp) repository, but the (entire) code was published as SI of the manuscript, with a documentation. We have however improved on that aspect, and we now have a proper website (<https://lumia.nateko.lu.se>) and git repository, from which the code can be downloaded. The repository also contains several pages of documentation/tutorial, in the form of jupyter notebooks.

So the authors do not think that their results are innovative or new, and they are not willing to publish the full code so I do not understand what is the strength of the paper?

The aim of the paper is to publish and document the new LUMIA inversion system, also in order to provide a reference for future research building on it.

Furthermore, it looks like there code is just for their own research.

We indeed developed the code primarily for our own research (as do most research groups?), but we do not think that this would be a sufficient reason not to document it and test it correctly.

Many groups have similar code and in some cases better documented fully functional codes are also publicly available. For e.g. see <https://www.geosci-model-dev-discuss.net/gmd-2019-185/> and the code is also publicly available: [https://github.com/greenhousegaslab/geostatistical\\_inverse\\_modeling](https://github.com/greenhousegaslab/geostatistical_inverse_modeling) Another one: <https://www.esrl.noaa.gov/gmd/ccgg/carbontracker-lagrange/doc/index.html>

These systems did not exist, or were only in early stage of their development when we started working on LUMIA (and they are not directly comparable to it either). And again, the fact that other systems with a similar purpose exist should not be a sufficient justification for disqualifying our work: replication of results is an important part of research.

Writing and keeping up to date an extensive documentation such as the one of CT-Lagrange is a time investment that is only justified if the user-base is large enough. There are indeed many other research groups developing internally similar types of codes, but not making them publicly available. This is not a good practice, but discouraging authors to publish on the basis that the documentation is insufficient leads to this situation.

The compromise that we had found was to distribute the code along with the manuscript, and to encourage interested users to contact us (which is always a good practice when using the research code of someone else). For this revised version, we have however made the code and documentation available online.

Please help me understand what is new in the manuscript!!

# Summary of the changes

## Major revisions

- Sections 2.2 and 2.3 have been merged and completely re-written, to reflect the changes in the code. Figure 1 (the flow diagram of the inversion) has also been edited to reflect these changes.
- Section 3.4.1 has been simplified (only the “gaussian” approach to construct the horizontal covariance length is now described, the “exponential” approach has been left out, as it did not add value to the paper). On the other hand, the section now lists two other approaches used to construct the prior uncertainty vector.
- Section 4 has been almost entirely re-written, and the selection of inversions that we show has been modified:
  - Inversions SB and SG have been removed; All the inversions now use covariance matrix constructed from Gaussian covariance decay functions (so SRef is now identical to the previous SG). The inversion SE.3H, SE.3Hcst and SE.x2 are new, and test different construction of the uncertainties (in response to reviewer’s 1 comment). All the inversion names have been modified (to reflect the grouping of sensitivity tests: inversions SE.3H, SE.3Hcst and SE.x2 test the impact of the prior uncertainties; SC.100 and SC.500 test the impact of changes in the prior error covariances; SO.A and SO.P test changes in the observation network).
  - The sub-structure of the section has been modified as well: we now first analyse the reference inversion, in terms of monthly, annual and pixel-scale fluxes and errors, and then look at the impact of the sensitivity runs. The analysis by regions has been removed.
  - Figure 5 has been modified, to show the sensitivity runs as an ensemble, for clarity and conciseness. The individual results can still be found in SI.
  - Figure 6 has been removed.
  - Figure 7 (now 6) has been highly simplified. The information from the previous plots can still be found in SI Figures, but is not discussed in the paper.
- Section 5 has been largely simplified, and the inversion selection has been modified as well:
  - The inversions RE.3H, RE.3Hcst and RE.x2 have been added (the selection of inversions is now similar in Section 4 and 5). The other inversions have been renamed following the naming convention used for Section 4.
  - The regional analysis has been removed, and as a consequence Figure 9 (now 8) has been simplified. The entire section has been made shorter and the text is now focusing only on the differences between OSSEs and real simulations. The sensitivity tests are not discussed individually, but only as an ensemble.
- The discussion Section 6.1 has been completely rewritten, in part to account for some of the comments of reviewer 2, and in part to reflect the changes in the results section.
- The content of the conclusion has now been moved to a new discussion section 6.3 (The LUMIA framework: conclusions and future perspectives), and enriched, in part also to account for the comments of reviewer 2.

## Minor revisions

- Minor edits in the abstract to better reflect the aims of the paper
- Section 2.1 renamed to “theoretical background”, and small textual changes.

- A paragraph has been added at the end of Section 3.2, to clarify the difference between observation operator and transport model.
- Figure 11 (now 10) has been replotted, since an error was found on the original figure (the bias instead of the RMSE reduction was shown in the left panel). The right panels of Figures 11 (now 10) and 8 (now 7) have also been edited, for clarity.

# Regional CO<sub>2</sub> inversions with LUMIA, the Lund University Modular Inversion Algorithm, v1.0

Guillaume Monteil<sup>1</sup> and Marko Scholze<sup>1</sup>

<sup>1</sup>Department of Physical Geography and Ecosystem Science, Lund University, Lund, Sweden

**Correspondence:** Guillaume Monteil (guillaume.monteil@nateko.lu.se)

**Abstract.** Atmospheric inversions are ~~commonly used for estimating large-scale (continental to regional)~~ used to derive constraints on the net sources and sinks of CO<sub>2</sub> and other stable atmospheric tracers from their observed concentrations. ~~Recently~~ The resolution and accuracy the fluxes can be estimated with depends, among other factors, on the quality and density of the observational coverage, on the precision and accuracy of the transport model used by the inversion to relate fluxes to observations, and on the adaptation of the statistical approach to the problem studied (Michalak et al., 2016).

In recent years, there has been an increasing demand from stakeholders for ~~robust estimates of greenhouse gases at country-scale (or higher)~~ resolution inversions at higher spatial resolution (country scale), in particular in the framework of the Paris agreement. This ~~increase step up~~ increase step up in resolution is in theory enabled by the growing availability of observations from surface in-situ networks (such as ICOS in Europe) and from remote sensing products (OCO-2, GOSAT-2). The increase in the resolution of inversions is also a necessary step to provide efficient feedback to the ~~process-based (bottom-up)~~ modelling community (vegetation models, fossil fuel emission inventories). ~~This, however,~~ etc.). It however calls for new developments in the inverse ~~modelling systems, mainly in terms of models;~~ diversification of the inversion approaches, shift from global to regional inversions, and improvement in the computational efficiency, etc.

We ~~have~~ developed the Lund University Modular Inversion Algorithm (LUMIA) as a tool to address some of these new ~~developmentst~~ topics. LUMIA is meant to ~~be~~ become a platform for inverse modelling developments at Lund University. It aims at being a flexible, yet simple and easy to maintain set of tools that the modellers can combine to build inverse modelling experiments. It is in particular designed to be transport model agnostic, which should facilitate isolating the transport model errors from those introduced by the inversion setup itself. ~~Here, we briefly describe~~

This paper describes briefly the LUMIA framework; the motivations for ~~developing LUMIA as well as the underlying building it, the~~ development principles, current status and future prospects. ~~We present~~ Then a first LUMIA inversion setup ~~for a~~ is presented, to perform regional CO<sub>2</sub> inversions ~~over Europe, based on a new~~ in Europe, using in-situ data from surface and tall tower observation sites. Since LUMIA doesn't come with its own transport model, the transport of fluxes is computed using a coupling between the Lagrangian FLEXPART transport model (high resolution foreground transport) and the global coarse resolution TM5 transport ~~models, using in-situ data from surface and tall tower observation sites~~ model (following the approach of Rödenbeck et al. (2009)). This particular coupling is new and therefore also described in this paper.

## 1 Introduction

The accumulation of greenhouse gases in the atmosphere is the main driver of climate change. The largest contribution of anthropogenic activities to global warming is through the release of fossil carbon (mainly as CO<sub>2</sub>) to the atmosphere, but other human activities such as land use change (for agriculture, deforestation, etc.) also play a significant role. The climate forcings from this increased CO<sub>2</sub> concentration is likely to induce feedbacks through reactions of the terrestrial ecosystems and of the oceans (Stocker et al., 2013). Our capacity to correctly predict climate change, anticipate and mitigate its effects depends therefore largely on our capacity to model and predict the evolution of carbon exchanges between the atmosphere and other reservoirs.

For future climate simulations, the only available option is through “direct” (bottom-up) modelling of the different components of the biogeochemical cycles, i.e. using models (numerical or statistical) that simulate, as accurately as possible (given the precision requirements of the simulation), greenhouse gas fluxes to and from the atmosphere. For past periods, however, the “inverse” (top-down) approach is also possible, in which the greenhouse gas fluxes are diagnosed from their observed impact on atmospheric greenhouse gas concentrations.

Direct and inverse approaches are complementary, the former can provide detailed estimates of the spatial and temporal variability of the fluxes, but often with large uncertainties on the total fluxes Sitch et al. (2015) . On the contrary, inverse approaches provide robust estimates of total fluxes at large scales consistent with the observations (e.g. Gurney et al. (2002)), but with poor sensitivity to smaller scales (e.g. Peylin et al. (2013)).

An atmospheric inverse model typically couples an atmospheric transport model (which computes the relationships between fluxes and concentrations) with an inversion algorithm, whose task is to determine the most likely set of fluxes, within some prior constraints and given the information from an observation ensemble (in a Bayesian approach). In practice, inversions are complex codes, computationally heavy. The complexity arises in a large part from the necessity to combine large quantities of informations from sometimes very heterogeneous datasets (various types of observations, flux estimates, meteorological forcings, etc.). The computational weight depends largely on that of the underlying transport model, which usually needs to be ran a large number of times (iteratively or as an ensemble).

In recent years, the availability of observations has grown by orders of magnitude, with the deployment of high-density surface observation networks (such as the Integrated Carbon Observation System, ICOS, in Europe) and the fast developments in satellite retrievals of tropospheric greenhouse gas concentrations (GOSAT, OCO-2, etc.). Meanwhile, the demand for inversions is increasing, in particular from stakeholders such as regional, national or trans-national governments who are interested in country-scale inversions as a means of quantifying their carbon emissions, in connection with emission reduction targets as defined in the Paris agreement (Ciais et al., 2015).

This context puts strains on the existing inverse models. The larger availability of high quality data means that fluxes can be constrained at finer scales, but it also means that models of higher definition and precision must be used. The development of regional inversions (of varying scales) allows in theory an efficient usage of high resolution data while preserving a reasonable computational cost, but comes with specific challenges such as the need of more boundary conditions and the lack of options



60 for cross-validation when the resolution increases and the domain size shrinks. The demands from various stakeholders (policy makers, bottom-up modellers, medias, etc.) also call for developments in the inversion techniques, with for instance a more pronounced focus on the quantification of anthropogenic sources (Ciais et al., 2015) or the optimization of ecosystem models parameters instead of CO<sub>2</sub> fluxes in carbon cycle data assimilation systems (CCDAS) (Kaminski et al., 2013).

To enable such progress in the method and quality of the inversions, it is important to have a robust and flexible tool. The purpose of LUMIA (Lund University Modular Inversion Algorithm) is to be ~~not an integrated, well defined inverse model, but~~ a development platform for top-down experiments. LUMIA was developed from the start as a model-agnostic inversion tool, with a clear isolation of the data stream between the transport model and the optimization algorithm in an interface module. One of the main aims is to eventually allow a better characterization of the uncertainty associated to the transport model. Strong emphasis was put on the usability (low barrier entry code for newcomers, high degree of modularity to allow users to build  
70 their experiments in a very flexible way) and sustainability of the code (small, easily replaceable one-tasked modules instead of large multi-option ones).

This paper presents the ~~general concepts of LUMIA~~ LUMIA inversion framework, as well as a first application of regional (European) CO<sub>2</sub> inversions for Europe. The inversions use in-situ CO<sub>2</sub> observations from European tall towers (now part of the ICOS network, see <https://www.icos-ri.eu>) and rely on a regional transport model based on a new coupling between the  
75 FLEXPART Lagrangian particle dispersion model (Seibert and Frank, 2004; Pisso et al., 2019) (foreground, high resolution transport) and TM5-4DVAR (Meirink et al., 2008; Basu et al., 2013b) (background concentrations). The paper is organized as follows: First, Section ~~???~~ presents the LUMIA framework (general principles and architecture). Then Section 3 presents the specific inverse modelling setup used here (including the FLEXPART-TM5 coupling). Sections 4 and 5 present the results from two set of inversions (against synthetic and real observations). Finally, a short discussion summarises the main outcomes  
80 of the paper in Section 6.

## 2 The LUMIA framework

### 2.1 Theoretical background

### 2.2 ~~General inversion principles, and implications on the code structure~~

The general principle of an atmospheric inversion is to determine the most likely estimate of a set of variables controlling the  
85 atmospheric content and distribution of a tracer (typically sources and sinks, but also initial or boundary conditions), given a set of observations of that tracer's distribution in the atmosphere. The link between the set of parameters to optimize (control vector  $\mathbf{x}$ , of dimension  $n_x$ ) and the observed concentrations (observation vector  $\mathbf{y}$ , of dimension  $n_y$ ) is established by a numerical model of the atmospheric transport (and of any other physical process relating the state and observation vectors):

$$\mathbf{y} + \varepsilon_y = H(\mathbf{x} + \varepsilon_x) + \varepsilon_H \quad (1)$$

90 ~~*H* is the observation operator, which includes the atmospheric~~

~~The observation operator  $H$  includes the transport model itself, but also any additional step between the physical quantities represented by  $\mathbf{x}$  and steps needed to express  $\mathbf{y}$  as a function of  $\mathbf{x}$  (aggregation/disaggregation of flux components, accounting of boundary conditions and of non-optimized fluxes, etc.). The error terms  $\varepsilon_y$ ,  $\varepsilon_x$  and  $\varepsilon_H$  are respectively the observation error, the control vector error and the model representation error (see Section ??).~~

95 3.3.1. In the simplest cases, the system can be solved for  $\mathbf{x}$  analytically, but most often inversions use the Bayesian inference approach: the optimal control vector  $\hat{\mathbf{x}}$  is the one that allows the best statistical compromise between fitting the observations and limiting the departure from a prior estimation of the control vector  $\mathbf{x}_b$  (accounting for the (prescribed) uncertainties in the observations and the prior). Mathematically, this means finding the vector  $\hat{\mathbf{x}}$  that minimizes a cost function  $J(\mathbf{x})$  defined (in our case) as

$$\begin{aligned} J(\mathbf{x}) &= \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) \\ &\quad + \frac{1}{2} \sum_j (\mathbf{H}\mathbf{x} - \mathbf{y}_j)^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}_j) \\ 100 \quad &= J_b + J_{obs} \end{aligned} \tag{2}$$

where the prior ( $\mathbf{B}$ ) and observation ( $\mathbf{R}$ ) error covariance matrices weigh the relative contributions to the cost function of each departure from each prior control variable  $x_b^i$  and from each observation  $y^j$ . The optimal control vector  $\hat{\mathbf{x}}$  is solved for analytically (for small scale problems) or approximated step-wise (variational and ensemble approaches are most common (Rayner et al., 2018)).

105 An inversion system is therefore the combination of an observation operator (i.e. transport model, sampling operator, etc.), an inversion technique and a set of assumptions on the prior values of the variables to estimates, their uncertainties and the uncertainties of the observations. Each of these components introduces its own share of uncertainty, which makes the results harder to interpret: which feature of the solution is real, and which is introduced by e.g. the transport model, or incorrect assumptions on some uncertainties?

110 ~~These questions are difficult to address with inversion systems tightly integrating a transport model and an implementation of a specific inversion technique. LUMIA is therefore developed as a flexible inversion library, with a clear isolation between~~

## 2.2 The lumia python package

The LUMIA system is designed with the aim to provide the modularity needed to quantify the impact of the inversion design choices on the inversion results themselves. The strict isolation of the transport model also enables the transport model and the  
115 inversion algorithm ~~itself. LUMIA is primarily designed of a collection of modules, as simple and independent as reasonable, that can be used as basic elements to construct (inverse or forward) transport modelling experiments.~~

120 While the name “LUMIA” refers to that code library, its development (and first application) stem from the need of performing regional CO<sub>2</sub> inversions based on observations using in-situ data from the ICOS network. At this stage, the existing code is therefore largely limited to that regional inversion setup. The distinction between the LUMIA inversion library and the specific LUMIA setup presented in the rest of this document is therefore somewhat theoretical but we maintain the prospect of future developments.

### 2.3 Architecture and technical implementation

125 The flowchart in Figure 1 represents the data flow in our current inversion setup (see Section 3). The blocks within the black box are part of the actual LUMIA library, while blocks outside are either pre-processing steps (import of fluxes and observations) or external tools called by LUMIA (transport model).

130 Among the LUMIA modules (in green), two are primarily databases. The *optimData* module stores data in the optimization space (control vectors (prior, posterior and intermediate steps), coordinates of the control variables (lat, lon, time, category), uncertainties); the *obsdb* module stores data in the observations space: observation vector, observations coordinates and uncertainties, model estimates for the observations, etc. (the exact list of fields stored here is totally user-dependent. In our case the path to the observation response functions are stored (See Section ??), but one could also store for instance some data quality flag, or different model estimates for a same observation dataset). The two modules include save to evolve independently. On the other hand, the modularity does lead to an increase in the overall complexity of the code (due to the need to develop and maintain generic interfaces), which can end up being counterproductive if it limits the performances and load functions and can be instantiated independently, which allows their use in pre- and post-processing stages (preparation of the inversion inputs, analysis of the results, diagnostics or usability of the system. We nonetheless believe that the benefit of a higher modularity outweighs the risks. The potential adverse effects can be mitigated by careful design choices. The code is distributed as a single python package, with the following structure (see also Figure 1):

140 – The *lumia* folder contains the *lumia* python library, which implements the basic components of the inversion such as data storage (control vector, fluxes, observations, uncertainty matrices) and functions (forward and adjoint transport, conversion functions between fluxes and control vector; cost function evaluation, etc.).

The core of the inversion setup is the *Optimizer* module, which implements the actual steps of the optimization algorithm (as described in Section ??). It does not itself perform any calculation but calls routines from the other modules and controls the data flow between them.

