

# Towards the Assimilation of Atmospheric CO<sub>2</sub> Concentration Data in a Land Surface Model using Adjoint-free Variational Methods

Simon Beylat<sup>1,2</sup>, Nina Raoult<sup>3</sup>, Cédric Bacour<sup>1</sup>, Natalie Douglas<sup>4</sup>, Tristan Quaife<sup>4</sup>, Vladislav Bastrikov<sup>5</sup>, Peter J. Rayner<sup>2</sup>, and Philippe Peylin<sup>1</sup>

<sup>1</sup>Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France

<sup>2</sup>School of Geography, Earth and Atmospheric Sciences, University of Melbourne, Parkville, 3010 Victoria, Australia

<sup>3</sup>European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK

<sup>4</sup>National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading, UK

<sup>5</sup>Science Partners, Paris, France

**Correspondence:** simon.beylat@lsce.ipsl.fr

## Abstract.

A comprehensive understanding and an accurate modelling of the terrestrial carbon cycle, are of paramount importance to improve projections of the global carbon cycle and more accurately gauge its impact on global climate systems. Land Surface Models, which have become an important component of weather and climate applications, simulate key aspects of the terrestrial carbon cycle such as photosynthesis and respiration. These models rely on parameterisations that require careful calibration. In this study we explore the assimilation of atmospheric CO<sub>2</sub> concentration data for parameter calibration of the ORCHIDEE Land Surface Model using a ~~4D~~~~EnVar~~~~EnVarDA~~, an adjoint-free ensemble-variational data assimilation method. By circumventing the challenges associated with developing and maintaining tangent linear and adjoint models, the ~~4D~~~~EnVar~~~~EnVarDA~~ method offers a very promising alternative. Using synthetic observations generated through a twin experiment, we demonstrate the ability of the ~~4D~~~~EnVar~~~~EnVarDA~~ to assimilate atmospheric CO<sub>2</sub> concentrations for model parameter calibration. We then compare the results to a ~~4D~~~~Var~~~~VarDA~~ method that uses finite differences to estimate tangent linear and adjoint models, which reveal that the ~~4D~~~~EnVar~~~~EnVarDA~~ is superior in terms of computational efficiency and fit to the observations as well as parameter recovery.

## 1 Introduction

Since the link between the increase in atmospheric CO<sub>2</sub> concentration and global warming was revealed, understanding the carbon cycle has become essential. This increase is mainly due to anthropogenic emissions (IPCC, 2023), half of which are absorbed by oceans and lands. To improve predictions of the carbon cycle and reduce its associated uncertainty in climate projections, it is essential to better understand the mechanism of the carbon sink, particularly its land component, which remains the most uncertain aspect of the global carbon budget (Friedlingstein et al., 2023).

Atmospheric CO<sub>2</sub> concentration data have long been considered a rich source of information to understand the global carbon cycle and characterise the spatio-temporal variation of natural CO<sub>2</sub> fluxes (Kaminski et al., 1999a; Rayner et al., 1999; Bousquet

et al., 2000; Gurney et al., 2002; Peylin et al., 2005, 2013; Chevallier et al., 2014). Given that the atmosphere is relatively well mixed, the observed concentration gradients (in space and time) can be used to identify the large-scale characteristics of the underlying surface fluxes. Indeed this surface fluxes are the primary drivers of these gradients. Studying these data gives us an overall view of all the components of the carbon cycle. For more than 25 years, atmospheric CO<sub>2</sub> inversions have been used to estimate natural CO<sub>2</sub> surface fluxes, using atmospheric transport models and Bayesian inversion frameworks (Kaminski et al., 1999b; Enting, 2002; Chevallier et al., 2005, 2007; Baker et al., 2006; Rayner et al., 2019; Berchet et al., 2021). Atmospheric transport models represent the transport of atmospheric tracers, making it possible to simulate the 3D fields of atmospheric CO<sub>2</sub> concentration based on a CO<sub>2</sub> surface flux scenario, including all components of the carbon cycle: natural land and ocean fluxes, and anthropogenic emissions from fossil fuels and cement. By inverting the atmospheric transport and using the CO<sub>2</sub> surface flux scenario as prior information, atmospheric inversions statistically adjust the surface CO<sub>2</sub> fluxes, minimising the differences between observed and modelled concentrations. This statistical optimisation generally assumes that the corrections to CO<sub>2</sub> surface fluxes are isotropic in time and space. This suggests that errors in surface fluxes are only correlated in space by distance between points, and not by direction. Furthermore, these errors are not strongly correlated in time. While this approach has been valuable for understanding the global carbon cycle, it only estimates the net surface fluxes, with no direct information on the underlying components (i.e., photosynthesis uptake, ecosystem respiration release, fire release, etc.). Consequently, this approach is also not suitable to make future projections.

Over the same period, land surface models (LSMs) have become an important component of Earth system models, representing a wide range of interactions between the land surface and the atmosphere. As their role has expanded, these models have incorporated an increasing number of complex processes (Fisher and Koven, 2020), and have come to play a key role in weather and climate applications. LSMs now simulate key aspects of the terrestrial carbon cycle, including soils and vegetation dynamics, providing valuable insights into the main drivers of the land carbon budget and enabling future projections. Given the complexity and the small-scale nature of many of these processes, they are represented using mechanistic and empirical formulations. To accurately model these processes, LSMs rely on parametrisations that must be carefully calibrated to ensure their simulations are consistent with actual observations. One promising approach for calibrating these parameters is the use of atmospheric CO<sub>2</sub> concentration data, which offers a global constraint for large-scale calibration, serving as an alternative to traditional atmospheric inversions (Knorr and Heimann, 1995; Kaminski et al., 2002, 2012; Rayner et al., 2005; Scholze et al., 2007; Peylin et al., 2016; Schürmann et al., 2016; Castro-Morales et al., 2019; Bacour et al., 2023). This assimilation enables the calibration of LSM parameters by adjusting the underlying process representations rather than directly modifying the fluxes themselves. Such an approach also helps to identify structural errors within the models and enhances our understanding of the various processes involved. Once calibrated and refined, these models can be applied to generate more reliable future projections.

There is a long history of using data assimilation frameworks to calibrate LSM parameters (Rayner, 2010; MacBean et al., 2022; Raoult et al., 2024b). ~~Popular methods—also used in atmospheric inversions—are variational Bayesian methods and, more specifically,~~ Most of the methods used for parameter calibration are derived from Bayesian formulations of inverse problems and defined here as variational data assimilation (VarDA) methods. The VarDA method is inspired by the four-

dimensional variational (4DVar) method. ~~This method, which~~ was originally developed in the field of meteorology and Earth sciences (Talagrand and Courtier, 1987; Courtier et al., 1994; Asch et al., 2016) and also employed in atmospheric inversions to correct the ~~initial state of the model and has been shown to be robust and very efficient~~ (Talagrand and Courtier, 1987; Courtier et al., 1994).

60 ~~The 4DVar approach involves defining surface CO<sub>2</sub> fluxes~~ (Chevallier et al., 2005; Basu et al., 2013; Liu et al., 2021). ~~This approach is characterised by the definition of a cost function~~ ~~(which is usually, which is typically~~ based on a ~~least-square criterion)~~ least-squares criterion. This cost function calculates two terms: (i) an observation term that computes the difference between observations and model outputs ~~distributed in space and time as well as,~~ and (ii) a background term that ~~accounts for~~ incorporates prior knowledge of the ~~parameters~~ state. The computation of both terms is performed in space and time. We define

65 here the VarDA method, as our focus is not on directly optimizing the prior state. Instead, we concentrate on time-invariant parameters used in the parameterisation that defines the variable of interest, such as the Net Carbon Flux. Therefore, while the observation term of the cost function incorporates time-distributed observations and model predictions - comparing them across multiple time points - the background term only compares prior parameter values once, as these values remain constant over time. Furthermore, with the VarDA method, a single assimilation cycle covering the entire observation period is used, which

70 differs from the conventional 4DVar framework, which generally uses sequential cycles with shorter assimilation windows. In order to minimise this cost function, the ~~4DVar~~ VarDA method calculates its gradient with respect to the different parameters to be calibrated. A precise calculation of the gradient of this cost function requires the tangent linear and the adjoint model (Plessix, 2006). To obtain these models, the code must be differentiated. This task can be performed using automatic differentiation software (Giering and Kaminski, 2003), but the model code must be cleaned up and small modifications made to ensure

75 differentiation (e.g. the reformulation of minimum and maximum computations to enable a smooth transition at the edge, Schürmann et al. (2016)). For some LSMs, it is possible to keep the model compliant using automatic differentiation software (Kaminski et al., 2012; Knorr et al., 2024), however, for complex community models such as ORCHIDEE or JULES LSMs (Raoult et al., 2016), maintaining the tangent linear and adjoint models is very challenging due to their continuous evolution. In this case, one approach to calculating the gradient is to use finite differences to estimate the gradient of the cost function in

80 order to use the ~~4DVar~~ VarDA method (Santaren et al., 2007; MacBean et al., 2015; Peylin et al., 2016; Bacour et al., 2019).