145 – The *transport* folder contains the code that was used to implement the TM5-FLEXPART transport model coupling, described further in Section 3.2.

– The *src* folder contains the FORTRAN source code for the conjugate gradient minimizer used in the example inversions (see Section 3). Replacing this external code by a native python equivalent is planned.

- The *doc* folder contains a documentation, mainly in the form of jupyter-notebooks, and example data and configuration files.
- 150 - The *minimizerGMDD* module handles the communication with an external conjugate gradient executable. folder contains the scripts and configuration files used for producing the results presented further down in this manuscript.

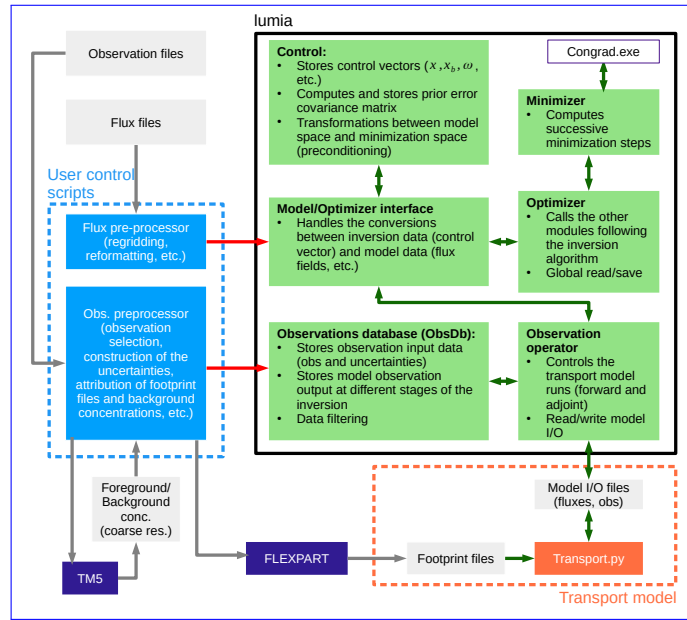
The *transport* module is the main interface between the inversion algorithm and the actual transport model. It constructs the transport model input based on a given control vector, triggers the actual transport modelling runs and constructs model-data mismatch vectors based on transport model output. The module is also designed so as to be instantiatable independently, for  
155 *instance to allow forward transport runs* package can be installed using the standard 'pip' command, which installs *lumia* and *transport* as python modules, which can then be imported from any python script. The *lumia* module itself has a relatively flat hierarchy, which limits the risk that replacing or changing one component prevents the others from working. The implementation of alternative features is preferably carried out via the development of alternative classes, which allows each individual class to remain compact and easy to understand and maintain.

160 The two pre-processing modules (in blue on the diagram) contain routines to import fluxes and observations from various data providers (handling of different file formats, units conversions, etc.); and are in charge of filling in the databases in *obsdb*, *lumia* and *optimData*. These preprocessing steps typically represent a large fraction of the total code of (any) inversion setup, but are very user-*transport modules* and application-specific. We chose to explicitly exclude them from the LUMIA library, as the benefit of including them is far outweighed by the associated code maintenance cost. These pre-processing steps are also  
165 not necessarily performed at the same time, and on the same computer, as their submodules can be used totally independently from the inversion scripts that are provided in the scripts folder. This allows their use in different contexts, such as development, pre/post-processing of the inversion data or during the analysis of the results (and eventually this helps keeping the inversion scripts compact, as they need only to focus on the inversion itself.

The execution of the actual inversion is controlled by a master python script, written by the user and for which an example  
170 is provided in SI. The transport operator itself (purple box) is formally independent of the LUMIA library, nonetheless our transport script, which relies on pre-computed observation response functions, is distributed and can be used as an example (in SI). The scope of the *lumia* library is intentionally vague: it should permit easily constructing inversion experiments and is primarily designed for it, but our current design choices should not over-constrain the alternative use cases (such as e.g. forward transport model experiments or optimization of land surface model parameters).

### 175 3 **Inversion-Test inversion setup**

Our test inversion setup is designed to optimize the monthly net atmosphere-ecosystem carbon flux (NEE, Net Ecosystem Exchange) over Europe at a target horizontal resolution of  $0.5^\circ$ , using  $\text{CO}_2$  observations from the European ICOS network (or similar/precursor sites). Two series of inversions are presented: First, a series of Observing System Synthetic Experiments (OSSEs), using known truth and synthetic observations; then a series of inversions constrained by real observations. All the



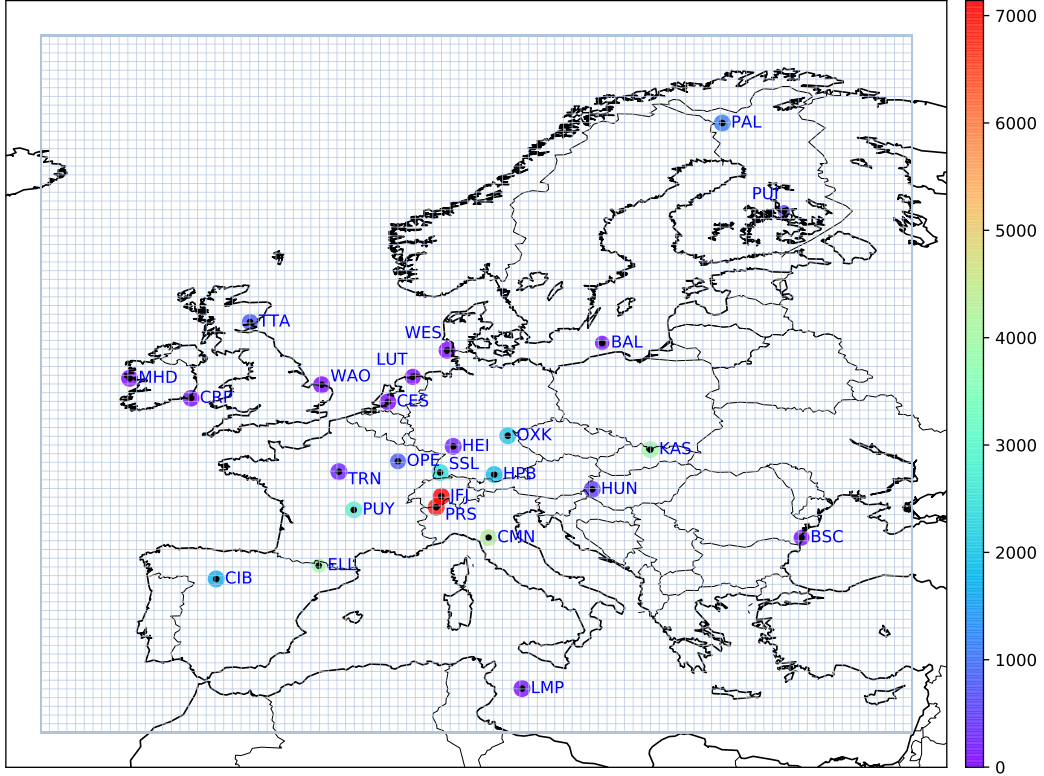
**Figure 1.** Inversion flow diagram. Green boxes are LUMIA modules, purple boxes are external codes and blue boxes are pre-processing scripts, typically user- and application-specific. The black thick line marks the limits of LUMIA itself. The red arrows show the flow green boxes represent code that is part of data in the inversion setup phase and the green arrows show lumia python module; the data flow during orange box shows operations performed by the inversion. The inversion inputs are prepared in a preparation phase (blue arrows), which involves external models (FLEXPART and TMS). The atmospheric transport model used in the inversion is considered an external code (here is in our case a simple python script that reads in observations, fluxes and footprints, but a full transport model could be plugged in instead); the blue boxes show code that is typically user- and application-specific (pre-processing of data and main inversion control script). The boxes in grey mark on-disk data, and the boxes in purple show external executables

180 inversions are performed on a domain ranging from 15°W, 33°N to 35°E, 73°N (illustrated in Figure 2, and hereafter referred to as the Regional Inversion Domain, RID) and cover the year 2011. The following sections describe the variational inversion technique, the transport model and the problem constraints (prior fluxes and observations).

### 3.1 Inversion approach

We use a Bayesian variational inversion algorithm, similar to that used in TM5-4DVAR inversions (Basu et al., 2013b; Meirink et al., 2008). In a variational inversion, the minimum of the cost function  $J(\mathbf{x})$  (Equation 2) is solved for iteratively:

- An initial "prior" run is performed to compute the concentrations ( $\mathbf{y}_m = \mathbf{H}\mathbf{x}_b$ ) corresponding to the prior control vector  $\mathbf{x}_b$  (since the transport is linear,  $H(\mathbf{x}) \equiv \mathbf{H}\mathbf{x}$ ).
- The local cost function ( $J(\mathbf{x} = \mathbf{x}_b)$ ) and cost function gradient ( $\nabla_{\mathbf{x}} J(\mathbf{x} = \mathbf{x}_b)$ ) are computed.
- A control vector increment ( $\delta\mathbf{x}$ ) is deduced from the gradient, and the process is repeated from step 1 (with  $\mathbf{x} = \mathbf{x}_b + \delta\mathbf{x}$ ), until a convergence criterion is reached.



**Figure 2.** Regional inversion domain and location of the observation sites. The area of the dots is proportional to the number of observations available at each site (the actual number of observations is reduced by the filtering described in Section 3.3.5) and their color represents the altitude of the sites.

The control vector increments are computed using an external library implementing the Lanczos algorithm (Lanczos, 1950). For efficiency (reduction of the number of iterations) and practicality (reduction of the number of large matrix multiplications) reasons, the optimization is performed on the preconditioned variable  $\omega = \mathbf{B}^{-1/2}(\mathbf{x} - \mathbf{x}_b)$  (following Courtier et al. (1994) and similar to the implementation in Basu et al. (2013b)). Equation 2 then becomes

$$195 \quad J(\omega) = \frac{1}{2}\omega^T\omega + \frac{1}{2}\left(\mathbf{HB}^{1/2}\omega + \mathbf{d}_0\right)^T \mathbf{R}^{-1}\left(\mathbf{HB}^{1/2}\omega + \mathbf{d}_0\right) \quad (3)$$

with  $\mathbf{d}_0 = \mathbf{H}\mathbf{x}_b - \mathbf{y}$  the prior model-data mismatches. In this formulation, the cost function gradient is given by

$$\begin{aligned}\nabla_{\omega} J &= \omega + \mathbf{B}^{\mathbf{T}/2} \mathbf{H}^{\mathbf{T}} \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y}) \\ &= \omega + \mathbf{B}^{\mathbf{T}/2} \nabla_{\mathbf{x}} J_{obs}\end{aligned}\quad (4)$$

The non-preconditioned observational cost function gradient  $\nabla_{\mathbf{x}} J_{obs} = \mathbf{H}^{\mathbf{T}} \mathbf{R}^{-1} (\mathbf{H}\mathbf{x} - \mathbf{y})$  is computed using the adjoint technique (Errico, 1997). The transformation matrix  $\mathbf{B}^{1/2}$  is obtained by eigen-value decomposition of  $\mathbf{B}$ . Note that in this formulation, the inverse of  $\mathbf{B}$  (or the square root of its inverse) is actually never needed, making it possible to constrain the inversion with a non invertible matrix.

In practice, the preconditioning adds two extra steps to the algorithm described above: conversion from  $\omega$  to  $\mathbf{x}$  ( $\mathbf{x} = \mathbf{B}^{1/2}\omega + \mathbf{x}_b$ ) before applying the transport operator (i.e. running the transport model), just before step 1; Conversion from  $\nabla_{\mathbf{x}} J_{obs}$  to  $\nabla_{\omega} J_{obs}$  (just after step 2). The initial preconditioned control vector is filled with zeros and corresponds to  $\mathbf{x}_0 = \mathbf{x}_b$ .

### 205 3.2 Observation operator (and transport model)

The observation operator ( $H$  in Equation 2) groups the ensemble of operations to compute the  $\text{CO}_2$  concentrations corresponding to a given control vector. In our case, this covers the disaggregation from the monthly fluxes in the control vector to a 3-hourly temporal resolution, the addition of prescribed fluxes (fossil, ocean and biomass burning categories), their transport to the observations location and the addition of background concentrations.

210 In this first implementation of  $\text{CO}_2$  inversions with LUMIA, we opted for a regional transport model based on pre-computed observational response functions (footprints):

$$y_m^i = y_{bg}^i + \sum_i \sum_c \langle \mathbf{K}_i, \mathbf{f}_c \rangle \quad (5)$$

where the footprint  $\mathbf{K}_i$  stores the sensitivity of the observation  $y^i$  to the surface fluxes  $\mathbf{f}_c$  (with the index  $c$  referring to the flux category), and  $y_{bg}^i$  is the contribution of background and historical fluxes to the model estimate  $y_m^i$  of observation  $y^i$ .

215 The flux vectors  $\mathbf{f}_c$  are constructed from the control vector for the optimized flux category (NEE) and prescribed for the other fluxes. The relation between the control vector and the  $\mathbf{f}_{nee}$  is given by:

$$\begin{aligned}\mathbf{f}_{nee}^t &= \frac{\mathbf{x}^m - \mathbf{x}_b^m}{n_t} + \mathbf{f}_{0,nee}^t \\ \mathbf{x}_b^m &= \sum_t \mathbf{f}_{0,nee}^t\end{aligned}\quad (6)$$

where  $\mathbf{f}_{nee}^t$  is the NEE flux map at time step  $t$  of the month  $m$ ,  $\mathbf{f}_{0,nee}^t$  is the corresponding prior NEE map,  $\mathbf{x}^m$  and  $\mathbf{x}_b^m$  are the control vector and prior control vector components corresponding to month  $m$  at the same spatial coordinates, and  $n_t$  is

220 the number of three-hourly intervals in the month  $m$ . In other words, the inversion adjusts an offset to the prior, high temporal resolution fluxes.

The adjoint operations corresponding to Equations 5 and 6 are summarized by

$$\mathbf{x}_{adj}^m = \frac{1}{n_t} \sum_t^{n_t} \sum_i \mathbf{K}_i^t \delta y^i \quad (7)$$

with  $\delta y^i$  the model-data mismatches weighted by their uncertainties (See Section 3.3.1).

225 Since  $\mathbf{K}$  and  $y_{bg}$  are constant throughout the inversion iterations, they can be pre-computed, which reduces the transport computations to a set of very simple matrix operations. This tremendously reduces the computational cost of the inversions but increases the I/O and storage requirements (one response function  $\mathbf{K}$  must be stored for each observation and is read at each forward and adjoint iteration).

230 Although the operations described above are all part of the observation operator ('lumia.obsoperator' module), the actual forward and adjoint flux transport (Equation 5 and its adjoint counterpart  $f^{adj} = H^{adj}(dy)$ ) are performed by the transport model ('transport' python library), called as a subprocess. The transport model relies on pre-computed observation footprints (response functions) and time series of background concentrations:

### 3.2.1 Response functions (regional transport model)

The response functions ( $\mathbf{K}$ ) were computed using the FLEXPART 10.0 Lagrangian transport model (Seibert and Frank, 2004; 235 Stohl et al., 2010). FLEXPART simulates the dispersion, backwards in time from the observation location, of a large number of virtual air "particles". The response function  $\mathbf{K}_i^\phi$  corresponds to the aggregated residence time of the particles released for observation  $y^i$ , in a given space-time grid box  $\phi$  of the regional inversion, and below a threshold altitude layer arbitrarily set to 100m).

240 The simulations were driven by ECMWF ERA-Interim reanalysis, extracted at a 3-hourly temporal resolution, and on a  $0.5^\circ \times 0.5^\circ$  horizontal resolution, on a regional domain ranging from  $25^\circ\text{W}$ ,  $23^\circ\text{N}$  to  $45^\circ\text{W}$ ,  $83^\circ\text{N}$ , slightly larger than the inversion grid, which allows for some accounting of particles re-entry (i.e. when an air mass leaves the inversion domain, and re-enters it later, which is not accounted for in the background).

245 One set of 3-hourly response functions was computed for each observation, up to seven days backward in time (less if all the particles leave the domain sooner). For plain or low altitude sites (see Table 1), the particles were released from the sampling height above ground of the observations. For high altitude sites (around which the orography is unlikely to be correctly accounted for), the particles were released from the altitude above sea level of the observation sites.

The response functions are stored in HDF5 files, following a format described in SI. ~~More detailed information on the configuration of the FLEXPART runs is also provided in SI.~~



### 3.2.2 Background concentrations (global transport model)

250 The background CO<sub>2</sub> concentrations ( $y^{bg}$  in Equation 5) result from the transport of CO<sub>2</sub>-loaded air masses from outside the regional inversion domain to the observation sites. One approach to compute these background concentrations has been proposed by Rödenbeck et al. (2009), and consists in extracting background concentrations time series at the observation sites from the model output of a global, coarse resolution Eulerian transport model, driven by a set of inversion-derived CO<sub>2</sub> fluxes.

The background extraction is done in three steps:   
255 a realistic set of prior CO<sub>2</sub> fluxes  $f_{apri}^{glo}$ , an initial atmospheric distribution of CO<sub>2</sub> concentrations ( $C^{imi}$ ), a set of global, background surface CO<sub>2</sub> observations and a subset of the observations to be used later in the regional, high resolution CO<sub>2</sub> inversion. The aim of this step is to obtain a set of CO<sub>2</sub> fluxes  $f^{glo}$  that leads to a very realistic atmospheric CO<sub>2</sub> distribution in and around the regional inversion domain (RID). The accuracy of the fluxes themselves has less importance.   
260 The CO<sub>2</sub> concentrations  $y^{tot}$  corresponding to the transport of the optimized coarse resolution fluxes  $f^{glo}$  to the observation sites within the RID are computed using a forward run of the global transport model used in step 1. The foreground CO<sub>2</sub> concentrations  $y^{fg}$  are computed using a modified version of that same model, in which the fluxes and concentrations are maintained as zero at all times outside the regional domain, so that the concentrations  $y^{fg}$  result only from the transport of the fraction of the fluxes  $f^{glo}$  that is within the RID.   
265 The background CO<sub>2</sub> concentrations are obtained by subtraction of the foreground concentrations to the total ones ( $y^{bg} = y^{tot} - y^{fg}$ ).

265 The underlying assumption is, that, by the time the air masses originating from outside the RID reach the observation sites, existing high resolution patterns of CO<sub>2</sub> at the regional domain boundaries concentrations would have been dispersed, and therefore the field of background CO<sub>2</sub> concentrations within the RID can be well represented with a coarse resolution transport model. On the other hand, this background CO<sub>2</sub> distribution should be as realistic as possible (within the limits of the model resolution), especially in and around the boundaries of the foreground domain, therefore the use of an inversion in step 1 above.   
270 We refer to Rödenbeck et al. (2009) for a much more complete description of the approach.

We implemented the Rödenbeck et al. (2009) approach in a TM5 model setup (Huijnen et al., 2010) with the initial global inversion (step 1) performed in a TM5-4DVAR setup, based on (Basu et al., 2013a)(Basu et al., 2013b). The NEE flux is optimized monthly on a global  $6^\circ \times 4^\circ$  grid, and three additional prescribed CO<sub>2</sub> flux categories are accounted for (fossil fuel, biomass burning and ocean sink). It covers the entire period of the LUMIA inversion, plus six extra months at the beginning   
275 and one at the end to limit the influence of the initial condition and to ensure that the background concentrations in the last month of the LUMIA inversion are well constrained by the observations (observations provide important constraints on the fluxes from the preceding month).

The inversion is constrained by flask observations from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network (Dlugokencky et al., 2019) outside the European domain, and by a subset of the observations used for the regional   
280 inversion within the European domain (see Section 3.3 for references, and [SI-Table S11](#) for a full list of the sites used in that step).

Since the focus of this inversion is to produce a realistic CO<sub>2</sub> distribution around the European domain, the choice of a prior matters a lot less than the selection of observations. For practical reasons, prior fluxes from the CarbonTracker 2016 release were used (Peters et al., 2007): the NEE prior is generated by the SibCASA model (Schaefer et al., 2008); fossil fuel emissions spatially distributed according to the EDGAR4.2 inventory (<https://edgar.jrc.ec.europa.eu/overview.php?v=42>); biomass burning emissions are based on the GFED4.1s product (Van Der Werf et al., 2017) and the ocean flux is based on the Takahashi et al. (2009) climatology. We refer to the official CarbonTracker 2016 documentation (<https://www.esrl.noaa.gov/gmd/ccgg/carbontracker/CT2016>) and to references therein for further documentation on these priors.

The total ( $y^{tot}$ ) and foreground ( $y^{fg}$ ) CO<sub>2</sub> concentration time series at the observation sites are extracted using a modified forward TM5 run implementing the step 2 of the background extraction approach described above. The foreground and total CO<sub>2</sub> time series were saved for each observation site, both as continuous (every 30 minutes) concentration time series, sampled at the actual altitude (above sea level) of the observation site, but also as vertical profiles between the surface (as defined in the TM5 orography) and 5000 m.a.s.l (with a vertical resolution of 250 m and a temporal resolution of 30 minutes). The latter is used to construct a part of the observation uncertainties.

### 295 3.3 Observations and ~~observation~~observational uncertainties

Observations from the GLOBALVIEWplus 4.2 obspack product were used in the inversions (NOAA Carbon Cycle Group ObsPack Team, 2019). For the year 2011, the product includes observations from 26 sites within our regional domain (in addition to observations from mobile platforms, which were not used). Continuous observations are available at 18 of these 26 sites and nine sites are high altitude. Most of these observation sites are now part of the European ICOS network. A list of the sites (coordinates, observation frequency, sampling height and data provider) is provided in Table 1, and the location of the observations is also reported in Figure 2.

#### 3.3.1 Observation uncertainties

The observation uncertainty matrix ( $\mathbf{R}$ ) accounts for both the measurement uncertainties ( $\varepsilon_{obs}$ ) and the model representation uncertainty ( $\varepsilon_H$ , i.e. the incapacity of the model to represent perfectly well the observations, even given perfect fluxes). In theory, the diagonal of the matrix stores the absolute total uncertainty associated to each observation while the off-diagonals should store the observation error correlations. In practice, these correlations are difficult to quantify, and the size of the matrix would anyway make it impractical to invert. The off-diagonals are therefore ignored in our system (as in most similar inversion setups) and the observation uncertainty is stored in a simpler observation error vector,  $\varepsilon_y$ .

Our inversion system uses an observation operator that decomposes the background and foreground components of the CO<sub>2</sub> mixing ratio, therefore the model uncertainty can itself be decomposed in foreground and background uncertainties:

$$\varepsilon_y = \sqrt{\max\{\varepsilon_{obs}, \varepsilon_{obs}^{min}\}^2 + \varepsilon_{bg}^2 + \varepsilon_{fg}^2} \quad (8)$$

The instrumental error ( $\varepsilon_y$ ) is provided by the data providers for most of the observations, and typically ranges between 0.1-0.7 ppm (see Figure 3). We enforced a minimum instrumental error ( $\varepsilon_{obs}^{min}$ ) of 0.3 ppm for all the observations.

315 The model representation error can not be formally quantified, as this would require knowing precisely the CO<sub>2</sub> fluxes that the inversion is attempting to estimate. One can, however, assign representation error estimates (foreground and background) based, in particular, on assumptions of situations that would normally lead to a degradation of the model performances (for instance late-night/early-morning observations, with a development of the boundary layer that may not be well captured by the model, or observations in regions with a complex orography). Transport model comparisons can also provide representation error estimates based on the difference in their results.

### 320 **3.3.2 Foreground model uncertainties:**

As described in Section 3.2.2, ~~background and foreground~~ the TM5 simulation used for computing background CO<sub>2</sub> concentrations ~~are computed for~~, also computes the foreground concentrations at each observation site ~~from a TM5 simulation~~. We performed a forward transport simulation with the regional transport model in LUMIA, using both the background concentrations and the foreground fluxes from that TM5 simulation, so that the two simulations differ only by their regional transport model. A comparison between the concentrations computed by the two models is shown in Figure 4. The bias between the two  
325 models is very contained during the summer months (it is below 0.2 ppm from April to September, and goes as low as 0.01 ppm in July), but rises during the winter months (up to 1.45 ppm in November). The mean average difference between the two simulations is also much larger in winter: it ranges from 0.82 ppm in September to 4.3 ppm in November, with a yearly average of 3.3 ppm.

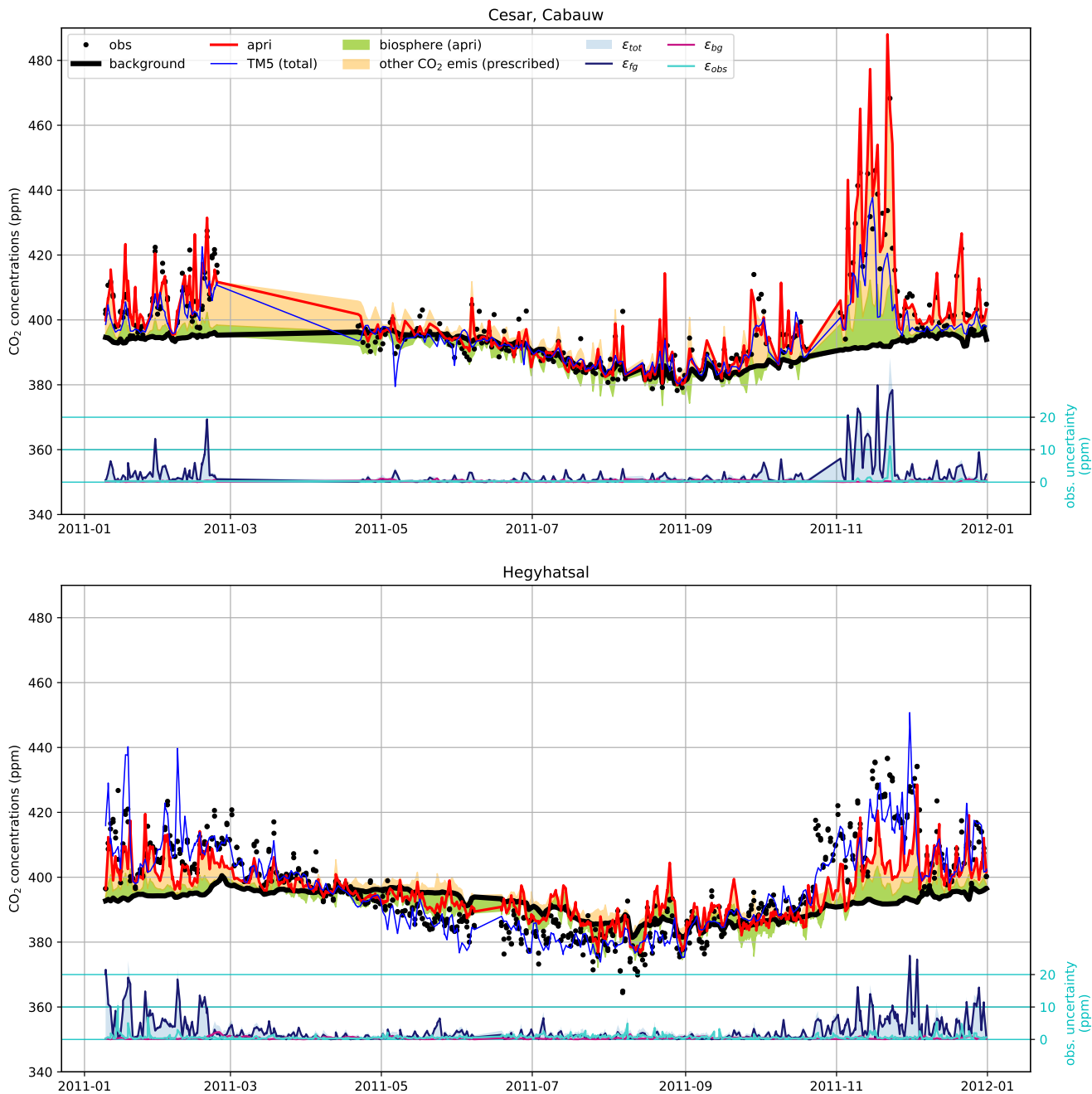
330 This comparison is not a formal performance assessment of either TM5 or of the FLEXPART-based transport used in LUMIA, and in particular the bias should be interpreted with care as the sign of the total net foreground flux changes during the year (which mechanically leads to a change of the sign of the bias). Nonetheless, it provides an indication on the order of magnitude of the foreground model transport errors. We use the absolute differences between the two models as a proxy for  $\varepsilon_{fg}$ .

### 335 **3.3.3 Background model uncertainties:**

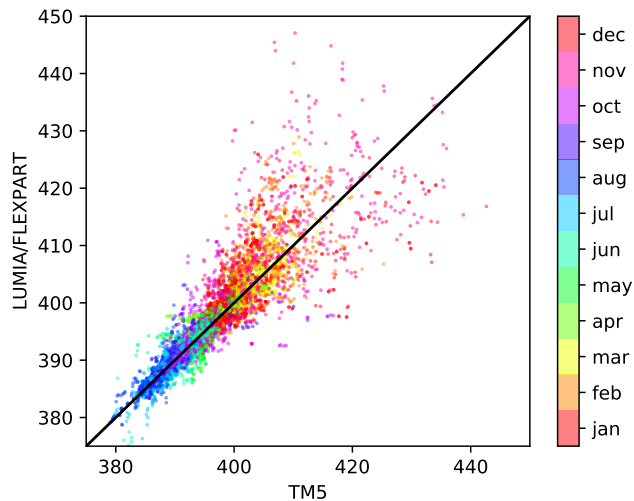
#### **3.3.4 Background model uncertainties**

Background concentrations are expected to be accurately estimated by the global TM5 inversion when the dominant winds are from the West and that any signal from a strong point CO<sub>2</sub> source or sink has had time to dissipate along the air mass trajectory over the Atlantic Ocean. In less favourable conditions, there can be entries of less well-mixed air inside the domain,  
340 in particular in case of Easterly winds or in events of re-entry of continental air that would have previously left the domain. These events are less likely to be well captured by the TM5 inversion and should be attributed a higher uncertainty.

There is no perfect and easy way to detect these events, but one of their consequences would be a less homogeneous background CO<sub>2</sub> distribution around the observation sites when they occur. As part of the TM5 simulation, vertical profiles of



**Figure 3.** Decomposition of the modelled mixing ratio and of the observation uncertainties at two sites (Cabauw, The Netherlands and Hegyhatsal, Hungary). The “TM5 total” line is the concentration computed in the coarse resolution TM5 inversion from which the background (thick black line) is extracted. The LUMIA prior concentration is shown in red and the green and orange shaded areas show respectively the contribution of the prior biosphere flux and of the other CO<sub>2</sub> fluxes to the difference between that prior and the background. The lower series of lines in each plot (with y-axis on the right) shows the total observation uncertainty (blue shaded area), and the contributions of the foreground, background and observational uncertainties.



**Figure 4.** LUMIA (CO<sub>2</sub> concentrations obtained with TM5-FLEXPART vs. with TM5, using the CO<sub>2</sub> fluxes used as prior of the TM5 inversion (background). The color of the dots show the observation month.

background concentrations were stored for each observation (from the surface to 5000 m.a.s.l, at a 250m vertical resolution).  
 345 We set the background uncertainty of each observation ( $\varepsilon_{bg}$ ) to the standard deviation of its corresponding background CO<sub>2</sub> vertical profile.  $\varepsilon_{bg}$  is on average 0.36 ppm, one order of magnitude lower than  $\varepsilon_{fg}$ , and it is also more constant (it ranges between 0.01 and 3.6 ppm). Note that these statistics are computed before the observation selection procedure, described in the following section. The different components of the observation uncertainty are compared in Figure 3 for two representative sites.

### 350 3.3.5 Observation selection

The inversions are performed on a subset of the observations included in the obspack product. Only observations for which the transport model simulation is expected to result in accurate concentrations are kept. In practice, one of the main difficulties of transport modelling is to correctly compute the mixing of air in the lower troposphere below the boundary layer. The lowest model representation error is expected for observations that are either within the boundary layer when it is most developed  
 355 (in the afternoon), or well above the boundary layer for high-altitude sites (during the night). For each site with continuous observations, we selected only observations sampled during the time range for which the model is expected to perform the best. The time ranges are based on the "dataset\_time\_window\_utc" flag in the metadata of the observation files from the obspack. For sites with discrete sampling, all observations were used.

A second filter, used in some of the inversions and in the background TM5 inversion (Section 3.2.2), is the limitation of a  
 360 maximum of one observation per 24 hours at each observation site (the one that has the lowest observation uncertainty, as per the definition in the previous above). This is justified by the fact that two observations at a same site, within a small time interval

have strongly correlated model representation errors and do therefore not provide independent information. In the absence of a proper accounting of observational error correlations, it may be preferable to limit the number of assimilated observations. The use of this second observation filter is discussed in further details in Section 5.

Code	Name	Lat (°E)	Lon (°N)	Alt (m.a.s.l)	Intake Height (m.a.g.l)	Nobs	Time range (h)	sets	Data provider
BAL	Baltic Sea	55.35	17.22	3	25	53	all	P	1
BSC	Black Sea, Constanta	44.18	28.66	0	5	17	all	P	1
CES200	Cesar, Cabauw	51.97	4.93	-1	200	306	11-15	P	2
CIB005	Centro de Investigacion de la Baja Atmosfera (CIBA)	41.81	-4.93	845	5	49	*	A	1
CMN	Mt. Cimone Station	44.18	10.70	2165	12	549	23-3	A	3
CRP	Carnsore Point	52.18	-6.37	9	14	589	12-16	P	4
ELL	Estany Llong	42.57	0.95	2002	3	8	11-15	A	5
HEI	Heidelberg	49.42	8.67	116	30	632	11-15	P	6
HPB054	Hohenpeissenberg	47.80	11.02	936	54	47	all	A	1
HUN115	Hegyhatsal	46.95	16.65	248	115	685	11-15	P	7
JFJ	Jungfrauoch	46.55	7.99	3570	10	461	23-3	A	8
KAS	Kasprowy Wierch	49.23	19.98	1989	5	481	23-3	A	9
LMP005	Lampedusa	35.52	12.62	45	5	35	all	P	1
LMP008	Lampedusa	35.52	12.62	45	8	418	10-14	PA	10
LUT	Lutjewad	53.40	6.35	1	60	289	11-15	P	11
MHD024	Mace Head	53.33	-9.90	5	24	352	12-16	P	12
OPE120	Observatoire Perenne de l'Environnement	48.56	5.50	390	120	405	11-15	P	12
OXK163	Ochsenkopf	50.03	11.81	1022	163	48	all	A	1
PAL	Pallas-Sammaltunturi	67.97	24.12	565	5	654	22-2	PA	13
PRS	Plateau Rosa Station	45.93	7.70	3480	10	445	23-3	A	14
PUI	Puijo	62.91	27.65	232	84	170	11-15	P	13
PUY010	Puy de Dome	45.77	2.97	1465	10	409	23-3	A	12
PUY015	Puy de Dome	45.77	2.97	1465	15	141	23-3	A	12
SSL	Schauinsland	47.92	7.92	1205	12	625	23-3	A	15
TRN180	Trainou	47.96	2.11	131	180	539	11-15	P	12
TTA	Tall Tower Angus	56.56	-2.99	400	222	435	12-16	PA	16
WAO	Weybourne, Norfolk	52.95	1.12	20	10	1078	12-16	P	17
WES	WES	54.93	8.32	12	0	1377	11-15	P	18

**Table 1.** Observation sites used in the inversions. The Data providers: 1:NOAA Carbon Cycle Group ObsPack Team (2019); 2:Vermeulen et al. (2011); 3:Ciattaglia et al. (1987); 4:D. Dodd (EPA Ireland); 5:J.A. Morgui and R. Curcoll (ICTA-UAB, Spain); 6:Hammer et al. (2008); 7:Haszpra et al. (2001); 8:Uglietti et al. (2011); 9:Rozanski et al. (2014); 10:A. G. di Sarra (ENEA, Italy); 11:van der Laan et al. (2009); 12:Yver et al. (2011); 13:Hatakka et al. (2003); 14:F. Apadula (RSE, Italy); 15:Schmidt (2003); 16:Ganesan et al. (2015); 17:Wilson (2012); 18:K. Uhse (UBA, Germany)

### 365 3.4 Prior and prescribed fluxes

In addition of the Net Ecosystem Exchange (NEE, net atmosphere-land CO<sub>2</sub> flux) that is optimized in the inversions, the simulations also account for anthropogenic CO<sub>2</sub> emissions (combustion of fossil fuels, bio fuels and cement production), for biomass burning emissions (large scale forest fires) and for the ocean-atmosphere CO<sub>2</sub> exchanges.