Several avenues of research have been explored for parameter calibration, including alternative methods to minimise the cost function and the application of new machine learning techniques (Raoult et al., 2024a, b). Ensemble methods have proven effective for the calibration of LSM parameters, such as Genetic Algorithm (GA) (Santaren et al., 2014; Bastrikov et al., 2018) or Markov chain Monte Carlo (MCMC) (Ziehn et al., 2012). These methods require a large number of simulations and

85 are primarily used with low-cost computational models and for on-site applications, as here they are relatively inexpensive. Parameter calibration in Earth system models has also been the subject of more intensive research (Hourdin et al., 2017). It has led to the development of new methods - emulator-based methods (Williamson et al., 2013; Couvreur et al., 2021) for instance - that have been used to calibrate components of Earth system models (Watson-Parris et al., 2021; Hourdin et al., 2023) such as ocean and atmospheric model (Williamson et al., 2017; Hourdin et al., 2021; King et al., 2024). In these methods, the model is

90 replaced by an emulator - a computationally efficient statistical model designed to reproduce the behaviour of complex models - to enable numerous simulations and rule out sets of parameters that are not plausible. These methods are gaining in popularity

for the calibration of LSM parameters (Dagon et al., 2020; Baker et al., 2022; McNeall et al., 2024; Raoult et al., 2024a) but they still require a large ensemble of simulations to build the emulator. More recently, an ensemble 4DVar method named 4DEnVar implemented in Pinnington et al. (2020) for LSM parameter estimation has proved very promising. This method uses a small ensemble to circumvent the necessity for a tangent linear and adjoint model. This 4DEnVar method has been used to estimate JULES LSM crop parameters at a single Nebraskan site (Pinnington et al., 2020) and to calibrate pedotransfer functions to improve JULES LSM soil moisture predictions over East Anglia (Pinnington et al., 2021) and the whole of the UK (Cooper et al., 2021). This method was also successfully used by Douglas et al. (2025) to calibrate the parameters of a simple carbon model in a twin experiment. Although the method was defined as an 4DEnVar in Pinnington et al. (2020) and Douglas et al. (2025), we choose to refer to it as EnVarDA to maintain consistency with the definitions previously presented.

The problem addressed in this article is the assimilation of atmospheric CO<sub>2</sub> data to calibrate the parameters of the ORCHIDEE LSM. For this application, we need to couple ORCHIDEE with an atmospheric transport model, which, in our case, is LMDZ, as they are historically linked and represent the land and atmospheric components of the IPSL (Institut Pierre-Simon-Laplace) Earth system model (Boucher et al., 2020). While tangent linear and adjoint models can be easily derived for the transport model (Hourdin et al., 2006; Hourdin and Talagrand, 2006), this is not the case for the ORCHIDEE LSM. Although tangent linear or adjoint models are not required for methods such as GA, MCMC, or emulator-based approaches, these methods necessitate defining a large ensemble, making them unfeasible for use in this study due to the time-consuming nature of model simulations. The purpose of this article is to present an adjoint-free data assimilation framework that facilitates the assimilation of atmospheric CO<sub>2</sub> concentrations. We demonstrate the potential of ~~4DEnVar~~ EnVarDA using synthetic observation data according to different criteria: i) the differences between synthetic observation and simulation of atmospheric CO<sub>2</sub> concentration, ii) the spatial distribution of carbon flux as well as their subcomponent, and iii) the recovery of the true parameters used to generate the synthetic observation. We also compare the performance of ~~4DEnVar~~ EnVarDA using these criteria with that of ~~4DVar~~ VarDA with finite differences Section 2 presents the methods, the models, the data and the experiments. Results are shown in Section 3, with discussions and conclusions in Sections 4 and 5, respectively.

## 2 Method

### 2.1 Models and datasets

#### 2.1.1 ORCHIDEE land surface model

ORCHIDEE (ORGanizing Carbon and Hydrology In Dynamic EcosystEms; originally described in Krinner et al. (2005)) is a process-based LSM that simulates the exchange of carbon, water and energy between the surface, vegetation, and the atmosphere. It is composed of different sub-models: a fast one that calculates photosynthesis, hydrology and energy balance every 30 minutes; and a slow one that simulates carbon allocation in plant reservoirs, soil carbon dynamics and litter decomposition every day. In this study, we used the ORCHIDEE version 2 used in the Coupled Model Intercomparison Project Phase 6 (CMIP6) (Boucher et al., 2020; Lurton et al., 2020). This version contains significant improvements over the original version

described by Krinner et al. (2005). The soil hydrology scheme is based on Richards' equation that describes vertical water  
125 fluxes for a soil depth of 2 m discretised into 11 layers (de Rosnay et al., 2002). The vertical discretisation for heat diffusion  
is identical to that used for water up to 2 m extended to 90 m with a zero flux condition at the bottom and with 18 calculation  
nodes in order to extrapolate the water content across the entire profile between 2 m and 90 m (Wang et al., 2016). The hydro-  
logical and thermal properties of the soil are determined by soil moisture and texture. The dominant soil texture for each  
model grid cell is derived from the ZOBLER map (Zobler, 1999) using a classification system comprising 3 categories. The  
130 set of equations governing the Soil Organic Matter (SOM) pools and their temporal evolution have analytical solution driven  
by litter input and climate conditions, including soil temperature and humidity (Lardy et al., 2011).

The carbon fixation scheme follows the approach presented by Yin and Struik (2009) based on the FvCB model (Farquhar  
et al., 1980) for C3 plants and Collatz et al. (1991) for C4 plants. The ORCHIDEE LSM uses different types of vegetation  
grouped into Plant Functional Types (PFT) with similar structural characteristics. It distinguishes 14 vegetation PFT classes de-  
135 scribed in Table A1. Each grid point in the model is associated with PFT fractions prescribed using annually varying PFT maps  
derived from ESA's Climate Change Initiative land cover (LC) products and a LC-to-PFT cross-walking approach (Poulter  
et al., 2015) (see <https://orchidas.lscce.ipsl.fr/dev/lccci/>).

In this study, ORCHIDEE is run offline using 3-hour ERA-Interim surface weather forcing fields (Dee et al., 2011) over  
2000-2001, and aggregated to the spatial resolution of the LMDZ atmospheric transport model ( $2.5^\circ$  latitude  $\times$   $3.75^\circ$  longitude)  
140 . The carbon pools are brought to equilibrium following the TRENDY protocol (Sitch et al., 2024). This involves spinning up  
the model for 200 years, employing an analytical spin-up for soil carbon pools to bring them to equilibrium. This process uses a  
constant CO<sub>2</sub> concentration of 1700, no land-use change (LUC), and recycled ERA-Interim meteorological data from 1990 to  
1999, as these are the only years where forcing data is available preceding the assimilation period. This spin-up run is followed  
by a transient simulation to account for the effects of disturbances, varying global atmospheric CO<sub>2</sub> concentration and LUC  
145 from 1800 to 1999, recycling the same meteorological data.

### 2.1.2 LMDZ atmospheric transport model

The atmospheric transport model used in this study is version 3 of the LMDZ General Circulation Model (GCM) (Hourdin  
and Armengaud, 1999). The LMDZ atmospheric model has been widely used to model the climate; it was implemented as the  
atmospheric component of the IPSL Earth System Model (Dufresne et al., 2013). Its derived transport model has been used  
150 to simulate gas, particle chemistry and greenhouse gas distributions in numerous studies (Peylin et al., 2005; Chevallier et al.,  
2005; Locatelli et al., 2015; Remaud et al., 2018). The advection is based on Van Leer scheme (Leer, 1977); the deep convection  
is parametrised following the scheme of (Tiedtke, 1989); and turbulent mixing in the planetary boundary layer is based on a  
second-order local closure formalism (Hourdin and Armengaud, 1999). It uses a horizontal resolution of  $2.5^\circ$  (latitude)  $\times$   $3.75^\circ$   
(longitude) and 19 sigma-pressure layers up to 3 hPa. The calculated winds ( $u$  and  $v$ ) used to drive LMDZ are provided by  
155 ERA-Interim reanalysis meteorological data in order to realistically account for the temporal dependence of meteorological  
events. In this study, we use pre-calculated transport fields, as described in Peylin et al. (2005): they quantify the sensitivity  
of atmospheric concentrations at a given atmospheric station according to the space-time variability of the surface fluxes.

The temporal resolution of the concentration is monthly, taking into account the daily surface fluxes of each grid cell in the model (as shown on the Fig. 1). These pre-calculated transport fields have proven to be very useful - they considerably reduce computing time, given that the model only needs to be run once. They have been used to assimilate atmospheric CO<sub>2</sub> data in a few data assimilation studies (Peylin et al., 2016; Bacour et al., 2023). Although the version of LMDZ used in this study is outdated, the main objective of this work is to develop a framework for atmospheric data assimilation that will support future research using an updated version of LMDZ. Therefore, these pre-calculated transport fields provide a low-cost experiment so as to address the methodological and technical challenges that were previously presented. Nevertheless, it is important to note that the use of these pre-calculated transport fields does not allow for the evaluation of dynamic feedbacks between the surface and the atmosphere that may occur due to parameter changes. The pre-calculated transport fields were originally calculated to assimilate atmospheric CO<sub>2</sub> concentration data using the NOAA Earth System Laboratory's collaborative product (GLOBALVIEW-CO<sub>2</sub>, 2013). They model average monthly concentrations at 53 stations over the period 1990-2009. The stations are located at different altitudes and in different locations on the continents and oceans around the world.