The NEE prior is taken from simulations of the LPJ-GUESS and ORCHIDEE vegetation models: in the OSSEs (Section 4)  
370 ORCHIDEE fluxes are used as prior and LPJ-GUESS fluxes are used as truth, while in inversions against real data LPJ-GUESS fluxes are used as prior. Both vegetation models provide 3-hourly fluxes, on a horizontal 0.5° × 0.5° grid.

LPJ-GUESS (Smith et al., 2014) is a dynamic global vegetation model (DGVM), which combines process-based descriptions of terrestrial ecosystem structure (vegetation composition, biomass and height) and function (energy absorption, carbon and nitrogen cycling). The vegetation is simulated as a series of replicate patches, in which individuals of each simulated plant  
375 functional type (or species) compete for the available resources of light and water, as prescribed by the climate data. The model is forced using the WFDEI meteorological data set (Weedon et al., 2014) and produces 3-hourly output of gross and net carbon fluxes.

ORCHIDEE is a global processed-based terrestrial biosphere model (initially described in Krinner et al. (2005)) that computes carbon, water and energy fluxes between the land surface and the atmosphere and within the soil-plant continuum. The  
380 model computes the Gross Primary Productivity with the assimilation of carbon based on the Farquhar et al. (1980) for C3 plants and thus account for the response of vegetation growth to increasing atmospheric CO<sub>2</sub> levels and to climate variability. The land cover change (including deforestation, regrowth and cropland dynamic) were prescribed using annual land cover maps derived from the Harmonized land use data set (Hurtt et al., 2011) combined with the the ESA-CCI land cover products. The net and gross CO<sub>2</sub> fluxes used for this project correspond to the one provided for Global Carbon Project inter-comparison  
385 (Le Quéré et al., 2018) with a model version that was updated recently (Peylin et al., in preparation).

Fossil fuel emissions are based on a pre-release of the EDGARv4.3 inventory for the base year 2010 (Janssens-Maenhout et al., 2019). This specific dataset includes additional information on the fuel mix per emission sector and thus allows for a temporal scaling of the gridded annual emissions for the inversion year (2011) according to year-to-year changes of fuel consumption data at national level (bp2, 2016), following the approach of Steinbach et al. (2011). A further temporal disaggregation into hourly emissions is based on specific temporal factors (seasonal, weekly, and daily cycles) for different emission  
390 sectors (?)([Denier van der Gon et al., 2011](#)).

The ocean-atmosphere flux is taken from the Jena CarboScope v1.5 product, which provides temporally and spatially resolved estimates of the global sea-air CO<sub>2</sub> flux, estimated by fitting a simple data-driven diagnostic model of ocean mixed-layer biogeochemistry to surface-ocean CO<sub>2</sub> partial pressure data from the SOCAT v1.5 database (Rödenbeck et al., 2013).

395 A biomass burning flux category was also included in the inversion, based on fluxes from the Global Fire Emission Database v4 (Giglio et al., 2013). In our European domain biomass burning emissions are negligible regarding the other CO<sub>2</sub> emission sources, however, we include it for completeness.

Category	Product	Original resolution	Data provider	Total (min/max) flux (PgC/year)
Biosphere	LPJ-GUESS	0.5° × 0.5°; 3-hourly	Lund University	-0.33 (-2.65 / 1.85)
Biosphere	ORCHIDEE	0.5° × 0.5°; 3-hourly	LSCE (P. Peylin, pers. comm)	-0.28 (-3.73 / 2.14)
Fossil	EDGARv4.3	0.1° × 0.1°; hourly	ICOS-CP + JRC	1.53
Ocean	CarboScopev (oc_v1.7)	5° × 3.83°; daily	Rödenbeck et al. (2013)	-0.11 (-0.05 / 0.01)
Fires	GFEDv4	0.5° × 0.5°; 3-hourly	Van Der Werf et al. (2017)	0.01

**Table 2.** Prior and prescribed CO<sub>2</sub> fluxes. Min/Max values are provided for the fluxes that have both positive and negative components, and correspond to the minimum and maximum values of the 3-hourly flux aggregated over the entire domain, in PgC/year.

All fluxes are regridded on the same 0.5° × 0.5°, 3-hourly resolution (by simple aggregation or re-binning). A summary of the prior fluxes sources, original resolution and yearly totals is provided in Table 2.

### 400 3.4.1 Prior uncertainties

The background error covariance matrix (**B** in Equation 2) is constructed following the “correlation length” approach used in many other inversion systems (e.g. Houweling et al. (2014); Thompson et al. (2015); Chevallier et al. (2005)): The error covariance between fluxes  $\mathbf{x}_1$  and  $\mathbf{x}_2$  at grideells grid cells with coordinates  $p_1 = (i1, j1, t1)$  and  $p_2 = (i2, j2, t2)$  is defined as:

$$405 \text{ cov}(\mathbf{x}_1, \mathbf{x}_2) = \sigma_{\mathbf{x}_1}^2 \cdot \sigma_{\mathbf{x}_2}^2 \cdot \underline{C_h(p_1, p_2)} \cdot \underline{C_t(p_1, p_2)} \underline{e^{-(d(p_1, p_2)/L_h)^2}} \underline{e^{-|t_1 - t_2|/L_t}} \quad (9)$$

where  $\sigma_{\mathbf{x}_1}^2$  and  $\sigma_{\mathbf{x}_2}^2$  are the variances assigned to the prior monthly NEE at coordinates  $p_1$  and  $p_2$ , and  $C_h$  and  $C_t$  are decay functions of the spatial ( $C_h$ ) and temporal ( $C_t$ ) distance between  $p_1$  and  $p_2$ , defined respectively as  $C_t(p_1, p_2) = e^{-|t_1 - t_2|/L_t}$  and  $C_h = e^{-\left(\frac{d(p_1, p_2)}{L_h}\right)^\beta}$ , with  $d(p_1, p_2)$  the geographic distance between the center of the two pixels.  $L_h$  and  $L_t$  are temporal and spatial correlation length. The decay functions are further referred as Gaussian or exponential decay functions, depending on the value of the  $\beta$  parameter (1 for exponential, 2 for Gaussian). We explore the two versions of decay functions in our OSSEs described in Section 4 are correlation lengths, which define how rapidly the correlation between two components drops as a function of their distance in time and space.

The true uncertainty of the prior fluxes ( $\sigma_{\mathbf{x}}^2$ ) is difficult to evaluate and is therefore constructed on reasonable but arbitrary assumptions. We assumed that the flux uncertainties scale linearly with the net flux, and therefore to set the individual flux tested several approaches, further discussed in Section 4: 1) Scaling the uncertainties linearly to a constant fraction of the absolute next monthly flux; 2) Scaling the uncertainties to the absolute net flux, possibly within a range. 3-hourly flux (and then cumulating these 3-hourly uncertainties to the monthly scale); 3) Enforcing constant monthly uncertainties throughout the year, at the domain-scale.



This approach is practical as it doesn't require any additional inputs, and justified by the fact that the uncertainties should be low in areas where the flux is known to be low (for instance, desert areas in a CO<sub>2</sub> inversion). One drawback however is that the net CO<sub>2</sub> flux (NEE) may be close to zero as a result from the gross fluxes (GPP and respiration) of same magnitude. An alternative approach is setting the uncertainty as a scaling factor of the vegetation respiration. We did not explore this here because of a lack of a respiration field at the beginning of the study.

### 3.5 Inversions performed

We performed two ensembles of inversions. The first ensemble consists of Observing System Simulation Experiments (OSSEs) to assess the theoretical performance of the system. Here, the LPJ-GUESS NEE dataset was taken as an arbitrary truth, and a dataset of synthetic pseudo-observations was generated at times and locations of the actual observations listed in Table 1, by forward propagation of the "true" NEE flux with the transport model (including also the contributions of non-optimized fluxes listed in Section 3.4). Random perturbations were then added, to mimic the measurement error ( $y = y^{truth} + \mathcal{N}(0, \sigma_y^2)$ , with  $\sigma_y^2$  the uncertainty of each observation as defined in the matrix **R**).

The OSSEs use this set of pseudo-observations as observational constraint and the ORCHIDEE NEE dataset as a prior. The reference OSSE, SRef, uses a prior error covariance matrix (**B**) constructed with prior uncertainties set to 25% of the absolute prior value ( $\sigma_{x_b}^2 = 0.25|x_b|$ ) and with covariances constructed from a horizontal correlation length ( $L_H$ ) of 200 km and a temporal correlation length ( $L_t$ ) of 30 days. In the sensitivity tests we vary the correlation lengths (SC.100 and SC.500), the prescribed prior uncertainties (SE.3H, SE.3Hcst, SE.x2) and the extent of the observation network (SO.A, SO.P).

## 4 ~~OSSEs~~

The second ensemble is essentially identical to the OSSE ensemble of inversions, except that it is using real observations, and the LPJ-GUESS flux dataset as a prior. The details of the two ensembles of inversions are listed in Table 3

## 4 OSSEs

A series of OSSEs was performed aiming primarily at verifying the absence of abnormal behaviour of the inversion system (i.e. verify that it converges towards the truth and responds as expected to changes in observational or prior constraints). The OSSEs share a common prior NEE (ORCHIDEE) and a common set of observations derived from a forward propagation of the "true" (We first analyze the capacity of SRef to reconstruct various characteristics of the "true" LPJ-GUESS) fluxes (at the same time and location as the actual observations listed in Table 1), perturbed by a random error ( $y = y^{truth} + \mathcal{N}(0, \sigma_y^2)$ , with  $\sigma_y^2$  the uncertainty of each observation as defined in the matrix **R**) NEE fluxes (monthly and annual NEE budget, aggregated at spatial scales ranging from the entire domain down to single pixels). Then we use a series of sensitivity experiments to verify how sensitive the results are to a range of reasonable assumptions in the inversion settings.

Simulation	Prior	Observations	$\sigma_x^2$	$L_h$	$L_t$
SRef	OR-CHIDEE	<del>Truth</del> + $N(\sigma_y^2)$ <del>pseudo</del>	25% of monthly prior	200 km ( <del>exp</del> )	1 month
<del>SB-SC.500</del>	-	<del>Truth</del> + R + B <del>~</del>	-	<del>200-500</del> km (exp)	-
<del>S500</del> <del>SC.100</del>	-	<del>Truth</del> + R <del>~</del>	-	<del>500-100</del> km (exp)	-
<del>S100-SE.x2</del>	-	<del>Truth</del> + R <del>~</del>	<del>100 km (exp)</del> <del>50% of monthly prior</del>	<del>200 km</del>	-
<del>SG-SE.3H</del>	-	<del>Truth</del> + R <del>~</del>	<del>12.5% of 3-hourly prior</del>	200 km ( <del>gauss</del> )	-
<del>SP-SE.3Hcst</del>	-	<del>Truth</del> <del>~</del>	12.5% of 3-hourly prior, scaled to the same total value every month	<del>~</del>	<del>~</del>
<del>SO.P</del>	<del>~</del>	<del>pseudo</del> (set P) + R <del>~</del>	-	200 km ( <del>exp</del> )	-
<del>SA-SO.A</del>	-	<del>Truth</del> <del>pseudo</del> (set A) + R <del>~</del>	-	200 km ( <del>exp</del> )	-
RRef	LPJ-GUESS	<del>Real</del> <del>real</del>	25% of monthly prior	200 km ( <del>gauss</del> )	1 month
<del>R500</del> <del>RC.500</del>	-	<del>Real</del> (set P) <del>~</del>	-	500 km ( <del>gauss</del> <del>exp</del> )	-
<del>R100</del> <del>RC.100</del>	-	<del>Real</del> (max 1 obs/24h/site) <del>~</del>	-	100 km ( <del>gauss</del> <del>exp</del> )	-
<del>RP-RE.x2</del>	-	<del>Real</del> (set A) <del>~</del>	<del>50% of monthly prior</del>	<del>200 km</del>	-
<del>RE.3H</del>	<del>~</del>	<del>~</del>	<del>12.5% of 3-hourly prior</del>	200 km ( <del>gauss</del> )	-
<del>RA</del> <del>RE.3Hcst</del>	-	<del>Real</del> <del>~</del>	12.5% of 3-hourly prior, scaled to the same total value every month	<del>~</del>	<del>~</del>
<del>RO.P</del>	<del>~</del>	<del>real</del> (set P)	-	200 km ( <del>gauss</del> )	-
<del>RIf-RO.A</del>	-	<del>Real</del> <del>real</del> (set A)	-	200 km ( <del>gauss</del> )	<del>~</del>

**Table 3.** List of inversion experiments performed. The R and B letters in the Observations column refer respectively to a random error perturbation (R, proportional to the assigned individual uncertainty of each observation) and to a systematic bias (B), described in the main text of Section 4. The restricted observation sets A and P are reported in Table 1.

## 4.1 Reference inversion (SRef)

450 Figure 5 shows monthly and annual time series of NEE and NEE error (with respect to the prescribed truth) aggregated over the entire domain.

In the reference inversion Sref, at the domain-scale, the prior error covariance matrix ( $\mathbf{B}$ ) is constructed with prior uncertainties set to 25% of the absolute prior value ( $\sigma_{x_b}^2 = 0.25|x_b|$ ) and exponentially decaying horizontal and temporal covariances, with a horizontal covariance length ( $L_h$ ) of 200 km and a temporal covariance length ( $L_t$ ) of one month.

455 Inversions S500 and S100 differ from Sref by the use of respectively longer (500 km) and shorter (100 km) horizontal covariance lengths. Inversion SG uses a Gaussian decay function instead of the exponential one used in Sref. Inversions SP and SA use respectively only the low altitude and only the high altitude sites (See Table 1). Inversion SB is similar to Sref but uses a perturbed set of background CO<sub>2</sub> concentrations:  $y_{pert}^{bg} = 0.75y_{truth}^{bg} + 0.25y_{prior}^{bg}$ , with  $y_{prior}^{bg}$  the background concentrations corresponding to the estimate for the annual NEE is very close to the prior fluxes of the coarse resolution TM5 inversion from which  $y_{truth}^{bg}$  are generated (See Section 3.2.2).

## 460 4.2 Domain-wide NEE budget

The net annual budget is one of the first quantities typically looked at in an atmospheric inversion. The net annual NEE optimized in the six OSSEs are compared to the prior and true NEE in “truth” (respectively -0.28 and -0.34 PgC/year), but the amplitude in the upper panel of Figure ???. At this annual scale, all the inversions degrade the flux estimate. Most inversions point to a near zero annual flux (from -0.04 PgC prior is more than double that of the truth, with monthly NEE ranging from +0.26 PgC/year in S100 to -0.1 PgC in SP and S500. SB is the closest to month in October to -0.66 PgC/month in June in the prior, compared to +0.08 PgC/month (in January) to -0.29 PgC/month (in May) in the truth. In total, the absolute prior error slightly exceeds 3 PgC and peaks in June and July and is the lowest in December-February.

470 The inversion improves the estimation of the seasonal cycle at the domain scale, with a seasonal cycle amplitude reduced to a range of -0.36 PgC (May) to +0.16 PgC (December), much closer to the truth and the ‘true’ flux, despite being constrained by biased background concentrations, with annual estimate of -0.4 absolute error is reduced by nearly 40% to 1.87 PgC/year (-20% larger than the -0.33 year. However since the positive flux corrections in the summer months largely exceed the negative corrections from September to April, this results in a strong degradation of the annual European NEE estimate, with a near-balanced posterior flux of -0.05 PgC/year of the synthetic truth).

On the other hand, the lower panel of Figure ??? compares the root mean square error of the inversions to that of the prior: 475 all the inversions lead to a reduction of the RMSE by  $\approx 30\%$ . Inversion S100 performs marginally better, and SP and SG marginally worse.

480 These contradictory analyses can be explained by looking at the fluxes at the monthly scale, shown in Figure 6 illustrates the spatial distribution of the upper error reduction. While the largest prior errors are found north of the Black Sea and in North-Africa, the error reduction is rather homogeneous, except for North-Africa and Turkey (which are not really constrained by the observation network), and some patches in Western Europe (mainly in the UK, but also in Ireland, France and the

Benelux) where the error actually increases. In total, these localized error enhancements amount to 0.16 PgC (lower panel of Figure ??). The prior NEE has a much more pronounced seasonal cycle than our assumed truth. The summer uptake is, for instance, almost three times larger in the prior and the winter release of carbon to the atmosphere is also much stronger than in the 'true' fluxes. 5). These isolated occurrences of error enhancements are not a sign of malfunction of the inversion system, but they highlight its limitations: they result from attributions of flux corrections to the incorrect grid cells, which can happen if the resolution of the inversion is not adapted to the constraints provided by the observation network (i.e. smoothing and aggregation errors, as defined in Turner and Jacob (2015)).