### 2.1.3 Atmospheric stations

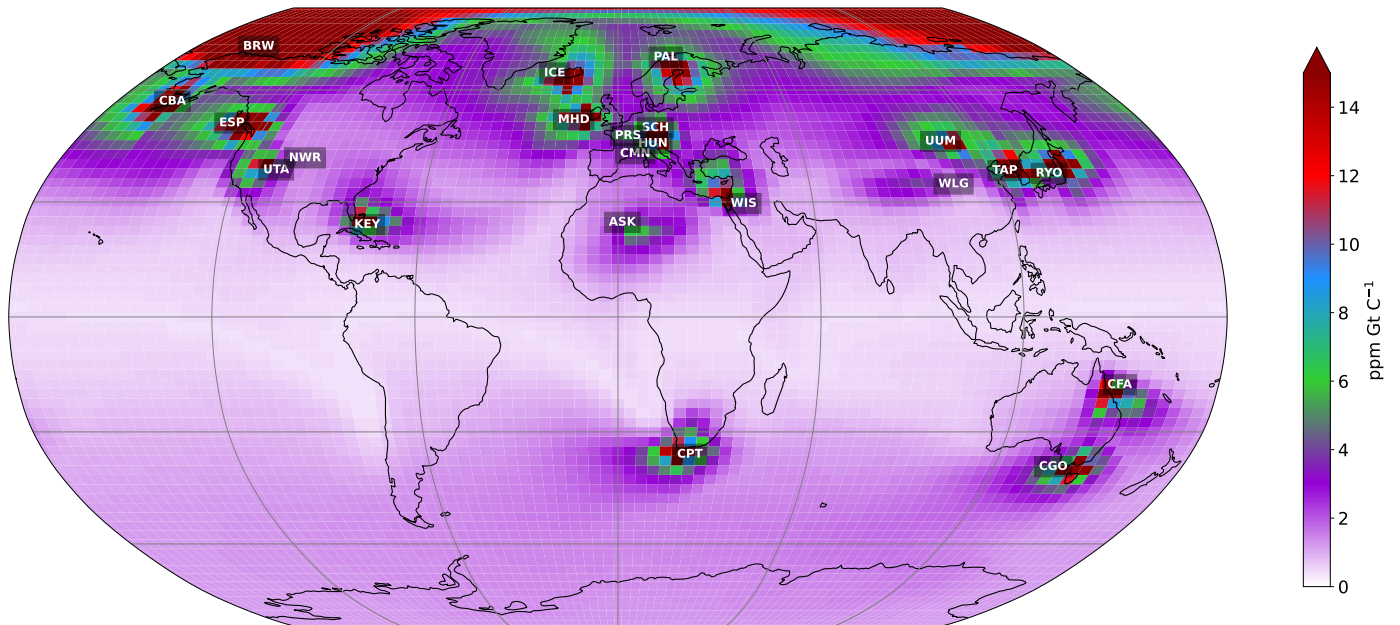
Fig. 1 shows the location of the 21 stations selected for this study. The stations were selected according to their sensitivity to continental fluxes (also shown in Fig. 1) in order to capture the temporal and spatial variations in fluxes over the continental surface. The selected stations are therefore mainly located above the land surface. The other stations, mainly located over the oceans, are less sensitive to continental fluxes, capturing mainly long term variations. As we are only assimilating 2 years of concentrations, we choose not to take them into account. The selected stations also provide a good overview of most PFTs. However, as we can see in Fig. 1, two PFTs appear to be less sensitive to the selected stations: TrBE, which is mainly found in the tropical forests of Amazonia and Central Africa, and BoND, which is mainly found in Siberia.

### 2.1.4 Other components of surface CO<sub>2</sub> fluxes

Other components contributing to the global surface fluxes are not optimised in this study:

- The oceanic flux component was derived from a neural networks model which estimated the spatial and temporal variations in CO<sub>2</sub> fluxes between the air and the sea (Peylin et al., 2016).
- The global maps of biomass burning emissions are taken from the Global Fire Emission Database version 3 (Randerson et al., 2013).
- The fossil fuel CO<sub>2</sub> emission products used here were developed by the University of Stuttgart/IER on the basis of EDGAR v4.2.

All the fluxes used are described in greater detail in previous studies (Peylin et al., 2016; Bacour et al., 2023) and are shown in Fig. A1.



**Figure 1.** Monthly mean sensitivity map of atmospheric CO<sub>2</sub> concentrations to land carbon fluxes at the 21 stations considered over the 2000-2001 period. The average sensitivity map is obtained by deriving, for each atmospheric station and each of the 24 months, the map of the average daily sensitivity of the atmospheric concentration of CO<sub>2</sub> to surface carbon flux (in ppm GtC<sup>-1</sup>) over the last six months, and then calculating the average of the 24 maps. The colour of the pixel indicates the influence of the surface fluxes given by the pixel on the atmospheric concentration of CO<sub>2</sub>, depending on the station. Red indicates a very strong influence of surface fluxes. The blue, green and violet colours indicate different influences, from strong to weak. White indicates no influence from surface fluxes (see full detail of the stations <https://gml.noaa.gov/dv/site/index.php> [Last access : 5 June 2025]).

## 2.2 Data assimilation framework

### 2.2.1 A Bayesian setup

190 First, let us define a general Bayesian framework, mainly following Tarantola (1987, 2005), that accounts both for model/observation error and an *a priori* background error. Taking the approach of Kennedy and O'Hagan (2001), for an observational

constraint  $\mathbf{y}$ , let

$$\mathbf{y} = \mathcal{Y} + \mathbf{e} \quad (1)$$

where  $\mathcal{Y}$  represents the relevant aspect of the observed system and  $\mathbf{e}$  represents the error on that observation, often due to instrument error but can include any error in the derivation of the data product. Let  $\mathcal{H}$  represent the model operator that takes the parameter vector  $\mathbf{x}$  as input. We then assume that there exists an input  $\mathbf{x}^*$  such that:

$$\mathbf{y} = \mathcal{Y} + \mathbf{e} = \mathcal{H}(\mathbf{x}^*) + \boldsymbol{\eta} + \mathbf{e} \quad (2)$$

where  $\boldsymbol{\eta}$  represents the model error, given an imperfect model. Here, the model operator output  $\mathcal{H}(\mathbf{x}^*)$  and the observation  $\mathbf{y}$  are defined in time and space. All observations are concatenated into a large vector of observations  $\mathbf{y}$ , in order to represent all observations available in a given time window. The same operation is performed for the output of operator  $\mathcal{H}(\mathbf{x}^*)$ . Note that, given no additional information about the errors, we assume that i)  $\mathbf{e}$  and  $\boldsymbol{\eta}$  are independent of  $\mathcal{Y}$  and  $\mathcal{H}(\mathbf{x})$  respectively and ii) both are random vector quantities following a multivariate normal distribution with a mean equal to 0 and a covariance matrix  $\boldsymbol{\Sigma}_i$  such that  $\mathbf{e} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_e)$  and  $\boldsymbol{\eta} \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\eta)$ . Furthermore, we assume that the parameter vector  $\mathbf{x}$  and the model/observation likelihood  $\mathbf{y}|\mathbf{x}$  both follow Gaussian multivariate distributions:

$$p(\mathbf{y}|\mathbf{x}) \propto \exp \left[ -\frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}) \right]; \quad p(\mathbf{x}) \propto \exp \left[ -\frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \right], \quad (3)$$

where  $\mathbf{x}_b$  represents prior knowledge of the parameter vector and  $\mathbf{B}$  and  $\mathbf{R}$  are respectively the covariance error matrix for the parameters vector and for the model/observation such that  $\mathbf{R} = \boldsymbol{\Sigma}_\eta + \boldsymbol{\Sigma}_e$ . We seek to find the posterior distribution  $p(\mathbf{x}|\mathbf{y})$  which quantifies the probability of parameters given the observations using Bayes' theorem:

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \propto \exp \left[ -\frac{1}{2}(\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1}(\mathcal{H}(\mathbf{x}) - \mathbf{y}) - \frac{1}{2}(\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}_b) \right]. \quad (4)$$

### 2.2.2 The ~~4DVar~~-VarDA method

#### Standard ~~4DVar~~VarDA

In this Section, we present the ~~4DVar-assimilation~~-VarDA method. Maximising the probability in equation (4) is equivalent to minimising the following function, usually referred to as the ~~4DVar~~-VarDA cost function:

$$J(\mathbf{x}) = \frac{1}{2} \left( \sum_t^{N_t} \mathcal{H}_t(\mathbf{x}_{\underline{t}}) - \mathbf{y}_t \right)^T \mathbf{R}_t^{-1} (\mathcal{H}_t(\mathbf{x}_{\underline{t}}) - \mathbf{y}_t) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) \quad (5)$$

where  $t$  refers to time steps  $0, \dots, N_t$ . Since the parameter must be constant over time, we consider only a single time window that includes all observation vector  $\mathbf{y}$  (in time and space). We therefore simplify ~~to the compact form the initial 4DVar the initial VarDA~~ cost to the compact form:

$$J(\mathbf{x}) = \frac{1}{2} (\mathcal{H}(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1} (\mathcal{H}(\mathbf{x}) - \mathbf{y}) + \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b). \quad (6)$$

where, for example, the concatenated vector  $y = (y_0, y_1, \dots, y_N)^T$  represents all available observations at all times over the time window. The minimum of the equation 6 can be reached iteratively using a descent algorithm that requires the computation of the gradient of  $J$  with respect to the parameter vector  $x$ . In addition, when the model is non-linear, it is common to use the quasi-Newton method to optimise the parameters vector:

$$x_{i+1} = x_i - (\nabla^2 J(x_i))^{-1} \times \nabla J(x_i). \quad (7)$$

The gradient of the cost function,  $\nabla J(x_i)$ , and the square matrix of partial second derivatives of the cost function (called the Hessian matrix),  $\nabla^2 J(x_i)$ , can be calculated as follows:

$$\nabla J(x_i) = \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H}x_i - \mathbf{y}) + \mathbf{B}^{-1} (x_i - x_b); \quad \nabla^2 J(x_i) = \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1}. \quad (8)$$

We can update equation (7) using (8):

$$x_{i+1} = x_i - [\mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} + \mathbf{B}^{-1}]^{-1} [\mathbf{H}^T \mathbf{R}^{-1} (\mathbf{H}x_i - \mathbf{y}) + \mathbf{B}^{-1} (x_i - x_b)]. \quad (9)$$

Here, the notation  $\mathcal{H}$  becomes  $\mathbf{H}$ , because it does not represent the use of the direct operator  $\mathcal{H}$ . Instead, we use the tangent linear model  $\mathbf{H}$  and the adjoint model  $\mathbf{H}^T$ . Usually, these two terms are coded directly, but for complex models, it is usually very difficult to code and maintain these terms, especially when the model is subject to many developments (which means that they quickly become obsolete).

#### Epsilon-based ~~4DVar~~-VarDA variant: ~~$\epsilon$ -4DVar~~-VarDA

To approximate the tangent linear and adjoint models, we can use finite differences:

$$\mathbf{H} = \frac{\mathcal{H}(x + \Delta x) - \mathcal{H}(x)}{\Delta x} \quad (10)$$

where  $\Delta x$  represents a small change in  $x$ . This estimate will not be as accurate as the exact tangent linear and adjoint models, but it can still help us in our minimisation objective. The accuracy of the tangent linear and adjoint models is then completely dependent on the choice of  $\Delta x$ . A selection of  $\Delta x$  that is too small may lead to  $\mathbf{H}$  being insensitive to the parameter vector, i.e.  $\mathcal{H}(x + \Delta x) - \mathcal{H}(x) = \mathbf{H}\Delta x \approx 0$ . This leads to the term corresponding to the difference between the observation and the output's operator ( $\mathbf{H}x_i - \mathbf{y}$ ) becoming negligible in equation (8) and hence resulting in an ineffective minimisation. By contrast, if the choice of  $\Delta x$  is too large, the result gives inaccurate tangent linear and adjoint models that lose their local vision around  $x$ . This results in a large loss of information and therefore a much less accurate minimisation. In our case, we define  $\epsilon$  such that:  $\Delta x = x_{range} * \epsilon$ , where  $x_{range} = x_{max} - x_{min}$  and we will refer to this method as  ~~$\epsilon$ -4DVar~~-VarDA. Due to this approximation,  ~~$\epsilon$ -4DVar~~-VarDA is therefore not entirely equivalent to standard ~~4DVar~~VarDA.

### 2.2.3 The ~~4DEnVar~~-EnVarDA method

#### From ~~4DVar~~-VarDA to ~~4DEnVar~~EnVarDA

We present here an implementation of ~~4DVar~~-VarDA that we do not use in this study, but that is important to understand the ~~4DEnVar~~EnVarDA method. This implementation is presented in several studies (Courtier et al., 1994; Gilbert and Lemaréchal,

1989; Liu et al., 2008; Bannister, 2017; Pinnington et al., 2020), and can be applied when the prior error covariance matrix  $\mathbf{B}$  becomes large and difficult to invert. It is possible to introduce a matrix  $\mathbf{U}$  and a vector  $\mathbf{w}$  to ensure that the ~~4DVar~~ VarDA cost function converges as efficiently as possible and avoids the explicit calculation of the matrix  $\mathbf{B}$  given by:

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T \quad (11)$$

and

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{U}\mathbf{w} \quad (12)$$

where  $\mathbf{x}_a$  represents the posterior value of the parameter vector. Consequently, this changes the  $J$  cost function, which is presented in detail in Courtier et al. (1994):

$$J(\mathbf{w}) = \frac{1}{2}(\mathbf{H}\mathbf{U}\mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \frac{1}{2}\mathbf{w}^T \mathbf{w} \quad (13)$$

and its gradient:

$$\nabla J(\mathbf{w}) = \mathbf{U}^T \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \mathbf{w}. \quad (14)$$

~~4DEnVar~~ EnVarDA

The ~~4DEnVar~~ The EnVarDA method described in Liu et al. (2008) and Pinnington et al. (2020) proposes to incorporate an aspect of the ensemble Kalman Filter (EnKF) in order to avoid the calculation of tangent linear or adjoint models necessary for ~~4DVar~~ VarDA. The EnKF is a Kalman filter, but uses a set of  $N$  parameter vectors, also known as ensemble members, to estimate the prior error covariance matrix  $\mathbf{B}$  (Evensen, 1994). A perturbation matrix:

$$\mathbf{X}'_b = \frac{1}{\sqrt{N-1}} (\mathbf{x}_1 - \mathbf{x}_b; \mathbf{x}_2 - \mathbf{x}_b; \dots; \mathbf{x}_N - \mathbf{x}_b). \quad (15)$$

where the ensemble members  $\mathbf{x}_i$  for  $i = 1, \dots, N$  are generated according to a multivariate normal distribution using  $\mathbf{x}_b$  as the mean and  $\mathbf{B}$  as the covariance matrix:  $\mathcal{N}(\mathbf{x}_b, \mathbf{B})$ . It follows that:

$$\mathbf{B} \approx \mathbf{X}'_b \mathbf{X}'_b{}^T. \quad (16)$$

Using the same logic as equation (12), we can use the perturbation matrix as follows:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{X}'_b \mathbf{w} \quad (17)$$

where  $\mathbf{w}$  is a vector of length  $N$ . The cost function in Equation (13) is updated accordingly:

$$J(\mathbf{w}) = \frac{1}{2}(\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y})^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \frac{1}{2}\mathbf{w}^T \mathbf{w}, \quad (18)$$

and the gradient in Equation (14) becomes:

$$\nabla J(\mathbf{w}) = \mathbf{X}'_b{}^T \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{X}'_b \mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \mathbf{w}. \quad (19)$$

275 Note that the minimisation problem changes. In both case we try to balance the cost function between the background term and the observation term but we no longer aim to find  $\mathbf{x}$  such that  $\mathcal{H}(\mathbf{x}) \approx \mathbf{y}$ , but we now look for  $\mathbf{w}$  that determines the linear combination  $\mathbf{H}\mathbf{X}'_{\mathbf{b}}\mathbf{w}$  which is equal to the distance  $\delta\mathbf{y}$  such that  $\delta\mathbf{y} \approx \mathcal{H}(\mathbf{x}) - \mathbf{y}$ . The  $\mathbf{H}\mathbf{X}'_{\mathbf{b}}$  term can be approximated by applying the  $\mathcal{H}$  operator to each parameter vector  $\mathbf{x}$  present in  $\mathbf{X}'_{\mathbf{b}}$  :

$$\mathbf{H}\mathbf{X}'_{\mathbf{b}} \approx \frac{1}{\sqrt{N-1}} (\mathcal{H}(\mathbf{x}_1) - \mathcal{H}(\mathbf{x}_b); \mathcal{H}(\mathbf{x}_2) - \mathcal{H}(\mathbf{x}_b); \dots; \mathcal{H}(\mathbf{x}_N) - \mathcal{H}(\mathbf{x}_b)). \quad (20)$$

280 where each  $\mathcal{H}(\mathbf{x}_i)$  is a concatenated vector of extracted simulations to correspond with all observations available at all times across the time window. Each coefficient  $w_i$  of  $\mathbf{w}$  multiplies a vector  $\mathcal{H}(\mathbf{x}_i) - \mathcal{H}(\mathbf{x}_b)$  present in the approximation of  $\mathbf{H}\mathbf{X}'_{\mathbf{b}}$  which represents the distance between a member of the ensemble and the prior information. The optimisation of  $\mathbf{w}$  is performed so that the linear combination  $\mathbf{H}\mathbf{X}'_{\mathbf{b}}\mathbf{w}$  converges around  $\delta\mathbf{y}$  and taking into account the background terms. Once optimised, the vector  $\mathbf{w}$  can be used for another linear combination  $\mathbf{X}'_{\mathbf{b}}\mathbf{w}$ , this time in the input space. This gives  $\mathbf{x}_a$ , the posterior value of the parameter vector, that can be obtained using equation (17). The great advantage of this method lies in the way the gradient is computed. In particular, the term  $\mathbf{X}'_{\mathbf{b}}{}^T\mathbf{H}^T$ , which is equivalent to  $(\mathbf{H}\mathbf{X}'_{\mathbf{b}})^T$ . This equivalence makes it possible to rewrite the gradient by "simply" transposing the matrix  $\mathbf{H}\mathbf{X}'_{\mathbf{b}}$  :

$$\nabla J(\mathbf{w}) = (\mathbf{H}\mathbf{X}'_{\mathbf{b}})^T \mathbf{R}^{-1} (\mathbf{H}\mathbf{X}'_{\mathbf{b}}\mathbf{w} + \mathcal{H}(\mathbf{x}_b) - \mathbf{y}) + \mathbf{w}. \quad (21)$$

Subsequently, tangent linear and adjoint models are no longer required. The subjective choice here is no longer related to the choice of the  $\epsilon$  that estimates the tangent linear and adjoint models, but to the number  $N$  of ensemble members used to generate  $\mathbf{X}'_{\mathbf{b}}$  and  $\mathbf{H}\mathbf{X}'_{\mathbf{b}}$ . A posterior ensemble can be obtained as it is described by Douglas et al. (2025) by calculating  $\mathbf{X}_a'$  where

$$\mathbf{X}_a' = \mathbf{X}_b' \left( \mathbf{I} + (\mathbf{H}\mathbf{X}_b')^T \mathbf{R}^{-1} \mathbf{H}\mathbf{X}_b' \right)^{-\frac{1}{2}}. \quad (22)$$

#### 2.2.4 Implementation into ORCHIDAS

The ORCHIDEE Data Assimilation System (ORCHIDAS) is a system desgined to calibrate the parameters of ORCHIDEE and is developed in Python. It has been used for over 15 years (MacBean et al., 2022) mainly for studies focusing on the carbon cycle and other terrestrial cycles such as water and energy budget, methane and nitrogen (see the full list of studies published at <https://orchidas.lscce.ipsl.fr/publications.php>).

This system has long used ~~4DVar~~ VarDA as described in Section ~~??2.2.2~~, but it also allows the use of several methods such as genetic algorithms (Bastrikov et al., 2018) or history matching (Raoult et al., 2024a). ORCHIDAS facilitates the testing of various data assimilation methods while maintaining a consistent configuration for ORCHIDEE execution. In this study, we implemented the ~~4DEnVar~~ EnVarDA method as described in Section ~~??2.2.3~~.

### 2.3 ~~Twin experiments~~ Experiment Design

#### 2.3.1 Twin Experiment Description

To test the ~~two data assimilation~~ methods presented in Section 2.2, we ~~use conducted~~ a so-called *twin experiment* ~~. This experiment significantly reduces all the~~ to evaluate their efficiency in calibrating parameters involved in calculating NBP

fluxes in the ORCHIDEE LSM model. This experimental framework reduces complexities associated with model-data errors and focuses on the efficiency, focusing on the performance of the assimilation method. In this methods. The known 'true' parameters being the default parameter values of the ORCHIDEE model are used to generate the synthetic observations. New values of *twin experiment, we aim to optimise the a priori* parameters are manually generated, ensuring physically meaningful values that differ from the 'true' parameters both presented in Table A2. The assimilation methods are then applied to assess how closely they converge toward the known solution (standard parameter values). The synthetic observations of atmospheric CO<sub>2</sub> concentrations from the 21 continental stations are assimilated simultaneously over a two-year window (2000–2001) to monitor spatial and temporal variations in carbon fluxes, as shown in Figure A4. A limited period was chosen for practical reasons to avoid computationally expensive simulations.