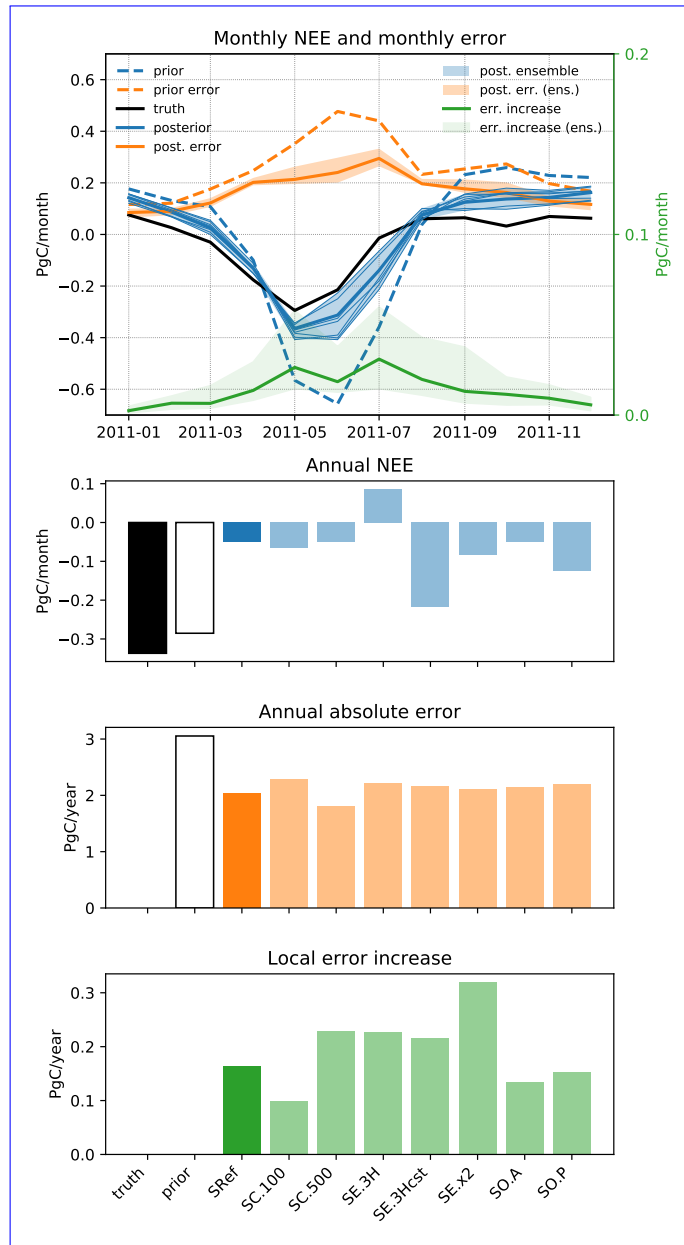
At the domain scale, all the inversions are able to bring the prior seasonal cycle in agreement with the one imprinted in the pseudo-observations, although performances differ slightly. Most of the time, inversion S500 leads to the most accurate representation of the true NEE, with a net error in a  $\pm 0.05$  PgC/month range. By comparison, inversion SB leads to similar performances in winter (August to April), but does not achieve the same error reduction in summer, with a net monthly summer error in a -0.11 to -0.15 PgC/month range from May to July. The better performance of inversion S500 in summer is explained by errors of opposite sign in summer and winter, compensating each other. More generally, the differences between the optimized annual budgets shown in Figure ?? are explained by errors of opposite sign, more or less compensating each other, and are not a good indication of the inversions performance. Although our control vector contains the flux estimates at the native spatial resolution of the transport model, the effective resolution of the inversion SB in estimating the annual budget of NEE is explained by errors of opposite sign in summer and winter, compensating each other. More generally, the differences between the optimized annual budgets shown in Figure ?? are explained by errors of opposite sign, more or less compensating each other, and are not a good indication of the inversions performance. Furthermore, the fluxes are only optimized monthly, while the actual prior error varies at a 3-hourly resolution. It may therefore be possible to reduce these errors by increasing the resolution of the inversion, but since the observation network is not homogeneous, this may lead to increased posterior error in other parts of the domain.

#### 4.1.1 Regional decomposition

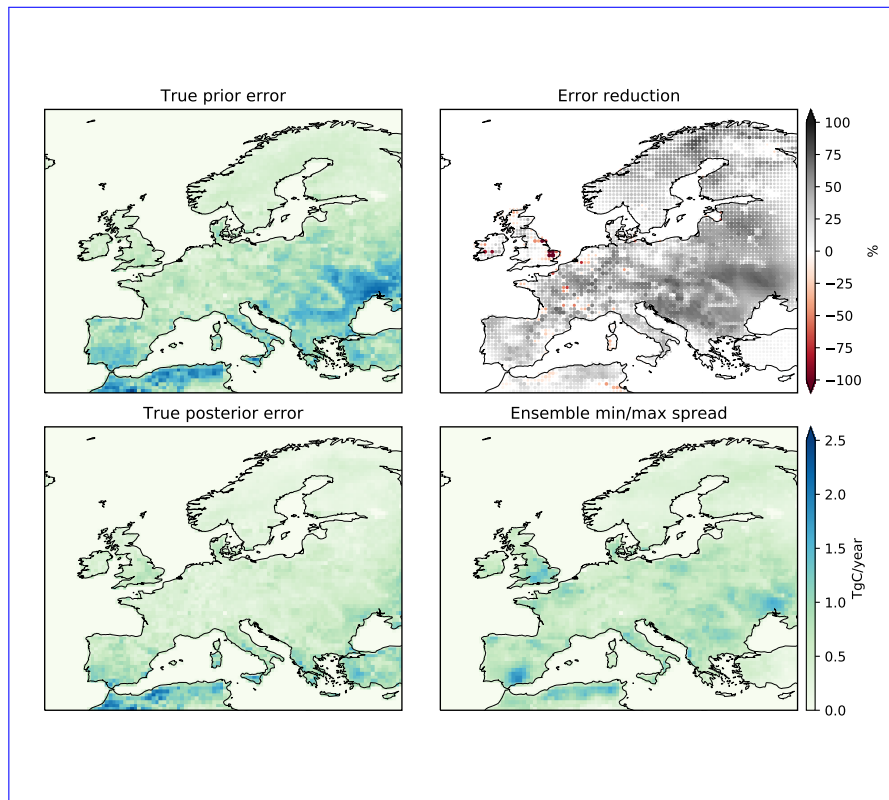
#### 4.2 Sensitivity tests

Aggregated monthly and annual regional totals are shown for all the inversions in Figure ?? (with the region definitions provided in SI). The bar plots on the left show the prior, true and posterior annual NEE budgets in each region, and the plots on the right show the monthly prior and posterior differences with the "true" regional NEE. We performed a series of sensitivity tests to assess the robustness of the results. The sensitivity experiments can be grouped in three series: First, SC.100 and SC.500 test the sensitivity to the spatial covariance lengths used to construct B. Then, SE.3H, SE.3Hcst and SE.x2 test the sensitivity of the results to the prior uncertainty themselves (i.e. diagonal elements of B). Finally, SO.A and SO.P test the sensitivity of the results to the network density.

As it is the case at the domain scale, the monthly regional NEE budgets are systematically improved by the inversions. This sometimes leads to degradation of the representation of the annual budget (in the Northern, Southern and Eastern Europe regions in particular), which confirms that the annual budget is not robustly constrained by our system, contrarily to the monthly NEE.



**Figure 5.** Top Upper row, left axis: Monthly prior NEE (dashed black-blue line), true NEE (thick solid black line) and posterior NEE aggregated on the whole inversion domain; Middle: Annual NEE budgets (truth blue), absolute prior error (dashed orange line) and posterior error (orange); Bottom Upper row, right axis: RMSE Total error increase (with respects to i.e. positive component of the arbitrary truth error reduction, green). The SRef inversion is shown as solid lines, the ensemble is shown as a shaded area. Second, third and fourth rows: same variables, but aggregated annually.



**Figure 6.** Left column: prior (top) and posterior (bottom) total error (with respect to the prescribed truth); Upper-right: percentage error reduction. The size of the dots is proportional to the size of the flux correction; Bottom-right: amplitude of the posterior ensemble spread.

515 ~~The largest discrepancies are found in the Eastern, Northern and South-Eastern Europe regions, where the observation network is the densest. On the other hand, the flux estimate is, at least at the monthly scale, very insensitive to~~ The total NEE flux, absolute error and error increase are shown in Figure 5, for the individual sensitivity experiments at the annual scale and as an ensemble shape for the monthly scale (the monthly-scale results of the individual simulations can be found in Figure S12).

#### 4.2.1 Sensitivity to the error distribution

520 Inversions SE.3H, SE.3Hcst and SE.x2 were designed to test the impact of the prescribed prior uncertainty vector (e.g. diagonal of **B**) on the inversion:

- In SE.3H, the prior uncertainty is set proportional to the sum of the uncertainties on the 3-hourly fluxes:  $\sigma_{x_b} = \frac{0.13}{P} \sum_t^T |f_t|$ . This avoids the situation where GPP and respiration are significant but compensate each other, leading to a near zero NEE as well as a near zero prior uncertainty, which can happen when the prior uncertainty is calculated following the approach used in SRef. The factor 0.13 was chosen to lead to a total annual uncertainty comparable to that of SRef. This

525

leads to an overall redistribution of the uncertainties from the winter to the summer period, which is closer to the actual distribution of differences between the prior and truth fluxes (see Figure S11).

- In SE.3Hcst, the different inversion settings in Central Europe. Reductions in the network density only leads to marginal prior uncertainty is computed as in SE.3H, but it is then scaled monthly, so as to lead to a flat distribution of the uncertainties across the year.
- In SE.x2, the prior uncertainty is simply doubled compared to SRef.

SE.3Hcst leads to an improved value of the annual budget of NEE at the domain scale, but this is due to a poorer estimation of the summer fluxes (since the uncertainty is lower in summer, the inversion sticks more to its prior). On the contrary, SE.3H leads to further degradation of the monthly regional NEE (at least in these large regions)-

The use of a biased set of background concentrations in inversion SB affects the Western Europe region more than any other: background air generally enters the domain from the West, annual budget, without achieving better performances than SRef at the monthly scale. For both inversions, this translates into a slightly larger total posterior error (2.15 and Western Europe fluxes are then used as a buffer to compensate for this biased background 2.20 PgC/year, respectively, compared to 2.03 in SRef). The doubling of the prior uncertainty in SE.x2 allows it to depart more from the prior and to derive better domain-scale flux estimates, both monthly and annually, but it also leads to an increase in the “added error” (lower panel of Figure 5).

Inversion S500 behaves noticeably different from the other inversions in Northern, Eastern and South-Eastern Europe (the three regions that are the least constrained by observations) : Inversion S500 uses longer spatial covariance lengths-

#### 4.2.2 Sensitivity to the error covariance structure

Inversion SC.100 and SC.500 use prior error covariance matrices constructed using respectively shorter (100 km) and longer (500 km) .Therefore, the flux adjustments will be spread over a larger area than in e.g. Sref (which uses a B matrix constructed by using 200 km covariance lengths) horizontal correlation lengths ( $L_H$ ) than SRef. The longer covariance length in SC.500 forces the inversion to favour large-scale, low amplitude flux corrections over localized strong adjustments. Since the prior error (i.e. the difference between ORCHIDEE and LPJ-GUESS) follows a relatively homogeneous pattern, S500-SC.500 effectively produces a better estimation of the NEE, especially in Eastern Europe .There is, however, no guarantee that such a conclusion would apply to an inversion constrained by real observations, as the difference between the prior and the real NEE fluxes may not follow such a simple pattern.-

This point is further illustrated in Figure ??, which shows maps of prior, posterior (Sref) and true fluxes, and the associated errors and error corrections, for the four seasons. Flux error reductions are more consistently obtained in Eastern and Southern Europe, where the network is the least dense. One straightforward explanation is that the prior errors are much larger in these regions and therefore easier to reduce. But in Western and Central Europe, where the observation network is dense and where neighbour sites occasionally provide contradictory signals, the inversions have to apply more complex corrections to the control vector, which is more difficult when the covariances lengths are long (i.e. in S500) and occasionally leads to locally increased error.-

560 ~~Prior, true and posterior annual regional NEE balance (bar plots) and monthly regional NEE error in the prior and posterior fluxes (line plots) where the network is sparse (Figure SI4b). The opposite happens with SC.100, which tends to concentrate the flux adjustments in the vicinity of the observation sites.~~

565 ~~From top to bottom: True fluxes (LPJ-GUESS), prior fluxes (ORCHIDEE), posterior fluxes (inversion S2), relative prior error (in percentage of the true fluxes), relative posterior error and relative error reduction. Each column corresponds to three months (sums for the fluxes and averages for the error and error reductions). The cyan dots show the location of the observation sites. At the domain scale, the annual budgets are nearly identical in SC.100, SC.500 and SRef. However the total error reduction is lower in SC.100 and higher in SC.500, compared to SRef (respectively 0.78, 1.28 and 1.02 PgC/year), but the “added error” is larger in SC.500 (0.23 PgC/year) and lower in SC.100 (0.10 PgC/year): this confirms the hypothesis that these are aggregation errors, that can be reduced by increasing the number of degrees of freedom in the inversion (for instance by reducing the covariance constraints).~~

### 570 4.2.3 Sensitivity to the observation network density

575 ~~Compared to SRef, SO.A uses only high-altitude observations (plus LMP and TTA as these were the only sites available in their region) and SO.P uses only low altitude sites. In terms of annual budget, SO.P outperforms most of the other inversions, but as for SE.3Hcst, this results from poorer flux corrections in summer rather than from a better overall reduction of the uncertainties. On the contrary SO.A leads to results very comparable to SRef at the domain scale, with a nearly identical seasonal cycle and net annual flux. The net error reduction remains however slightly better in SRef (see also Figure SI1 for the seasonal cycles of SO.A and SO.P).~~

### 4.3 Evolution of the fit to the observations

580 The comparison of the prior and posterior model fit to observations is a classical diagnostics of atmospheric inversions (Michalak et al., 2017). The inversion is expected to improve the overall fit to the observation ensemble, and a lack of statistical improvement would generally be a sign of a malfunctioning inversion algorithm. At a finer scale, analysis of the when and where the representation of the observations is most improved (or degraded), can provide useful insights on the performances of the inversion (adequacy of the definition of uncertainties) and on those of the underlying transport model.

585 In the right panel of Figure 7, we compare the statistical distribution of prior and posterior observation fit residuals for inversion Sref. The plot confirms that the inversion leads to an overall improvement of the representation of observations, albeit modest (prior bias (model-obs): 0.2 ppm; posterior bias: 0.05 ppm; prior RMSE: 4.9 ppm; posterior RMSE: 3.75 ppm).

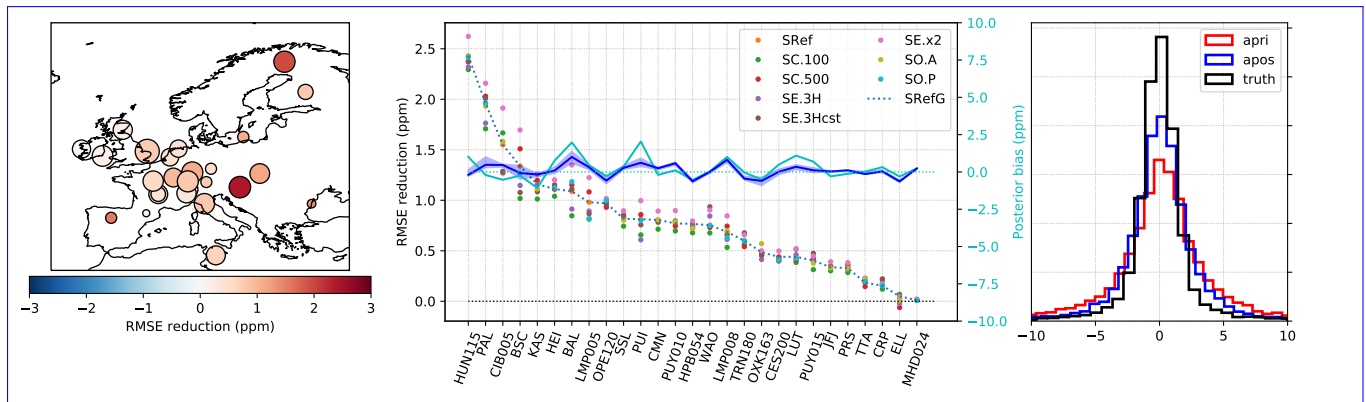
The left panel shows the RMSE reduction at each observation site ~~for inversion Sref~~ (the size of the dots is proportional to the number of assimilated observations at each site, and the color shows the net RMSE reduction). At all sites the inversion leads to improvements in the fit, but those are generally much more modest in Western Europe.

590 ~~Finally, the center panel compares the RMSE reduction of inversion Sref to that of the other OSSEs: The best performance is logically achieved by inversion S1 at most of the sites because it is constrained by exact concentrations, and therefore it is easier for the inversion to find a solution. But overall the inversion performances are very comparable.~~



The comparatively low misfits in continental Europe, which can be explained by the (coincidental) good performance of the prior in that region (see previous section and Figure ?? Figure 6), but also by the strong sensitivity of these sites to background concentrations. Sites in the UK and, in particular, Ireland sample very little continental air, which leaves little margin for the inversion to improve the representation of their observations.

The center panel of Figure 7 compares the RMSE reduction of inversion SRef to that of the other OSSEs. The best performances are logically achieved by SE.x2, which can depart much more from its prior than the other inversions. On the other hand, SC.100 systematically underperforms the ensemble, which is coherent with its poorer flux error reduction. In general however, the reduction of misfits are very similar and are not good indicators for the quality of the optimized fluxes.



**Figure 7.** Left: Map of the observation sites in Sref, with the area of the dots proportional to the number of assimilated observation at each site, and the color proportional to the RMSE reduction (prior RMSE minus posterior RMSE). Center: RMSE reduction at each site, for the five sensitivity OSSEs. Right: Prior and posterior distribution of the observation mismatches in Sref (irrespective of residuals with the site prior, posterior (SRef) and truth fluxes). The cyan and blue line plots in the center plot show the prior (blue) and posterior (cyan) mean biases at each site (right axis)

## 600 5 Inversions against with real observations

The OSSEs presented above neglect several complications of real inversions, in particular transport model errors (the observations were generated using the same transport model than as the one used in the inversions). Quantifying precisely the impact of such transport model errors is not While it is not within the scope of this paper. Nonetheless, we present results from a set of inversions similar to the OSSEs of Section 4 but to quantify precisely these errors, we nonetheless performed a series of inversions constrained by real observations. The objectives are to verify the stability of the solution with regards to typical user choices, such as the selection of observations and the definition of prior uncertainties, and to provide some indicators of the transport model performance, to assess to which extent the characteristics of the inversions results identified with the ensemble of OSSEs remain under a more realistic situation.