### 2.3.2 Generation of Synthetic Observations

To generate synthetic observations for the twin experiment, we simulate net biome productivity (NBP) fluxes of at the global scale using the ORCHIDEE LSM by calibrating the parameters involved in their calculation. The with default parameter values, referred to as the 'true' parameters (see Table A2). These NBP fluxes represent the net carbon fluxes of the land component, i.e. calculated as the difference between the emission fluxes of emission fluxes (heterotrophic and autotrophic respiration as well as the, and disturbance fluxes due to LUC and the sink fluxes mainly land-use change) and sink fluxes (primarily due to photosynthesis. To do this, we simulate the NBP fluxes at the global scale using the ORCHIDEE LSM with the default parameter values (see Tab. A2) – which we will refer to as 'true' in the remainder of this article – and include the. The concentration given by the surface fluxes (the simulated NBP fluxes, along with other fluxes described in Section 2.1.4 over the year 2000–2001. We then transport the concentration given by the surface fluxes using the) are transported using pre-calculated transport fields of LMDZ. We focus on the LMDZ model over the period 2000–2001. We then extract atmospheric CO<sub>2</sub> concentrations at 21 continental atmospheric stations, as shown in Fig. 1. These stations shown in Figure 1, which are highly sensitive to carbon fluxes over the continents, which provides a significant constraint on continental fluxes and therefore on our continental carbon fluxes, providing significant constraints on the parameters. This has enabled us to generate 'synthetic' process enabled the generation of synthetic observations of monthly average atmospheric CO<sub>2</sub> concentrations given by at these 21 atmospheric stations between 2000 and 2001. A limited period was chosen for practical reasons – to avoid expensive simulations. A new vector of *a priori* parameters was generated manually, ensuring that it differed from the 'true' parameter values while retaining physically meaningful values. We then applied the various assimilation methods to see how closely they converge towards the known solution (standard parameter values). The assimilation of atmospheric CO<sub>2</sub> concentrations at the 21 stations is performed simultaneously using a two-year assimilation window in order to assimilate all observations and thus monitor variations in carbon fluxes in space and time, as shown in Fig. A4.

stations over the two-year period. It is important to note that the steps taken here to generate the synthetic observations are exactly the same as those used to perform a simulation. This means that there is at least one solution where the model can perfectly match the synthetic observation.

### 2.3.3 Simplified case

340 First, we focus on a simplified case involving the calibration of only one PFT-dependant parameter:  $V_{\text{cmax}}$ , which controls the maximum rate of carboxylation limited by Rubisco activity at 25°C. This parameter was chosen because its impact on the atmospheric  $\text{CO}_2$  concentration is well understood: when its value increases, the quantity of carbon absorbed by photosynthesis increases and atmospheric concentrations decrease - and vice versa. The aim of the assimilation is to recover the 'true' values of  $V_{\text{cmax}}$  for the 14 PFTs resulting in the calibration of 14 parameters. This simplified case is very useful to perform several  
345 tests allowing for a better understanding of the behaviour of the different data assimilation methods.

### 2.3.4 Complex case

To assess the performance of the different approaches in conditions resembling real cases, we perform another twin experiment in which we calibrate four PFT-dependent parameters and one global parameter involved in different bio-geophysical processes. The parameters selected was already optimised in previous data assimilation studies using atmospheric  $\text{CO}_2$  concentrations  
350 (Peylin et al., 2016; Bacour et al., 2023). In addition to  $V_{\text{cmax}}$ , we choose:

- the PFT-dependent parameter  $\text{SLA}$  (Specific Leaf Area) that impacts leaf biomass and hence ecosystem photosynthetic capacity;
- the global parameter  $Q_{10}$  which controls the thermal dependence of heterotrophic respiration;
- the PFT-dependent parameter  $m_{\text{maint.resp}}$  that defines the slope of the maintenance respiration coefficient, which controls autotrophic respiration;  
355
- the PFT-dependent parameter  $\text{LAI}_{\text{max}}$  which controls the maximum leaf area index for carbon allocation. Once the  $\text{LAI}$  reaches  $\text{LAI}_{\text{max}}$ , no carbon is allocated to the leaf. It impacts the vegetation biomass and therefore acts on both photosynthesis and respiration.

A total of 57 ( $14 \times 4 + 1$ ) parameters are being calibrated. As they interact within the same modelled processes, the degree of  
360 equifinality is significant.

### 2.3.5 Error covariance matrices

~~We need to define the two~~ To implement the two data assimilation methods,  $\epsilon$ -VarDA and EnVarDA, we define two error covariance matrices,  $\mathbf{R}$  and  $\mathbf{B}$ , ~~in order to use the two data assimilation methods. Since~~. These matrices are configured to be diagonal, as we are assimilating 'synthetic' observations, these matrices can be diagonal. "synthetic" observations, and are  
365 common to both methods to ensure comparable experiments. Their configurations are informed by previous data assimilation studies using ORCHIDEE and a simplified carbon model (Kuppel et al., 2012, 2013; Bastrikov et al., 2018; MacBean et al., 2016), with Peylin et al. (2016) specifically applying diagonal matrices for atmospheric  $\text{CO}_2$  observations.

#### R Matrix

The  $\mathbf{R}$  matrix ~~is used to represent the model/observation error~~ represents the model-structural and observation errors. In our case, we define small diagonal terms of 0.01 ppm for the  $\mathbf{R}$  matrix as we are in a perfect model scenario. Indeed since no errors were included in the pseudo-observation and we are in a twin experiment, twin experiment setup, we choose to add only very small errors in the synthetic observations to compare both methods in an ideal case. In this context, structural errors in the ORCHIDEE LSM (i.e. missing processes, etc.) and in the transport model (i.e., coarse spatial resolution, wind biases, etc.), or measurement errors are discarded. Indeed, since the synthetic observations are generated by a simulation, as detailed in Section 2.3.2, there exists at least one solution where all observations can be matched perfectly. For this reason, we use a simplified  $\mathbf{R}$  matrix ~~was used. The~~ with the same small diagonal terms of 0.01 ppm for all stations. The rationale behind this choice is that, as all stations can be matched perfectly, we do not want to introduce any spatial or temporal preferences.

## B Matrix

The  $\mathbf{B}$  matrix ~~contains~~ represent the background errors associated with ~~the~~ prior knowledge of the parameters. We set ~~an error~~ corresponding the error to 30% of the parameter range for the simple case and ~~20%~~ 20% for the complex case (as we use larger parameter ranges ). ~~To ensure that the experiments are comparable, the  $\mathbf{R}$  and  $\mathbf{B}$  matrices are common to the two methods:  $\epsilon$ -4DVar and 4DEnVar . The configuration of the  $\mathbf{R}$  and  $\mathbf{B}$  matrices was based on previous data assimilation studies with ORCHIDEE and a simplified carbon model (Kuppel et al., 2012, 2013; Peylin et al., 2013; Bastrikov et al., 2018; MacBean et al., 2016) . These studies employed diagonal matrices for  $\mathbf{R}$  and  $\mathbf{B}$  to assimilate in-situ observations, while (Peylin et al., 2016) specifically used them for atmospheric  $\text{CO}_2$  observations for this case). The background errors of each parameter can be seen in Fig. 3 for the simple case and in Fig. 6 as well as Table A2.~~

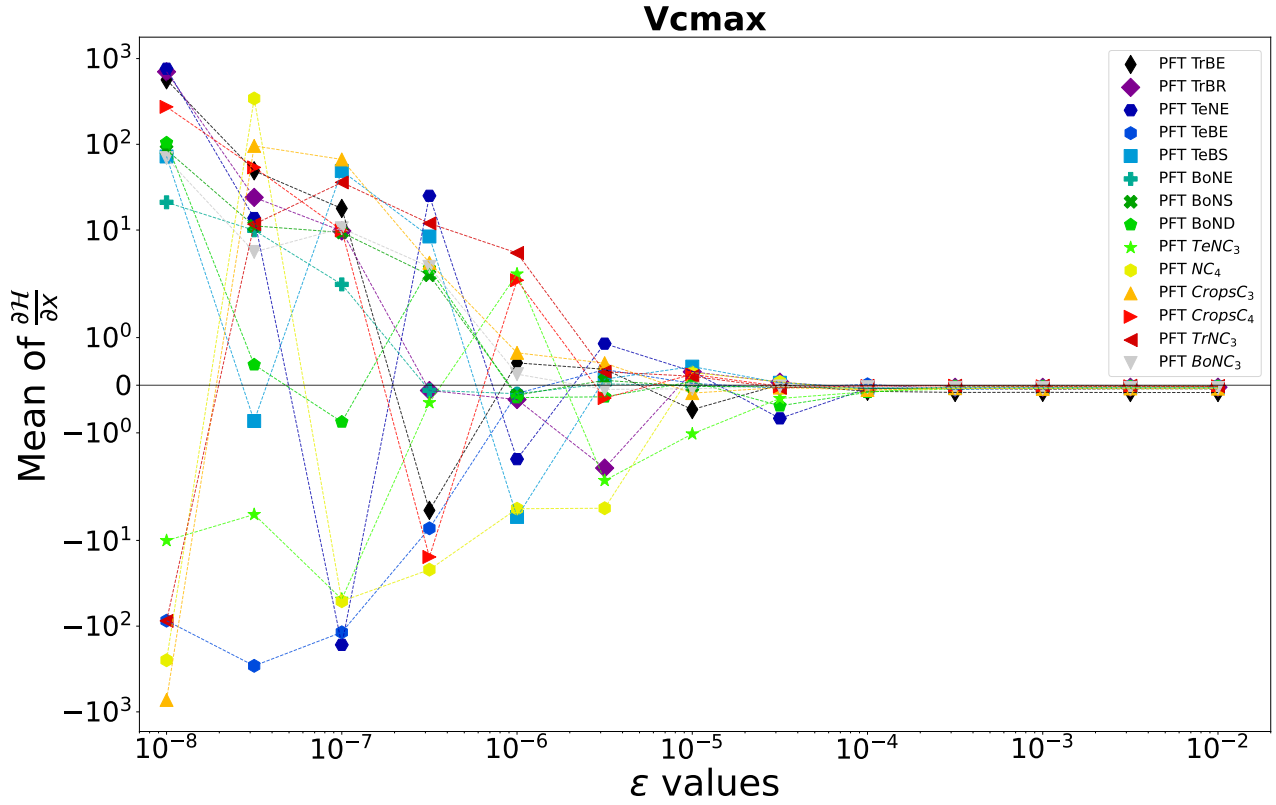
## ~~2.4 Tuning $\epsilon$ for gradient calculation~~

### 2.3.1 Tuning $\epsilon$ for gradient calculation

As explained in Section ~~??2.2.2~~, the choice of  $\epsilon$  is essential for effective  ~~$\epsilon$ -4DVar~~ VarDA performance. One way to select an appropriate  $\epsilon$  is to perform a  $\epsilon$ -test which calculates the partial derivative of  $\mathcal{H}$  for each of the parameters and using different  $\epsilon$ . We calculate the partial derivative as follows:

$$\frac{\partial \mathcal{H}}{\partial x} = \frac{\mathcal{H}(x + \Delta x) - \mathcal{H}(x)}{\Delta x} \quad (23)$$

where  $\epsilon$  defines  $\Delta x$  as explained in Section ~~??2.2.2~~. By changing the  $\epsilon$  we change  $\Delta x$  and we can seek to find the value of  $\epsilon$  for which the derivative becomes stable. Fig. 2 shows the sensitivity of  $\epsilon$  ranging from  $10^{-8}$  to  $10^{-2}$  on the calculates the partial derivative of each  $\mathbf{Vcmax}$ . We see that the partial derivative of  $\mathbf{Vcmax}$  is unstable with an  $\epsilon$  below  $10^{-3}$  for all PFT. Therefore, we need a value of  $\epsilon$  greater than  $10^{-3}$  to ensure correct gradient calculation with respect to the  $\mathbf{Vcmax}$  parameter. Table A3 shows the values of the mean of the partial derivatives for all parameters and PFTs using an  $\epsilon$  allowing for a stable derivative. This also allows us to check the consistency of the derivation calculation. For example, the increase in  $\mathbf{Vcmax}$  leads to an increase in the photosynthetic capacity and subsequently in the carbon uptake by ~~vegetations~~ vegetation. This leads to a reduction in atmospheric  $\text{CO}_2$  concentration. We can see in Table A3 that the values obtained for  $\mathbf{Vcmax}$  are negative



**Figure 2.**  $\epsilon$ -test: Spatial and temporal average of partial derivative of  $\mathcal{H}$  as a function of  $\epsilon$ . The partial derivative of  $\mathcal{H}$  is calculated with respect to the parameter  $V_{cmax}$  for each PFT. It is calculated on the concentration space using every station over 2 years. The mean of the partial derivative is then calculated over space and time in order to visualise the local derivative. The derivative of  $\mathcal{H}$  is calculated for several  $\epsilon$ .

which is the expected response. The same  $\epsilon$ -test was carried out for the four other parameters used in the complex case, and the results are shown in Fig. A2 and in Table A3 :

- The partial derivative of **SLA** diverges with an  $\epsilon$  below  $10^{-3}$  for all PFT. **SLA** has the same impact that  $V_{cmax}$  has on atmospheric  $CO_2$  concentration, so the negative mean values obtained are expected;
- The partial derivative of **Q10** does not diverge for any values of  $\epsilon$ . The mean value of its derivation is negative as expected. Increasing **Q10** increases the thermal dependence of heterotrophic respiration and consequently reduces it; with less heterotrophic respiration the atmospheric  $CO_2$  concentration decreases;
- The partial derivative of **m<sub>maint.resp</sub>** diverges with different  $\epsilon$  values depending on the PFT, ranging from  $10^{-5}$  for PFT TrBE to  $10^{-2}$  for PFT CropsC<sub>4</sub>. This may be due to different distributions and proportions of PFTs (see Table A1).

410 However, the mean values at  $10^{-2}$  are all positive.  $\mathbf{m}_{\text{maint.resp}}$  has an impact on autotrophic respiration, increasing this parameter increases vegetation respiration and therefore the atmospheric  $\text{CO}_2$  concentration;

- The partial derivative of  $\text{LAI}_{\text{max}}$  diverges with an  $\epsilon$  below  $10^{-2}$  for all PFT. Determining the sign of the mean values of the partial derivative of this parameter is not trivial here.  $\text{LAI}_{\text{max}}$  influences vegetation biomass and therefore photosynthesis and respiration. All PFTs gives a negative mean values for their partial derivative, only the PFT TrBR gives  
415 a positive mean value.

## 2.4 Defining the impact of the configuration

### 2.3.1 Defining the impact of the configuration

For both methods,  $\epsilon$ -4DVar and 4DEnVar-VarDA and EnVarDA, the configuration used plays an important role in the quality of the minimisation of the associated cost function and so the calibration of the parameter. Whether it is the choice of  $\epsilon$  for  
420 the  $\epsilon$ -4DVar-VarDA or the number of members used to generate the ensemble in the 4DEnVar-EnVarDA, it is up to the user to make a choice that can only be subjective. To assess their impact, we launch the twin experiment using different configurations:

- for the simple case:
  - 5 different values of  $\epsilon$  for the  $\epsilon$ -4DVar-VarDA based on the sensitivity test presented in Section 2.3.1 ;
  - 5 different ensemble sizes in the 4DEnVar-EnVarDA;
- 425 – for the complex case:
  - 5 different ensemble sizes in the 4DEnVar-EnVarDA;
  - 1 different values of  $\epsilon$  for the  $\epsilon$ -VarDA.

For the complex case using  $\epsilon$ -4DVar-VarDA, the  $\epsilon$  is selected relative to the results in the simple case and Fig. A2. To re-tune  $\epsilon$  for each parameter requires too many simulations and so it is not feasible for the complex case.

430 For each minimisation, the L-BFGS-B (limited memory Broyden–Fletcher–Goldfarb– Shanno) algorithm with bound constraints; (Byrd et al., 1995)) algorithm is used. For the  $\epsilon$ -4DVar-VarDA, we set a maximum number of iterations at 40, due to computing costs. Indeed, each iteration requires  $N_{\text{param}} + 1$  model simulations. In each case, a solution is reached after 20 iterations (subsequent iterations are only minor corrections of the solution obtained). For 4DEnVar-EnVarDA, no maximum iteration limit is chosen, since an iteration does not requires further simulation of the model (all required information is con-  
435 tained in the pre-calculated ensemble). We can therefore wait for the L-BFGS-B minimiser to converge, i.e. until the gradient becomes null.

3 Results

3.1 Comparing the different configurations

440 The results in terms of 1) mean reduction in root mean square difference (RMSD) calculated between the pseudo-observation and the simulation over the two years of the assimilation window for the 21 atmospheric stations, 2) Mean Absolute Differences (MAD) on parameter space and 3) computational demand of each experiment using the simple case are summarised in Table 1. We see that for the  $\epsilon$ -4DVar-VarDA method the best results are obtained with an  $\epsilon$  equal to  $5 * 10^{-2}$  where the mean RMSD reduction is 82.3% and the MAD score is 1.7. The best results for the 4DEnVar-EnVarDA method are obtained using an ensemble of 100 members where the mean RMSD reduction is 97% and the MAD score is 0.3. These two configurations are therefore considered for the simple case of the twin experiment in Section 3.2.1.

**Table 1.** Mean RMSD reduction score between "synthetic" observation and posterior simulation of the atmospheric CO<sub>2</sub> concentration at 21 atmospheric stations, Mean Absolute Difference (MAD) score computed between the "true" parameter values used to generate the "synthetic" observations and the posterior parameters and number of simulations used for each configuration of  $\epsilon$ -4DVar-VarDA and 4DEnVar-EnVarDA for the simple case.

| $\epsilon$ -4DVar-VarDA | Epsilon       | Mean RMSD reduction | MAD score | Number of ORCHIDEE simulations needed |
|-------------------------|---------------|---------------------|-----------|---------------------------------------|
|                         | $10^{-1}$     | 79.7%               | 1.84      | 300                                   |
|                         | $5 * 10^{-2}$ | 82.3%               | 1.7       | 300                                   |
|                         | $10^{-2}$     | 75.1%               | 2.05      | 300                                   |
|                         | $5 * 10^{-3}$ | 73.1%               | 1.91      | 300                                   |
|                         | $10^{-3}$     | 69.5%               | 2.0       | 300                                   |
| 4DEnVar-EnVarDA         | Ensemble      |                     |           |                                       |
|                         | 50            | 81.1%               | 0.44      | 50                                    |
|                         | 75            | 91.8%               | 0.67      | 75                                    |
|                         | 100           | 97.0%               | 0.3       | 100                                   |
|                         | 150           | 91.0%               | 0.34      | 150                                   |
|                         | 200           | 96.3%               | 0.29      | 200                                   |

445

For the complex case, results are presented in Table 2. We see that the best results for the 4DEnVar-EnVarDA method are obtained with an ensemble of 300 members giving a mean RMSD reduction of 94.4%. This configuration is considered for the complex case in Section 3.2.2 using the 4DEnVar-EnVarDA method.