The reference inversion (Rref) is constrained by the observations presented in Section 3.3 and uses a prior terrestrial flux from ensemble of inversions used here is identical to the OSSEs ensemble, except that real observations are used and that the LPJ-GUESS model described in Section 3.4. The prior uncertainties ( $\sigma_{x_b}^2$ ) are set to 25% of the net monthly flux in each pixel, and gaussian spatial correlation length of 200 km and exponential temporal correlation lengths of one month are used. Besides Rref, two groups of inversions were performed (with the specific settings listed flux is used as a prior (instead of ORCHIDEE in the OSSEs). The inversion settings are reported in Table 3):-

615 – The sensitivity of the fluxes to the observation selection is tested in experiments RP, Rlf and RA. In RP, high altitude sites (see Table 1) have been excluded from the inversion. In Rlf the same observation network as in Rref has been used, but the frequency of observations has been limited to a maximum of one for each day and site. Finally, in RA only high altitude sites plus LMP, PAL and TTA have been used (because of the lack of any altitude site in the regions of these three sites):-

620 – Inversions R500 and R100 were designed to assess the impact of the structure of the prior covariance matrix on the results. In inversions R500 and R100, the spatial correlation length used to build  $\mathbf{B}$  was changed to 500 km and 100 km

As for the OSSEs, we analyze first the optimized fluxes and then the evolution of the fits to the observations. Note that this time the true fluxes are not known.

## 5.1 Optimized fluxes

### 625 5.1 Posterior fluxes

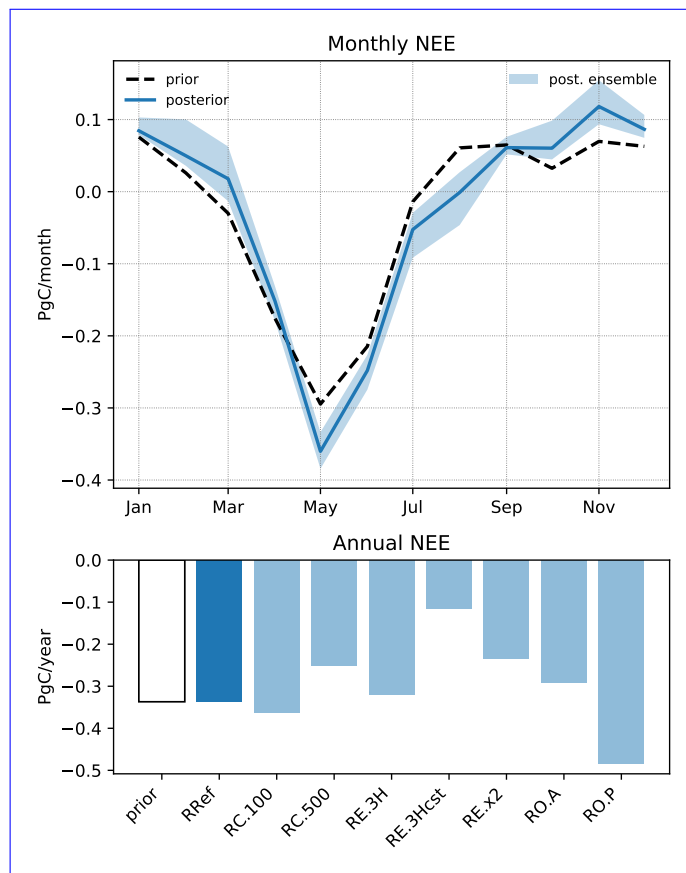
Optimized fluxes are shown in Figure ??, as monthly (left) and annual totals (right), aggregated on the entire domain (first row) and in smaller regions (following the same region definitions used for the OSSEs in Section ??). At the domain scale, the number of observations has the largest impact on the posterior flux. While the three inversions that use the full dataset (Rref, R500, R100) find a 50 to 60% reduction of the net NEE sink, inversions RP (which uses only low altitude sites) and Rlf (which uses the same number of sites but a reduced number of observations) lead only to a 25% reduction. The monthly and annual prior and posterior NEE are shown in Figure 8, for the reference RRef inversion and for the sensitivity tests. The inversion leads to a slight increase in the seasonal cycle amplitude, with a peak summer uptake increased by 24% in May (-0.36 PgC/month, instead of -0.24 PgC/month in the prior) and a nearly doubling of the CO<sub>2</sub> emissions in winter (+0.12 PgC/month instead of +0.07 PgC/month in the prior in November). It also leads to a delayed date for the change of sign of the net sink, and RA (altitude sites only) does not adjust the total flux significantly. Focusing on the monthly totals, the inversions consistently point to a larger seasonal cycle than simulated by the LPJ-GUESS model, with a more positive NEE in the winter and autumn months (especially in March, September, October and November), and a stronger carbon uptake in the European summer months. Especially the near-zero NEE in August is clearly contradicted by all the inversions. flux, both in the spring and in the autumn (the net prior flux becomes negative in March in the prior, and positive again in August, while it only becomes negative in April, and positive in October in RRef).

The regional decomposition shows a more complex picture. The increased early summer (May-July) uptake is largely attributed to three regions: Northern, South-Eastern and Central Europe, and the August shift towards negative NEE is almost entirely due to adjustments in the Southern and Northern Europe regions. On the contrary, the NEE is corrected towards more positive values (reduced uptake in summer and increased release in winter) most of the year (except in May) in Western Europe (and more prominently so in the summer and fall months). In fall and winter CO<sub>2</sub> fluxes to the atmosphere are increased in most of the regions, except in Northern Europe where it is reduced in the fall and unchanged in winter (January to April). These monthly flux adjustments do not result in a change in the net annual flux (-0.33 PgC/year, both in the prior and in the RRef posterior). As seen when analysing the OSSEs results, the net annual budget is not well constrained by the inversions and the absence of change is here purely coincidental.

Contrarily to what happened in In contrast to the OSSEs, the spread of the inversions ensemble is not narrower where the network density is the highest. The results in Central Europe are on the contrary extremely sensitive to the restriction transport model error is not zero, which may explain the slightly higher sensitivity of the results to the extent of the observation ensemble to high (RA) or low (RP) altitude sites, and changes in the observation frequency (Rf) also have an impact (but in a predictable sense: they lead to a correction of the flux of smaller amplitude than e.g. Rref). Changes in the spatial covariance lengths (Rref, R500, R100) correspond roughly to what was seen with the OSSEs: the impact is mostly in regions with low observational coverage network: RO.P and RO.A differ by, on average, 0.02 PgC/month, the double of the average difference between SO.P and SO.A. However the overall spread of results in that second ensemble of inversions is on the same order of magnitude to that obtained with the OSSEs, with a monthly spread ranging from 0.02 PgC/month (January and September) to 0.07 PgC/month (March and August). This indicates that the conclusions of the OSSEs regarding the robustness of the results can be generalized to these inversions with real data.

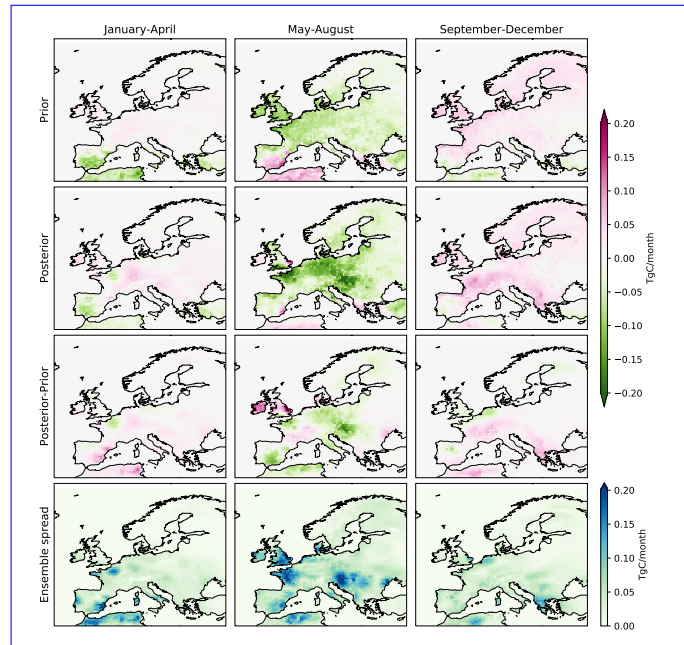
Maps of Maps of the prior and posterior fluxes, as well as the flux adjustments obtained with RRef are shown in the upper and central row of Figure ?? for the reference inversion Rref Figure 9, for three 4-months periods. The January to April and September to December periods correspond approximately to the time of the year when a positive NEE correction is obtained by the inversion, while May to August is the period when the inversion finds increased uptake compared to the prior. While at large scales, the inversion preserves relatively well the spatial NEE distribution, it tends to concentrate a large part of the flux adjustments around some of the observation sites, in particular in Western and Central Europe. On the contrary, flux adjustments are relatively smooth in areas at the domain edges distribution of NEE, the flux adjustment is not as homogeneous as what was obtained with the OSSEs (see also monthly flux adjustments maps in Figure SI5b).

The ensemble variability (lower row of Figure ?? shows the standard deviation of the ensemble at each pixel (and for each 4-month period). The largest discrepancies between the inversions are found around Northern Germany and Poland (in the "Central Europe" region). One likely explanation is site-specific transport model errors: Observations from two sites sensitive partly to the same fluxes, but affected by different systematic model error (biases) may provide apparently contradictory information on the direction and/or amplitude of 9) is much higher than in the OSSEs in North-Western Europe (Northern France, Ireland and the flux adjustments. The inversion can only reconcile such contradictory observational information by clustering strong flux adjustments in small areas and UK), and in Hungary, around the Hegyhatsal observation site (see also



**Figure 8.** Annual Top: Monthly prior (bar plots dashed black line) and monthly posterior NEE (solid line plots: RRef; shaded blue area: ensemble spread); Bottom: Annual NEE for the prior LPJ-GUESS and posterior fluxes aggregated on the entire domain (upper row) or in smaller regions seven inversion posteriors.

Figure 6). In the latter case, this is mainly due to the inclusion or not of this site in the inversions (i.e. RO.A/or time periods: The problem is stronger where the network density is the highest, and the ensemble spread varies the most where the network density is the most impacted by the sites selection in inversions RA and RP. In the RO.P inversions). The discrepancies in North-West Europe were already present in the OSSEs, similar changes in network density did not lead to similar problem because of the use of a perfect transport model but here with real observations the inversions additionally have to compensate for the inaccuracy of the transport model. In particular, errors in the prescribed background concentrations will have a stronger impact on the optimized fluxes in the vicinity of sites that sample predominantly background concentrations, such as the sites in Ireland and the UK. But also, observation sites downwind of large urban areas are more susceptible to be impacted by errors in the prescribed fossil fuel emissions, either because the emission scenario itself is incorrect, or the transport model resolution is too coarse to correctly represent the impact of these emissions at the observation site.



**Figure 9.** Prior Total prior NEE (top) and posterior NEE (middle second row) and NEE in the Rref adjustment (reference third row) scenario for the Rref inversion, in and for three four-months periods 4-months period (JFMA, MJJA, SOND left to right); Variability (standard deviation) of the posterior fluxes across the Low row: Posterior ensemble of simulations (Rref, R500, R100, RP, RA, R1f) spread

## 5.2 Reduction of the observation misfits

The comparison of prior and posterior fit residuals is more critical in this series of inversions than in the OSSEs since here the true fluxes are not known. Figure 10 shows observation fits for the R-denoted series of inversions in a similar way than done in Section ?? for the OSSEs. The overall change in the fit to the observations (right hand panel) from prior Figure 10 provides an overview of the model-data mismatches for Rref and, at the site level, for the sensitivity experiments. As expected, the inversion leads to a reduction in the RMSE, from 5.8 ppm in the prior to 4.8 ppm in the posterior, and to a slight reduction of the mean bias (from -0.2 ppm to posterior is still an improvement, but this time very modest (mean prior bias: -0.75 ppm; mean posterior bias: -0.03 ppm; prior RMSE: 5.74 ppm; posterior RMSE: 5.40 -0.1 ppm).

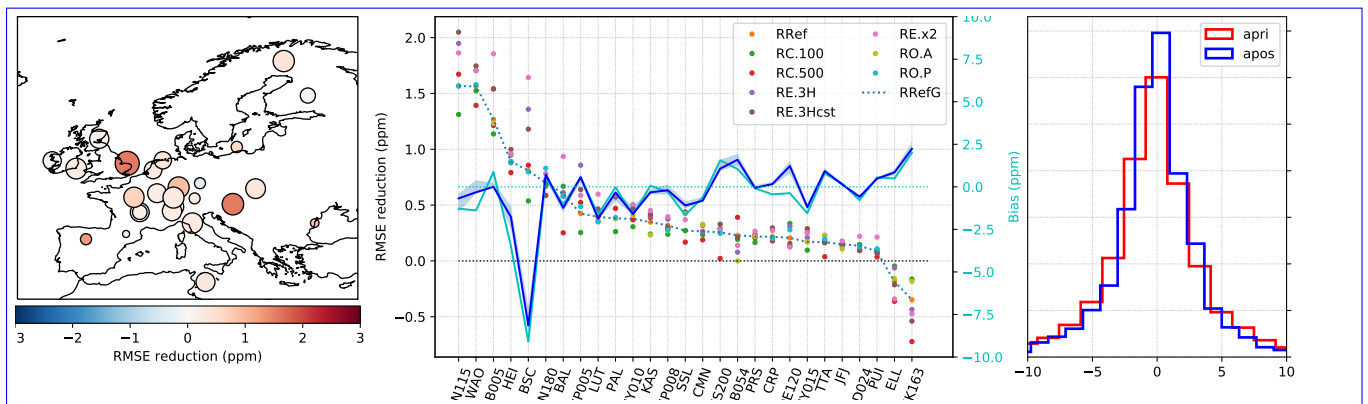
Looking at each site separately (left and middle panels) shows, that, while at most sites the representation of the observations is improved by the inversion (red circles), it is strongly degraded at two sites: OXK (-1.15 ppm) and OPE (-0.74 ppm). At another six sites, smaller degradations of the RMSE occur (from -0.02 ppm at PUI, to -0.25 ppm at CES).

The central plot in Figure 10 shows a comparison of the RMSE reduction at the sites with other inversions, but also the evolution of the model-observation biases for inversion Rref (blue lines, right axis). These values are slightly larger than the ones obtained in the OSSEs, which is consistent with the presence of a non-perfect transport model and boundary conditions.

At the site level, the prior biases are more variable than in the OSSEs ensemble, from -9.1 ppm at Baltic Sea (BSC) to +2ppm at Ochsenkopf (OXK). The posterior biases at OXK and OPE are respectively 3.19 ppm (prior: 2.00) and 2.83 ppm (prior: -0.37)

705 Meanwhile the posterior biases of neighbouring sites HEI and SSL are respectively  $-1.81$  ppm (bias corrections remain very modest at most sites (prior  $-3.48$ ) and  $-1.22$  ppm (prior  $-1.67$ ) the bias even slightly increase at a few sites). The observation sites are close to each other and sample largely the same air masses; it is not possible for the inversion to find a combination of fluxes that improves the fit at all sites (at least not without departing too much from the prior). The large ( $7.5$  ppm) bias at BSC (Black Sea) is computed from a very small number of observations is very small (17 in total, with observational errors up to 8 ppm), which have therefore very little weight in the inversion. The RMSE is generally reduced, except at ELL (Estany LLong, Spain) and OXK (Ochsenkopf, Germany), where the fit to the observations is slightly degraded. Both sites are located in relative proximity to other observation sites, with which their footprints overlap: the degradation of the RMSE results from contradictory constraints provided to the inversions by these different sites. The inversion does not have sufficient degrees of freedom to improve simultaneously the fit at all sites, and therefore degrades the fit to the OXK and ELL observations, which have only few observations (48 and 8, respectively). The problem is common to all the sensitivity runs, and the mean posterior biases are also very similar across the ensemble.

715 Likely causes for these poor misfit reductions are transport model errors (diverging biases at neighbouring sites, introduced either by the foreground transport model (FLEXPART) or by the background one (TM5-4DVAR)) and/or a locally insufficient resolution of the optimized state vector. These causes will need to be investigated thoroughly in a future study, but are not a sign of a malfunction of the optimization method itself. As seen with the OSSEs, a better performance in the fit to observations is not necessarily an indication of a more accurate optimized solution. The site-by-site analysis of the misfits might point to limitations of the transport operator, but a more in-depth analysis would be required, which is out of the scope of this paper.



**Figure 10.** Left: Map of the observation sites in Ref, with the area of the dots proportional to the number of assimilated observation at each site, and the color proportional to the RMSE reduction (prior RMSE minus posterior RMSE). Center: RMSE reduction at each site, for the five sensitivity OSSEs. Right: Prior and posterior distribution of the observation mismatches in Ref (irrespective of the site)

## 720 6 Discussion and conclusions

We have setup an atmospheric inversion system based on an implementation of the variational inversion approach (Section ~~???~~3.1) with a transport model based on an offline coupling between FLEXPART (high-resolution regional transport) and TM5 (coarse-resolution transport of the background fluxes and historical atmospheric CO<sub>2</sub> burden). The inversion was tested through a series of synthetic experiments and realistic inversions, which show that it is working as expected. In this section we discuss separately three aspects of the paper. First the inversion results themselves, then the TM5-FLEXPART coupling and finally the LUMIA system.