**Table 2.** Mean RMSD reduction score between "synthetic" observations and posterior simulation of the atmospheric CO<sub>2</sub> concentration at 21 atmospheric stations for each configuration of the 4DEnVar-EnVarDA method for the complex case

| <u>4DEnVar-EnVarDA</u> | <i>Ensemble</i> | Mean RMSD reduction |
|------------------------|-----------------|---------------------|
|                        | 100             | 81%                 |
|                        | 200             | 89.8%               |
|                        | 300             | 94.4%               |
|                        | 350             | 94.0%               |
|                        | 400             | 90.9%               |

### 3.2 Comparing $\epsilon$ -4DVar-VarDA and 4DEnVar-EnVarDA

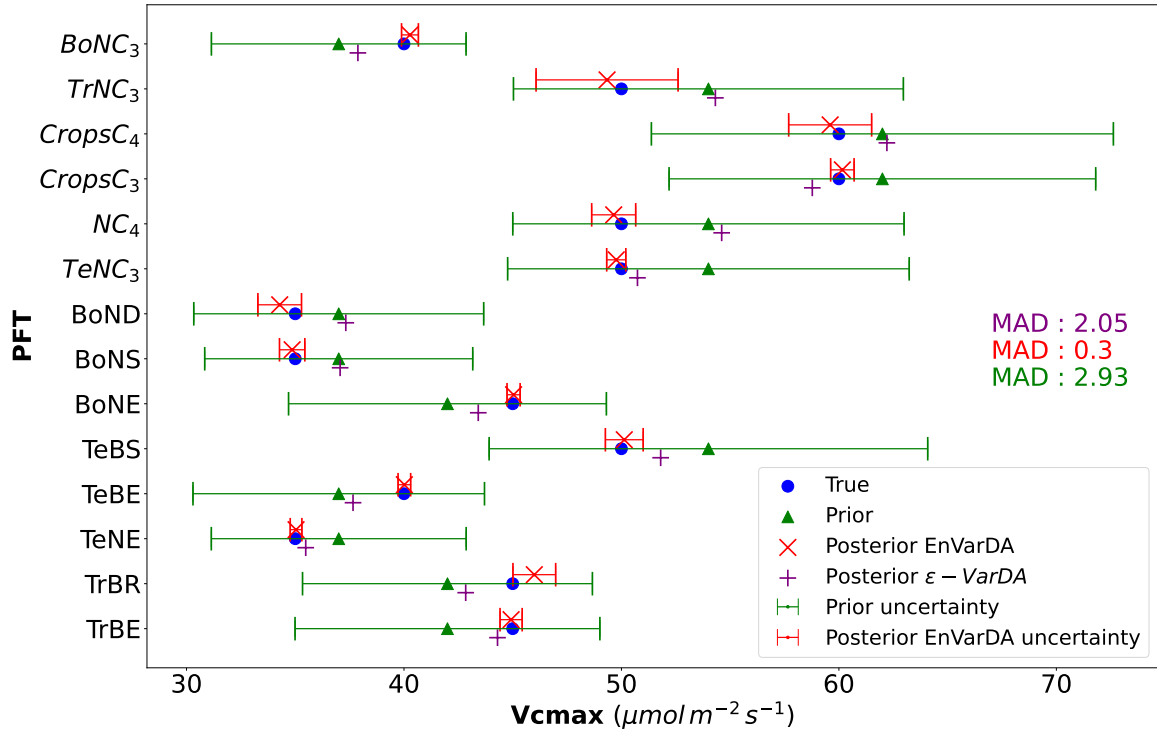
#### 3.2.1 Simple case

Fig. 3 and Fig. 4 compare the results obtained for the  $\epsilon$ -4DVar and the 4DEnVar-VarDA and the EnVarDA methods using the configurations chosen in Section 3.1. Fig. 3 shows that the parameter values obtained by the 4DEnVar-EnVarDA method is almost equal to the "true" parameters used to generate the "synthetic" observations with a mean absolute difference (MAD) score of 0.3. This shows that the 4DEnVar-EnVarDA method is able to almost recover the "true" parameters. The parameter values obtained by the  $\epsilon$ -4DVar-VarDA method have a MAD score of 2.05, which reduces the prior MAD by 30% but remains far from the "true" parameter values. Only the **V**max of PFTs TeNC<sub>3</sub>, TeNE and TrBE are close to the "true" value of the parameters whereas PFTs BoNC<sub>3</sub>, BoNE, TeBS, TeBE and TrBR give values that are between the prior and the "true" value; the **V**max of other PFTs have either maintained or increased the distance between the prior and the "true" values. This shows that the  $\epsilon$ -4DVar-VarDA method falls into a local minimum and is therefore unable to recover the "true" parameters.

Fig. 4 shows the different RMSD scores between the synthetic observations and prior/posterior simulations for each of the 21 atmospheric stations. The average reduction of RMSD for  $\epsilon$ -4DVar-VarDA methods is 82% with a mean RMSD of 0.1 ppm. The largest reduction of RMSD is for the German station, Shauinsland (SCH) (87%) and the lowest is for the Australian station, Cape Grim (CGO) (49.8%). Comparatively, the average reduction of RMSD for 4DEnVar-EnVarDA methods reaches 97% with a mean RMSD of 0.01 ppm across all stations. The highest reduction of RMSD is for the Chinese station, Waliguan (WLG) (99%) and the lowest one is for the Australian station, Cape Cleveland (CFA) (92.7%). We see that the 4DEnVar-EnVarDA method outperforms the  $\epsilon$ -4DVar-VarDA method: the 4DEnVar-EnVarDA method has the best fit to the "synthetic" observations assimilated and can find the value of the "true" parameters used to generate the "synthetic" observations.

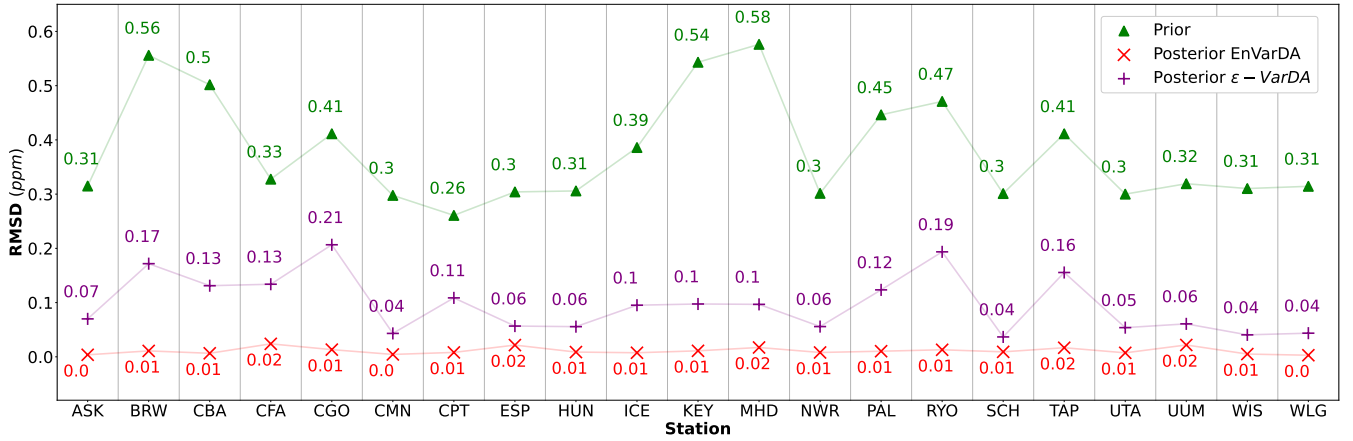
#### 3.2.2 Complex case

For the complex case, Fig. 5 shows the prior/posterior RMSD at each atmospheric station for the 4DEnVar-EnVarDA method using a ensemble of 300 members and the  $\epsilon$ -4DVar-VarDA method using an  $\epsilon$  of  $5 \times 10^{-2}$  for all parameters. We stop the  $\epsilon$ -4DVar-VarDA method after 25 iterations, which already represents 1450 model simulations, as it shows no significant



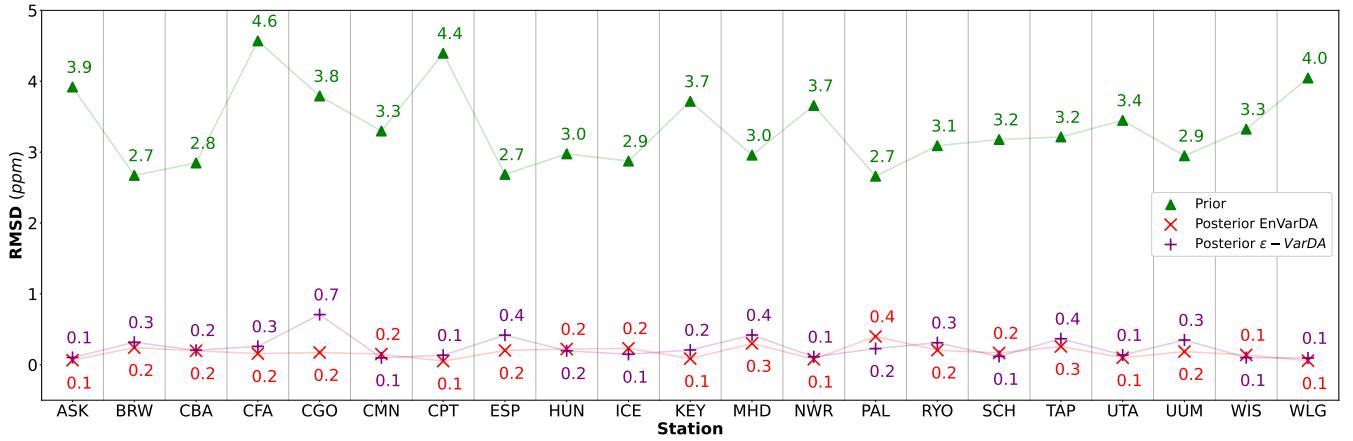
**Figure 3.** Result in parameter space for the simple case: the prior parameter values are represented by the green triangles and the posterior parameter values after optimization are represented by the purple + symbol for the  $\epsilon$ -4DVar-VarDA method and the red × symbol for the 4DEnVar-EnVarDA method. The blue circles represent the “true” values used to produce the assimilated ‘synthetic’ observations. The green error bar represents the prior uncertainty, which is also equal to the standard deviation of the prior ensemble used for 4DEnVar-EnVarDA. The red error bar represents the standard deviation of the posterior ensemble obtained by 4DEnVar-EnVarDA, which can be interpreted as the posterior uncertainty. The Mean Absolute Difference (MAD) score shown is calculated between the “true” parameter values used to generate the “synthetic” observations and the different parameter values following the same color code (green score using prior parameter, purple score using posterior parameter of  $\epsilon$ -4DVar-VarDA, red score posterior parameter of 4DEnVar-EnVarDA).