### 6.1 Inversion approach and results

~~The inversion technique used in this study is by design not innovative (the definition of the control vector, the specification of the uncertainties, etc. replicate what has been done in previous studies (e.g. Kountouris et al. (2018))), as the aim is We~~  
730 ~~have setup a variational inversion framework, to optimize European NEE at a monthly, 0.5° scale. The setup is intentionally simple: the aim at this stage was to develop a robust technical base and to have a reference setup .The scientific results are therefore at this stage limited (as it also wasn't the aim of the paper), but the analysis of the OSSEs results show that the inversions are~~ for future developments. The transport model is a transposition to TM5 and FLEXPART of the off-line coupling developed by Rödenbeck et al. (2009) for TM3 and STILT, and the optimization itself shares many similarities with existing  
735 ~~inversion systems, e.g. TM5-4DVAR (Basu et al., 2013b), TM3-STILT (Kountouris et al., 2018) or even PyVAR-CHIMERE (Broquet et al., 2011), which should facilitate the comparison of results with these systems.~~

~~The first inversion results suggest that the inversion system is working as expected. Fluxes are well resolved at the monthly scale and large error reductions (with respect to the known truth) occur where the network is the densest, as expected~~In the context of OSSEs, the inversions enable on average a 40% reduction of the flux error at the grid-cell, monthly scale, and the  
740 ~~differences between the optimized fluxes obtained from different sensitivity runs are in line with what could be expected from the different settings used. However, these local error reduction can be of opposite sign, and do not always add up to a net error reduction at larger scales. In particular, while the NEE estimate is generally always improved at the monthly scale, the positive corrections in summer are much stronger than the negative corrections in winter, which results in an overall degradation of the annual NEE.~~

~~The aggregated annual flux estimates are, on the contrary, generally degraded by the inversions compared to the prior. This is because, while both positive and negative prior errors are correctly adjusted in the inversions, the error reduction is much stronger in summer (when the prior strongly overestimates the (prescribed) carbon uptake) . The OSSEs are an ideal setup for such studies with no transport and no observational errors, such that this bias in the annual posterior NEE results entirely from uneven seasonal constraints on the flux adjustments (the prior uncertainties are larger in summer than in winter, because~~  
750 ~~NEE is larger in summer). Using an even month-to-month distribution of the uncertainties (SE.3Hcst inversion) leads to a more realistic annual estimate, but also to a higher occurrence of local degradations of the solution, which further complicates the interpretation of the results. This high sensitivity of the annual NEE to the different choices of prior uncertainty show that~~



755 this metric is not well constrained in our inversions: unless additional constraints are introduced in the system, it will remain difficult to provide a more robust estimate of the annual European NEE than what can be obtained through other estimation techniques (bottom-up modelling, global inversions, etc.). In particular, making use of constraints on the large-scale gradients from the global inversion (background) to the regional one would be beneficial.

760 ~~As mentioned earlier, the focus of the paper is not on the inversion results themselves~~ Another approach is to accept that this metric (the annual European NEE) can remain under-constrained in our inversions, and focus on the aspects of the solution that the inversions really improve: in the OSSEs, regardless of the specific inversion setup, the posterior provides a much more realistic depiction of the seasonal cycle of NEE, and of its spatial variability. Furthermore the results gain in consistency (i.e. become less sensitive to sensitivity experiments) where the observation network is dense, which is encouraging since the observation network in Europe has significantly expanded compared to the data selection used in this paper. The inversions provide relevant information that can help identifying specific shortcomings in the prior flux estimates, and the ~~results shown here might be a worst case scenario: scaling the prior uncertainty to the respiration instead of to the NEE itself, as done here, would for instance partly even out the uncertainties. Nonetheless,~~ practice of performing an ensemble of sensitivity runs greatly helps in identifying the most robust features. It is here facilitated by the low marginal computational cost of computing additional inversions.

770 The OSSEs systematically lead to some degradation of the solution in the parts of the domain that are very densely covered by the observation network, which is counter-intuitive. It may be partly because the prior was already very close to the truth in this part of the ~~results have potential implications for existing inversions (in the absence of a known truth, it is very difficult to tell how adapted is a particular prior error covariance matrix to a given problem), which makes it a worthy topic for a dedicated study~~ domain, which makes it difficult for the inversion to further optimize the solution, but a complementary explanation is that the system may not have sufficient degrees of freedom to adjust the fluxes to simultaneously improve the fit at all observation sites. In particular, the optimization of monthly fluxes is very restrictive. The implementation of an optimization at a higher

775 temporal resolution will therefore be an important next step. In addition, varying the resolution of the optimization according to the density of the observation network may also help (either by varying the resolution of the optimized fluxes, or by varying the covariance lengths in the prior error-covariance matrix).

780 The application of the same inversion approach to real observations leads to smaller flux adjustments than in the OSSEs. This could be a sign that the difference between the LPJ-GUESS prior (used in this second set of inversions) and the true fluxes is smaller than that between the prior and synthetic truth in the OSSEs, but the analysis of the observation misfits reduction also point to potential site-dependent transport model errors. One of the next steps towards improving our inversions will therefore have to be a thorough model calibration effort. In that sense, the flexibility of LUMIA with regards to the transport model is particularly adapted.



## 6.2 TM5-FLEXPART coupling

785 The inversions rely on an offline coupling between the FLEXPART Lagrangian transport model (for regional, high resolution transport) and TM5-4DVAR for providing background concentrations. The setup replicates the 2-step scheme of Rödenbeck et al. (2009) but with different models.

A succinct comparison between this "TM5-FLEXPART" transport model and TM5 itself was performed (Section 3.3.1) and is used as a proxy for the transport model error. It doesn't show any global bias between the two models, but a possible seasonal offset towards the month of November. The prescribed observation uncertainties are scaled up to account for this possible larger model error, so the impact on inversions should be limited. Nonetheless, that possible seasonal bias would need to be investigated and accounted for before deriving scientific conclusions from inversions against real observations.

The choice of the models and of that specific coupling was driven in part by the perspective of exchanges with other groups using similar setups. In the current stage, replacing the FLEXPART response functions from another similar Lagrangian transport model (STILT, NAME (Lin et al., 2003), NAME (Jones et al., 2007), etc.) or the TM5 background time series by data generated with a different model (using either the same or a different technique to estimate background concentrations at the observation sites) is straightforward and will facilitate a better evaluation of the model performance.

Note also that the Rödenbeck et al. (2009) approach means that there is no 'hard' coupling between the two models meaning that there is no risk of having to use an older version of one model because of the lack of implementation of the coupling in newer code. This, of course, also facilitates the exchange of one transport model against another as mentioned above.

From a practical and technical point of view, the current setup presents the advantage of speed and scalability: the application of the transport operator is done independently for each observation and therefore can be distributed on as many CPUs as available. Inversions can thus be performed in very limited (user) time (5-8 hours wall time per inversion on 24 CPUs for the inversions in this paper). This time efficiency is critical for running not only single inversions, but inversion ensembles, which provide a better representation of the real uncertainties. Evolutions of the code for very large ensemble of observations (such as from satellite retrievals) may, however, benefit from further developments (aggregation of observations and footprints; reduction of the number of grid points where possible; etc.).

## 6.3 The LUMIA framework: conclusions and future perspectives

~~LUMIA is primarily designed as a library of (python) code to support the development of inverse modelling activities at Lund University. It aims at being not an integrated atmospheric inversion system (i.e. transport model + inversion algorithm), but rather a toolbox for inversions (and for model-data fusion approaches in general). Currently, the available code is mainly limited to the modules used in the inversions presented in this paper, but applications such as inversions with other tracers (CH<sub>4</sub>, <sup>14</sup>CO<sub>2</sub>), with other regional transport models (STILT) and coupling with a vegetation model (towards a CCDAS) are envisioned.~~

~~The existing (and future) modules are build with the aim to maximize flexibility (We have developed the LUMIA inversion framework, and performed a first set of inversions with it. The framework is initially designed for the purpose of performing~~

regional CO<sub>2</sub> inversions in Europe, however it is designed and developed as a flexible and adaptable inversion system, which enables the easy exchange of major components of the system, such as the transport model or the minimization algorithm, to isolate and study their impact on the inversion results. LUMIA is designed to be transport-model agnostic, i.e. the capacity for the user to build their own inversion experiments, rather than just replicate existing ones), usability and sustainability of the code. The concept of usability refers to the need to limit the time spent by users in understanding the code itself (rather than the algorithms implemented in the code), while the concept of sustainability refers to the capacity for the code to retain its flexibility and usability characteristics throughout future developments. Concretely, it means that the code should attempt to follow a number of simple but essential coding practices (one function, one task; explicit naming of variables; code as close as possible as the original equations; clean I/O; usage of only standard libraries to facilitate portability, etc.). The aim is in particular that a module can be removed/replaced without risking to break other modules.

The LUMIA code is not meant to be a “key-in-hand” system, it targets users having or willing to acquire robust understanding of inverse modelling (it is perfectly usable as a toy model for learning). We therefore do not publish the code in a public repository, but we are very open to collaborations and distribute the code on demand. An archive it is not constructed on top of an existing model and it calls the transport model via a well-defined interface.

Technically, the inversion framework presented in this paper includes three major components: the lumia python library, which contains most of the actual inversion code in the form of independent modules; a transport component, which relies on pre-computed observation footprints and background concentrations; the inversion scripts themselves, which use the lumia library and the transport model to implement the inversion experiment.

The lumia python library defines an ensemble of classes corresponding to basic elements of the inversion setup, e.g. control vector, gradient descent algorithm, transport model interface, observations database, etc. The library is distributed on a git server and is installable via the standard pip tool, which means it can be installed in one single command on a new computer. Although the library is developed and designed for the purpose of inversions, we have made particular effort to ensure the modularity of the code: the different modules can be imported independently and be used to construct new experiments. On the short-term, this design facilitates the re-use of the code in it's current shape is nonetheless included as SI of this document.

## 7 Summary and conclusions

We have presented the LUMIA inversion library pre-/post-processing steps of the inversions, as well as a first set of regional CO<sub>2</sub> inversions performed with LUMIA, relying on a new coupling between the FLEXPART and TM5-4DVAR transport model. LUMIA is designed to be a toolbox for inversions and is meant to support future inverse modelling work at Lund University during the analysis of inversion results. On the longer term, the intention is to avoid that our initial design choices restrict the implementation of future experiments.

The inversions intentionally follow a very classical approach, so as to facilitate comparison with other systems and to set a point of reference. For this initial paper, we have performed regional CO<sub>2</sub> inversions, intentionally using a rather 'classical' inversion design to ensure comparability with other similar setups and to have a reference of comparison for future inversions.

850 but also because it enabled us to focus on the technical robustness of the code. The transport is performed by a script which  
relies on pre-computed FLEXPART observation footprints, and on background concentrations pre-computed with the global  
coarse resolution TM5-4DVAR inverse model (although technically, nothing limits the use of alternative models to compute  
these footprints and background concentrations). ~~The aim is, however, to use LUMIA in the future to develop more innovative~~  
~~approaches (multiple transport models, use of satellite data, optimisation of vegetation model parameters (CCDAS), etc.)~~ This  
855 TM5-FLEXPART setup replicates the 2-step TM3-STILT inversion proposed by Rödenbeck et al. (2009).

Although the inversion setup lacks the maturity of established systems, it offers promising computational performances and the results suggest interesting scientific questions regarding the capacity of regional inversion systems to constrain the annual budget of CO<sub>2</sub>, and point to specific improvements of the inversion approach, which will be implemented in the near future,  
e.g. the optimization of fluxes at a higher temporal resolution. On the longer term, the aim is to use LUMIA as a platform for  
860 testing innovative inversion approaches (multiple transport models, use of satellite data, multi-tracer inversions, optimisation  
of vegetation model parameters (CCDAS), etc.). The code corresponding to the inversions in this manuscript is provided for  
the research community on <https://lumia.nateko.lu.se>, and the access to the git server can be granted on demand.

*Code availability.* The LUMIA source code used in this paper as well as updates can be downloaded from the lumia website: <https://lumia.nateko.lu.se>

865 *Author contributions.* G.M. and M.S designed the experiments and G.M. developed the code and performed the simulations. G.M. prepared the manuscript and M.S. provided corrections and suggestions for improvements.

*Acknowledgements.* G.M. has been funded by the Swedish Research Council project 'Development of regional ecosystem-atmosphere models assimilating the ICOS data for a European-scale intercomparison of net CO<sub>2</sub> fluxes - Eurocom' (DNR 349-2014-6576). The research is part of three Swedish strategic research areas: Modelling the Regional and Global Earth system (MERGE), the e-science collaboration  
870 (eSENCE), and Biodiversity and Ecosystems in a Changing Climate (BECC).

We thank Michael Mischurow for providing the LPJ-GUESS net ecosystem exchange data, Philippe Peylin for providing the ORCHIDEE NEE fluxes, and Greet Janssens-Maenhout for providing the fossil fuel product. We also thank the National Supercomputer Centre at Linköping University (part of the Swedish National Infrastructure for Computing, SNIC) for providing computer resources for the project.

We thank the FLEXPART and TM5 developers for providing the transport models source codes.

875 Finally, we thank all the observation data providers cited in Table 1 for providing the observations used in this paper, and we thank NOAA/GMD for collecting and distributing the data in obspack format.

## References

- BP: Statistical Review of World Energy 2016, Tech. rep., <http://www.bp.com/content/dam/bp/excel/energy-economics/statistical-review-2016/bp-statistical-review-of-world-energy-2016-workbook.xls>, 2016.
- 880 Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO<sub>2</sub> fluxes estimated from GOSAT retrievals of total column CO<sub>2</sub>, *Atmospheric Chemistry and Physics*, 13, 8695–8717, <https://doi.org/10.5194/acp-13-8695-2013>, <http://www.atmos-chem-phys.net/13/8695/2013/>, 2013a.
- Basu, S., Guerlet, S., Butz, A., Houweling, S., Hasekamp, O., Aben, I., Krummel, P., Steele, P., Langenfelds, R., Torn, M., Biraud, S., Stephens, B., Andrews, A., and Worthy, D.: Global CO<sub>2</sub> fluxes estimated from GOSAT retrievals of total column CO<sub>2</sub>, *Atmospheric Chemistry and Physics*, 13, 8695–8717, <https://doi.org/10.5194/acp-13-8695-2013>, <http://www.atmos-chem-phys.net/13/8695/2013/>, 2013b.
- 885 Broquet, G., Chevallier, F., Rayner, P., Aulagnier, C., Pison, I., Ramonet, M., Schmidt, M., Vermeulen, A. T., and Ciais, P.: A European summertime CO<sub>2</sub> biogenic flux inversion at mesoscale from continuous in situ mixing ratio measurements, *Journal of Geophysical Research: Atmospheres*, 116, n/a–n/a, <https://doi.org/10.1029/2011JD016202>, <http://doi.wiley.com/10.1029/2011JD016202>, 2011.
- Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F.-M., Chédin, A., and Ciais, P.: Inferring CO<sub>2</sub> sources and sinks from satellite observations: Method and application to TOVS data, *Journal of Geophysical Research*, 110, D24309, <https://doi.org/10.1029/2005JD006390>, 2005.
- 890 Ciais, P., Crisp, D., Denier van der Gon, H., Engelen, R., Janssens-Maenhout, G., Heimann, M., Rayner, P., and Scholze, M.: Towards a European Operational Observing System to Monitor Fossil CO<sub>2</sub> emissions - Final Report from the expert group, vol. 19, European Commission, Copernicus Climate Change Service, [https://www.copernicus.eu/sites/default/files/2018-10/CO<sub>2</sub>{ }Report{ }22Oct2015.pdf](https://www.copernicus.eu/sites/default/files/2018-10/CO2_{ }Report_{ }22Oct2015.pdf), 2015.
- 895 Ciattaglia, L., Cundari, V., and Colombo, T.: Further measurements of atmospheric carbon dioxide at Mt. Cimone, Italy: 1979–1985, *Tellus B*, 39 B, 13–20, <https://doi.org/10.1111/j.1600-0889.1987.tb00266.x>, 1987.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Q. J. R. Meteorol Soc*, 120, 1367–1387, <http://onlinelibrary.wiley.com/doi/10.1002/qj.49712051912/abstract>, 1994.
- 900 Denier van der Gon, H. D., Hendriks, C., Kuenen, J., Segers, A., and Visschedijk, A.: Description of current temporal emission patterns and sensitivity of predicted AQ for temporal emission patterns, in: EU FP7 MACC deliverable report D\_D-EMIS\_1.3, [https://atmosphere.copernicus.eu/sites/default/files/2019-07/MACC{ }TNO{ }del{ }1{ }3{ }v2.pdf](https://atmosphere.copernicus.eu/sites/default/files/2019-07/MACC_{ }TNO_{ }del_{ }1_{ }3_{ }v2.pdf), 2011.
- Dlugokencky, E., Mund, J. M., Crotwell, A., Crotwell, M., and Thoning, K.: Atmospheric Carbon Dioxide Dry Air Mole Fractions from the NOAA ESRL Carbon Cycle Cooperative Global Air Sampling Network, 1968–2018, Version: 2019-07, <https://doi.org/https://doi.org/10.15138/wkgj-f215>, 2019.
- 905 Errico, R. M.: What Is an Adjoint Model ?, *Bulletin of the American Meteorological Society*, 78, 2577–2591, 1997.
- Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO<sub>2</sub> assimilation in leaves of C<sub>3</sub> species, *Planta*, 149, 78–90, <https://doi.org/10.1007/BF00386231>, <https://doi.org/10.1007/BF00386231>, 1980.
- Ganesan, A. L., Manning, A. J., Grant, A., Young, D., Oram, D. E., Sturges, W. T., Moncrieff, J. B., and O’Doherty, S.: Quantifying methane and nitrous oxide emissions from the UK and Ireland using a national-scale monitoring network, *Atmospheric Chemistry and Physics*, 15, 6393–6406, <https://doi.org/10.5194/acp-15-6393-2015>, <http://www.atmos-chem-phys.net/15/6393/2015/>, 2015.
- 910

- Giglio, L., Randerson, J. T., and Van Der Werf, G. R.: Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4), *Journal of Geophysical Research: Biogeosciences*, 118, 317–328, <https://doi.org/10.1002/jgrg.20042>, 2013.
- 915 Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of annual mean {CO}<sub>2</sub> sources and sinks, *Nature*, 415, 626–630, 2002.
- Hammer, S., Glatzel Mattheier, H., Müller, L., Sabasch, M., Schmidt, M., Schmitt, S., Schönherr, C., Vogel, F., Worthy, D., and Levin, I.: A  
920 gas chromatographic system for high-precision quasi-continuous atmospheric measurements of CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, SF<sub>6</sub>, CO and H<sub>2</sub>, 2008.
- Haszpra, L., Barcza, Z., Bakwin, P., Berger, B., Davis, K. J., and Weidinger, T.: Measuring system for the long-term monitoring of biosphere/atmosphere exchange of carbon dioxide, *Journal of Geophysical Research Atmospheres*, 106, 3057–3069, <https://doi.org/10.1029/2000JD900600>, 2001.
- Hatakka, J., Aalto, T., Aaltonen, V., Aurela, M., Hakola, H., Komppula, M., Laurila, T., Lihavainen, H., Paatero, J., Salminen, K., and  
925 Viisanen, Y.: Overview of the atmospheric research activities and results at Pallas GAW station, *Boreal Environment Research*, 8, 365–383, 2003.
- Houweling, S., Krol, M., Bergamaschi, P., Frankenberg, C., Dlugokencky, E. J., Morino, I., Notholt, J., Sherlock, V., Wunch, D., Beck, V., Gerbig, C., Chen, H., Kort, E. A., Röckmann, T., and Aben, I.: A multi-year methane inversion using SCIAMACHY, accounting for systematic errors using TCCON measurements, *Atmospheric Chemistry and Physics*, 14, 3991–4012, <https://doi.org/10.5194/acp-14-3991-2014>, <https://www.atmos-chem-phys.net/14/3991/2014/>, 2014.  
930
- Huijnen, V., Williams, J., Van Weele, M., Van Noije, T., Krol, M., Dentener, F., Segers, A., Houweling, S., Peters, W., De Laat, J., Boersma, F., Bergamaschi, P., Van Velthoven, P., Le Sager, P., Eskes, H., Alkemade, F., Scheele, R., Nédélec, P., and Pätz, H. W.: The global chemistry transport model TM5: Description and evaluation of the tropospheric chemistry version 3.0, *Geoscientific Model Development*, 3, 445–473, <https://doi.org/10.5194/gmd-3-445-2010>, 2010.
- 935 Hurtt, G. C., Chini, L. P., Frolking, S., Betts, R. A., Feddema, J., Fischer, G., Fisk, J. P., Hibbard, K., Houghton, R. A., Janetos, A., Jones, C. D., Kindermann, G., Kinoshita, T., Klein Goldewijk, K., Riahi, K., Shevliakova, E., Smith, S., Stehfest, E., Thomson, A., Thornton, P., van Vuuren, D. P., and Wang, Y. P.: Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands, *Climatic Change*, 109, 117–161, <https://doi.org/10.1007/s10584-011-0153-2>, 2011.
- 940 Janssens-Maenhout, G., Crippa, M., Guizzardi, D., Muntean, M., Schaaf, E., Dentener, F., Bergamaschi, P., Pagliari, V., Olivier, J. G. J., Peters, J. A. H. W., van Aardenne, J. A., Monni, S., Doering, U., Petrescu, A. M. R., Solazzo, E., and Oreggioni, G. D.: EDGAR v4.3.2 Global Atlas of the three major greenhouse gas emissions for the period 1970–2012, *Earth System Science Data*, 11, 959–1002, <https://doi.org/10.5194/essd-11-959-2019>, <https://www.earth-syst-sci-data.net/11/959/2019/>, 2019.
- Jones, A., Thomson, D., Hort, M., and Devenish, B.: The U.K. Met Office’s next generation atmospheric dispersion model, NAME III, in:  
945 *Air pollution modeling and its applications XVII (Proceedings of the 27th NATO/CCMS International Technical Meeting on Air Pollution Modelling and its Application)*, edited by Borrego, C. and A.-L. N., pp. 580–589, Springer, 2007.
- Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P. J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J. P., Heimann, M., Hooker-Stroud, A., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E. N., Köstler, C., Mathieu, P. P., Pinty, B., Reick, C. H., Rödenbeck, C., Schnur, R., Scipal, K., Sebald, C., Stacke, T., Van Scheltinga, A. T., Vossbeck, M., Widmann,

- 950 H., and Ziehn, T.: The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges, *Journal of Geophysical Research: Biogeosciences*, 118, 1414–1426, <https://doi.org/10.1002/jgrg.20118>, 2013.
- Kountouris, P., Gerbig, C., Rödenbeck, C., Karstens, U., Koch, T. F., and Heimann, M.: Technical Note: Atmospheric CO<sub>2</sub> inversions on the mesoscale using data-driven prior uncertainties: methodology and system evaluation, *Atmospheric Chemistry and Physics*, 18, 3027–3045, <https://doi.org/10.5194/acp-18-3027-2018>, <https://www.atmos-chem-phys.net/18/3027/2018/>, 2018.
- 955 Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, *Global Biogeochemical Cycles*, 19, 1–33, <https://doi.org/10.1029/2003GB002199>, 2005.
- Lanczos, C.: An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators, *Journal Of Research Of The National Bureau Of Standards*, 45, 255–282, 1950.
- 960 Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., Pickers, P. A., Korsbakken, J. I., Peters, G. P., Canadell, J. G., Arneeth, A., Arora, V. K., Barbero, L., Bastos, A., Bopp, L., Chevallier, F., Chini, L. P., Ciais, P., Doney, S. C., Gkritzalis, T., Goll, D. S., Harris, I., Haverd, V., Hoffman, F. M., Hoppema, M., Houghton, R. A., Hurtt, G., Ilyina, T., Jain, A. K., Johannessen, T., Jones, C. D., Kato, E., Keeling, R. F., Goldewijk, K. K., Landschützer, P., Lefèvre, N., Lienert, S., Liu, Z., Lombardozi, D., Metzl, N., Munro, D. R., Nabel, J. E. M. S., Nakaoka, S.-i., Neill, C., Olsen, A., Ono, T., Patra, P., Peregon, A., Peters, W., Peylin, P., Pfeil, B., Pierrot,
- 965 D., Poulter, B., Rehder, G., Resplandy, L., Robertson, E., Rocher, M., Rödenbeck, C., Schuster, U., Schwinger, J., Séférian, R., Skjelvan, I., Steinhoff, T., Sutton, A., Tans, P. P., Tian, H., Tilbrook, B., Tubiello, F. N., van der Laan-Luijkx, I. T., van der Werf, G. R., Viovy, N., Walker, A. P., Wiltshire, A. J., Wright, R., Zaehle, S., and Zheng, B.: Global Carbon Budget 2018, *Earth System Science Data*, 10, 2141–2194, <https://doi.org/10.5194/essd-10-2141-2018>, <https://www.earth-syst-sci-data.net/10/2141/2018/>, 2018.
- Lin, J. C., Gerbig, C., Wofsy, S. C., Andrews, A. E., Daube, B. C., Davis, K. J., and Grainger, C. A.: A near-field tool for simulating the up-
- 970 stream influence of atmospheric observations: The Stochastic Time-Inverted Lagrangian Transport (STILT) model, *Journal of Geophysical Research: Atmospheres*, 108, <https://doi.org/10.1029/2002JD003161>, <https://doi.org/10.1029/2002JD003161>, 2003.
- Meirink, J. F., Bergamaschi, P., and Krol, M. C.: Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, *Atmospheric Chemistry and Physics*, 8, 6341–6353, 2008.
- Michalak, A. M., Randazzo, N. A., and Chevallier, F.: Diagnostic methods for atmospheric inversions of long-lived greenhouse gases, *Atmospheric Chemistry and Physics Discussions*, pp. 1–27, <https://doi.org/10.5194/acp-2016-800>, <http://www.atmos-chem-phys-discuss.net/acp-2016-800/{#}discussion>, 2016.
- 975 Michalak, A. M., Randazzo, N. A., and Chevallier, F.: Diagnostic methods for atmospheric inversions of long-lived greenhouse gases, *Atmospheric Chemistry and Physics*, 17, 7405–7421, <https://doi.org/10.5194/acp-17-7405-2017>, <https://www.atmos-chem-phys.net/17/7405/2017/>, 2017.
- 980 NOAA Carbon Cycle Group ObsPack Team: Multi-laboratory compilation of atmospheric methane data for the period 1957-2017; `obspack_ch4_1_GLOBALVIEWplus_v1.0_2019_01_08`; <https://doi.org/https://doi.org/10.25925/20190108>, <https://www.esrl.noaa.gov/gmd/ccgg/obspack/data.php?id=obspack{ }ch4{ }1{ }GLOBALVIEWplus{ }v1.0{ }2019-01-08>, 2019.
- Peters, W., Jacobson, A. R., Sweeney, C., Andrews, A. E., Conway, T. J., Masarie, K., Miller, J. B., Bruhwiler, L. M. P., Pétron, G., Hirsch, A. I., Worthy, D. E. J., van der Werf, G. R., Randerson, J. T., Wennberg, P. O., Krol, M. C., and Tans, P. P.: An atmospheric perspective
- 985 on North American carbon dioxide exchange: CarbonTracker., *Proceedings of the National Academy of Sciences of the United States of America*, 104, 18 925–18 930, <https://doi.org/10.1073/pnas.0708986104>, <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2141884{ }& }tool=pmcentrez{ }& }rendertype=abstract>, 2007.

- Peylin, P., Law, R. M., Gurney, K. R., Chevallier, F., Jacobson, A. R., Maki, T., Niwa, Y., Patra, P. K., Peters, W., Rayner, P. J., Rödenbeck, C., Van Der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: Results from an ensemble of atmospheric CO<sub>2</sub> inversions, *Biogeosciences*, 10, 6699–6720, <https://doi.org/10.5194/bg-10-6699-2013>, 2013.
- 990 Pisso, I., Sollum, E., Grythe, H., Kristiansen, N., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaafink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhardt, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and Stohl, A.: The Lagrangian particle dispersion model FLEXPART version 10.3, *Geoscientific Model Development Discussions*, pp. 1–67, <https://doi.org/10.5194/gmd-2018-333>, <https://www.geosci-model-dev-discuss.net/gmd-2018-333/>, 2019.
- 995 Rayner, P. J., Michalak, A. M., and Chevallier, F.: Fundamentals of Data Assimilation applied to biogeochemistry, *Atmospheric Chemistry and Physics Discussions*, pp. 1–32, <https://doi.org/10.5194/acp-2018-1081>, <https://www.atmos-chem-phys-discuss.net/acp-2018-1081/>, 2018.
- Rödenbeck, C., Gerbig, C., Trusilova, K., and Heimann, M.: A two-step scheme for high-resolution regional atmospheric trace gas inversions based on independent models, *Atmospheric Chemistry and Physics Discussions*, 9, 1727–1756, <https://doi.org/10.5194/acpd-9-1727-2009>, 1000 2009.
- Rödenbeck, C., Keeling, R. F., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., and Heimann, M.: Global surface-ocean pCO<sub>2</sub> and sea-Air CO<sub>2</sub> flux variability from an observation-driven ocean mixed-layer scheme, *Ocean Science*, 9, 193–216, <https://doi.org/10.5194/os-9-193-2013>, 2013.
- Rozanski, K., Necki, J., Chmura, L., Sliwka, I., Zimnoch, M., Bielewski, J., Galkowski, M., Bartyzel, J., and Rosiek, J.: Anthropogenic changes of CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, CFC1<sub>3</sub>, CF<sub>2</sub>Cl<sub>2</sub>, CCl<sub>2</sub>FCClF<sub>2</sub>, CHCl<sub>3</sub>, CH<sub>3</sub>CCl<sub>3</sub>, CCl<sub>4</sub>, SF<sub>6</sub> and SF<sub>5</sub>CF<sub>3</sub> mixing ratios in the atmosphere over southern Poland, *Geological Quarterly*, 58, <https://doi.org/10.7306/gq.1163>, 2014.
- Schaefer, K., Collatz, G. J., Tans, P., Denning, A. S., Baker, I., Berry, J., Prihodko, L., Suits, N., and Philpott, A.: Combined simple biosphere/carnegie-ames-stanford approach terrestrial carbon cycle model, *Journal of Geophysical Research: Biogeosciences*, 113, 1–13, <https://doi.org/10.1029/2007JG000603>, 2008.
- 1010 Schmidt, M.: The Schauinsland CO<sub>2</sub> record: 30 years of continental observations and their implications for the variability of the European CO<sub>2</sub> budget, *Journal of Geophysical Research*, 108, 1–7, <https://doi.org/10.1029/2002jd003085>, 2003.
- Seibert, P. and Frank, a.: Source-receptor matrix calculation with a Lagrangian particle dispersion model in backward mode, *Atmospheric Chemistry and Physics*, 4, 51–63, <https://doi.org/10.5194/acp-4-51-2004>, 2004.
- Sitch, S., Friedlingstein, P., Gruber, N., Jones, S. D., Murray-Tortarolo, G., Ahlström, A., Doney, S. C., Graven, H., Heinze, C., Huntingford, C., Levis, S., Levy, P. E., Lomas, M., Poulter, B., Viovy, N., Zaehle, S., Zeng, N., Arneth, A., Bonan, G., Bopp, L., Canadell, J. G., Chevallier, F., Ciais, P., Ellis, R., Gloor, M., Peylin, P., Piao, S. L., Le Quééré, C., Smith, B., Zhu, Z., and Myneni, R.: Recent trends and drivers of regional sources and sinks of carbon dioxide, *Biogeosciences*, 12, 653–679, <https://doi.org/10.5194/bg-12-653-2015>, <https://www.biogeosciences.net/12/653/2015/>, 2015.
- 1015 Smith, B., Wärlind, D., Arneth, A., Hickler, T., Leadley, P., Siltberg, J., and Zaehle, S.: Implications of incorporating N cycling and N limitations on primary production in an individual-based dynamic vegetation model, *Biogeosciences*, 11, 2027–2054, <https://doi.org/10.5194/bg-11-2027-2014>, 2014.
- 1020 Steinbach, J., Gerbig, C., Rödenbeck, C., Karstens, U., Minejima, C., and Mukai, H.: The CO<sub>2</sub> release and Oxygen uptake from Fossil Fuel Emission Estimate (COFFEE) dataset: effects from varying oxidative ratios, *Atmospheric Chemistry and Physics*, 11, 6855–6870, <https://doi.org/10.5194/acp-11-6855-2011>, <https://www.atmos-chem-phys.net/11/6855/2011/>, 2011.

- 1025 Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P.: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, in: IPCC, 2013, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.
- Stohl, A., Sodemann, H., Eckhardt, S., Frank, A., Seibert, P., and Wotawa, G.: The Lagrangian particle dispersion model FLEXPART version 8.2, Tech. rep., 2010.
- 1030 Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., Hales, B., Friederich, G., Chavez, F., Sabine, C., Watson, A., Bakker, D. C., Schuster, U., Yoshikawa-Inoue, H., Ishii, M., Midorikawa, T., Nojiri, Y., Körtzinger, A., Steinhoff, T., Hoppema, M., Olafsson, J., Arnarson, T. S., Johannessen, T., Olsen, A., Bellerby, R., Wong, C., Delille, B., Bates, N., and de Baar, H. J.: Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea–air CO<sub>2</sub> flux over the global oceans, *Deep Sea Research Part II: Topical Studies in Oceanography*, 56, 554–577, <https://doi.org/10.1016/J.DSR2.2008.12.009>, <https://www.sciencedirect.com/science/article/pii/S0967064508004311>, 2009.
- 1035 Thompson, R. L., Stohl, A., Zhou, L. X., Dlugokencky, E., Fukuyama, Y., Tohjima, Y., Kim, S. Y., Lee, H., Nisbet, E. G., Fisher, R. E., Lowry, D., Weiss, R. F., Prinn, R. G., O’Doherty, S., Young, D., and White, J. W.: Methane emissions in East Asia for 2000–2011 estimated using an atmospheric Bayesian inversion, *Journal of Geophysical Research*, 120, 4352–4369, <https://doi.org/10.1002/2014JD022394>, 2015.
- Turner, a. J. and Jacob, D. J.: Balancing aggregation and smoothing errors in inverse models, *Atmospheric Chemistry and Physics*, 15, 7039–7048, <https://doi.org/10.5194/acp-15-7039-2015>, <http://www.atmos-chem-phys.net/15/7039/2015/>, 2015.
- 1040 Uglietti, C., Leuenberger, M., and Brunner, D.: European source and sink areas of CO<sub>2</sub> retrieved from Lagrangian transport model interpretation of combined O<sub>2</sub> and CO<sub>2</sub> measurements at the high alpine research station Jungfraujoch, *Atmospheric Chemistry and Physics*, 11, 8017–8036, <https://doi.org/10.5194/acp-11-8017-2011>, 2011.
- van der Laan, S., Neubert, R. E. M., and Meijer, H. A. J.: A single gas chromatograph for accurate atmospheric mixing ratio measurements of CO<sub>2</sub>, CH<sub>4</sub>, N<sub>2</sub>O, SF<sub>6</sub> and CO, *Atmospheric Measurement Techniques Discussions*, 2, 1321–1349, <https://doi.org/10.5194/amtd-2-1321-2009>, 2009.
- 1045 Van Der Werf, G. R., Randerson, J. T., Giglio, L., Van Leeuwen, T. T., Chen, Y., Rogers, B. M., Mu, M., Van Marle, M. J., Morton, D. C., Collatz, G. J., Yokelson, R. J., and Kasibhatla, P. S.: Global fire emissions estimates during 1997–2016, *Earth System Science Data*, 9, 697–720, <https://doi.org/10.5194/essd-9-697-2017>, 2017.
- 1050 Vermeulen, a. T., Hensen, a., Popa, M. E., Van Den Bulk, W. C. M., and Jongejan, P. a. C.: Greenhouse gas observations from Cabauw Tall Tower (1992–2010), *Atmospheric Measurement Techniques*, 4, 617–644, <https://doi.org/10.5194/amt-4-617-2011>, 2011.
- Weedon, G. P., Balsamo, G., Bellouin, N., Gomes, S., Best, M. J., and Viterbo, P.: Data methodology applied to ERA-Interim reanalysis data, *Water Resources Research*, 50, 7505–7514, <https://doi.org/10.1002/2014WR015638>.Received, 2014.
- Wilson, P.: Insight into the Carbon Cycle from Continuous Measurements of Oxygen and Carbon Dioxide at Weybourne Atmospheric Observatory, UK, Ph.D. thesis, University of East Anglia, 2012.
- 1055 Yver, C., Schmidt, M., Bousquet, P., and Ramonet, M.: Measurements of molecular hydrogen and carbon monoxide on the Trainou tall tower, *Tellus, Series B: Chemical and Physical Meteorology*, 63, 52–63, <https://doi.org/10.1111/j.1600-0889.2010.00520.x>, 2011.