improvement in the minimisation of its cost function. We find that 4DEnVar-EnVarDA gives a mean reduction in RMSD of 94.3% across all stations with a maximum reduction of RMSD at the South African station, Cape Point (CPT) (98.8% ) and minimum RMSD reduction at the Finland station, Pallas (PAL) (85%). The  $\epsilon$ -4DVar-VarDA gives a mean reduction in RMSD of 92.5% across all stations with a maximum reduction of RMSD at the Chinese station, Walinguan (WLG) (96.9% ) and a minimum RMSD reduction at the Australian station, Cape Grim (CGO) (81.3%). The average RMSD drops from 3.35 ppm to 0.17 ppm after assimilation for 4DEnVar-EnVarDA and to 0.24 ppm for  $\epsilon$ -4DVar-VarDA. Since the posterior RMSDs obtained were close, we performed a paired t-test (Student, 1908) between the two posterior RMSDs to determine whether they were significantly different. We obtained a t-value of -2.125 between the posterior RMSDs obtained by 4DEnVar



**Figure 4.** The Root Mean Squared Differences (RMSD) scores between synthetic observations and prior simulations for the simple case for each of the 21 atmospheric stations are represented by green triangles. The RMSD scores between synthetic observations and posterior simulations given by the  $\epsilon$ -4DVar-VarDA (4DEnVarEnVarDA) method are represented by the purple + symbol (red × symbol)

and  $\epsilon$ -4DVarEnVarDA and  $\epsilon$ -VarDA, with a p-value of 0.046. This confirms that the average posterior RMSD obtained by 4DEnVarEnVarDA is significantly lower than the posterior RMSD obtained by  $\epsilon$ -4DVar-VarDA, with a confidence level of 95%. We computed the mean squared difference (MSD) between the synthetic observations concatenated across all stations and the prior simulation, as well as the two posterior simulations. Following Hodson et al. (2021) and Geman et al. (1992), we decomposed the MSD into bias and variance terms as presented in Section A. The prior MSD is 11.49 ppm<sup>2</sup> and is reduced to 0.04 ppm<sup>2</sup> using the EnVarDA method and to 0.08 ppm<sup>2</sup> using the VarDA method. The decomposition of the prior MSD indicates a squared bias of 4.96 ppm<sup>2</sup> and an error variance equal to 6.53 ppm<sup>2</sup>. The same decomposition for the posterior simulations yields a squared bias of 0.006 ppm<sup>2</sup> and an error variance equal to 0.03 ppm<sup>2</sup> for the EnVarDA method, and a squared bias of 0.002 ppm<sup>2</sup> and an error variance equal to 0.07 ppm<sup>2</sup> for the VarDA method. We calculate the MAD score between the “true” parameter and the prior/posterior parameters after normalising between 0 and 1 (because the parameters do not have the same units). This normalisation allows us to bound the MAD score between 0 and 1. The normalised MAD score between the “true” parameters and the prior parameters is 0.17. After assimilation using 4DEnVarEnVarDA, a 53% reduction in this score is obtained, giving a normalised MAD score of 0.08. The  $\epsilon$ -4DVar-VarDA method gives a reduction in the normalised MAD of 15% giving a normalised MAD score of 0.14. Fig. 6 shows the prior, “true” and posterior parameter values obtained using both methods. For each parameter, we calculate the MAD score independently. The 4DEnVarEnVarDA method gives a MAD reduction of 44.7% for V<sub>cmax</sub>, 78.2% for SLA, 36.3% for LAI<sub>max</sub>, 54.2% for m<sub>maint.resp</sub> and a reduction of the absolute difference (AD) of 98.8% for Q10. The  $\epsilon$ -4DVar-VarDA method gives a MAD reduction of 11.3% for V<sub>cmax</sub>, 32.7% for SLA, 9.6% for LAI<sub>max</sub>, 4.2% for m<sub>maint.resp</sub> and a reduction of the AD of 97.5% for Q10. Fig. 7 illustrates the spatial disparities in net land carbon fluxes between the “synthetic” fluxes and the prior/posterior estimation of the



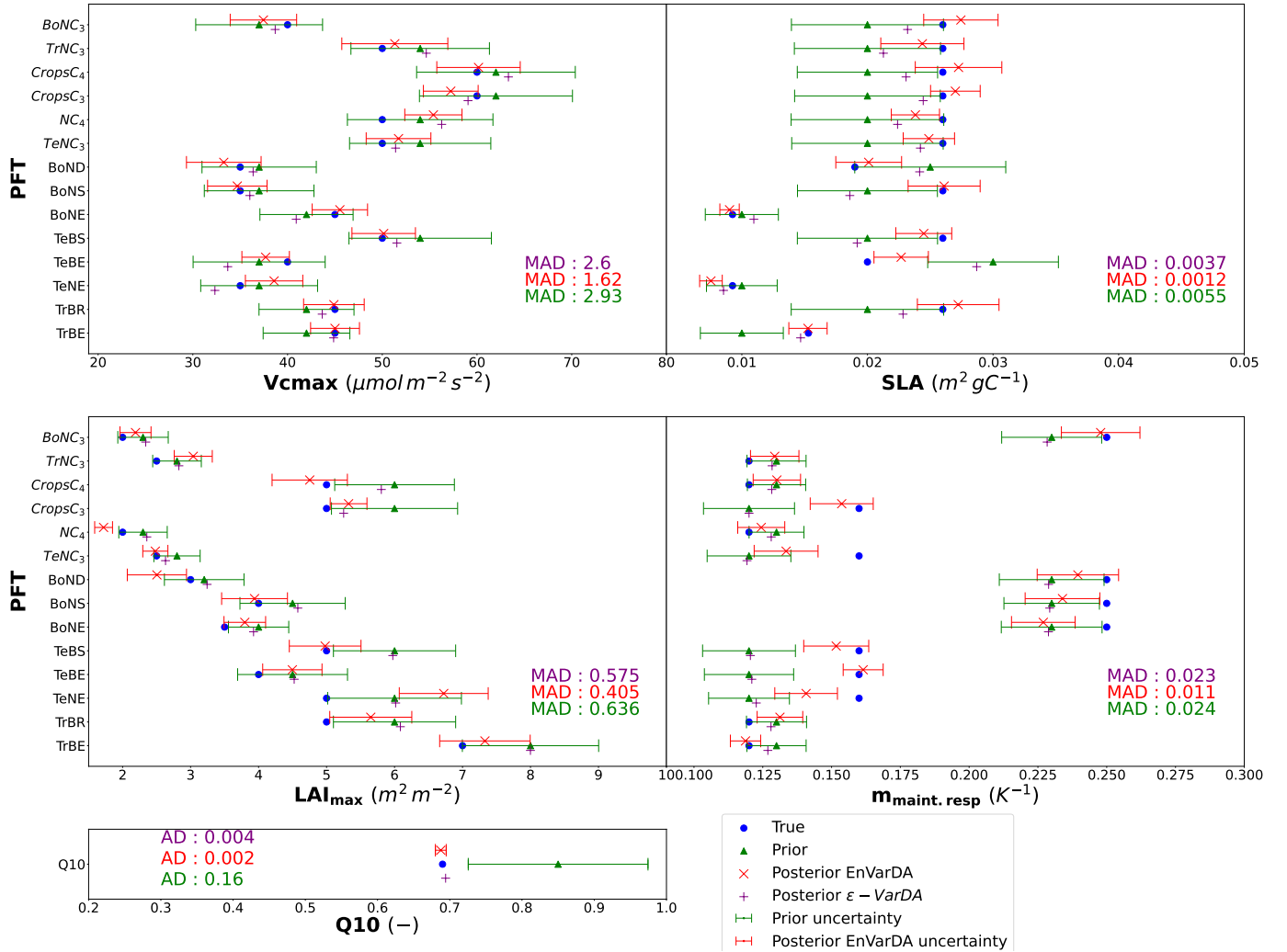
**Figure 5.** The Root Mean Squared Differences (RMSD) scores between synthetic observations and prior simulations for each of the 21 atmospheric stations for the complex case are represented by green triangles. The RMSD scores between synthetic observations and posterior simulations given by the  $\epsilon$ -4DVar-VarDA (4DEnVarEnVarDA) method are represented by the purple + symbol (red × symbol)

two methodologies, in addition to their mean annual global net carbon flux. The 4DEnVarEnVarDA method achieved a mean annual global net flux of  $-2.62 \text{ Gt C year}^{-1}$ , with a difference of  $0.05 \text{ Gt C year}^{-1}$  compared to the "synthetic" fluxes. Spatial differences were limited to an absolute maximum of  $0.28 \text{ g C m}^{-2} \text{ day}^{-1}$ , with an absolute mean of  $0.009 \text{ g C m}^{-2} \text{ day}^{-1}$ . In contrast, the  $\epsilon$ -4DVar-VarDA method produced a mean annual global net flux of  $-2.43 \text{ Gt C year}^{-1}$ , with a difference of  $0.24 \text{ Gt C year}^{-1}$  relative to the "synthetic" fluxes. Spatial differences for this method reached an absolute maximum of  $0.6 \text{ g C m}^{-2} \text{ day}^{-1}$ , with an absolute mean of  $0.031 \text{ g C m}^{-2} \text{ day}^{-1}$ . The Pearson correlation coefficient between the 'synthetic' NBP and the prior NBP is 0.87 in time and 0.17 in space. The posterior NBP obtained by the 4DEnVarEnVarDA method shows a Pearson correlation coefficient against the 'synthetic' NBP of 0.99 in time and 0.98 in space. In comparison, the posterior NBP obtained by the  $\epsilon$ -4DVar-VarDA method has correlation coefficients of 0.98 in time and 0.84 in space.

## 4 Discussion

### 4.1 Experiments

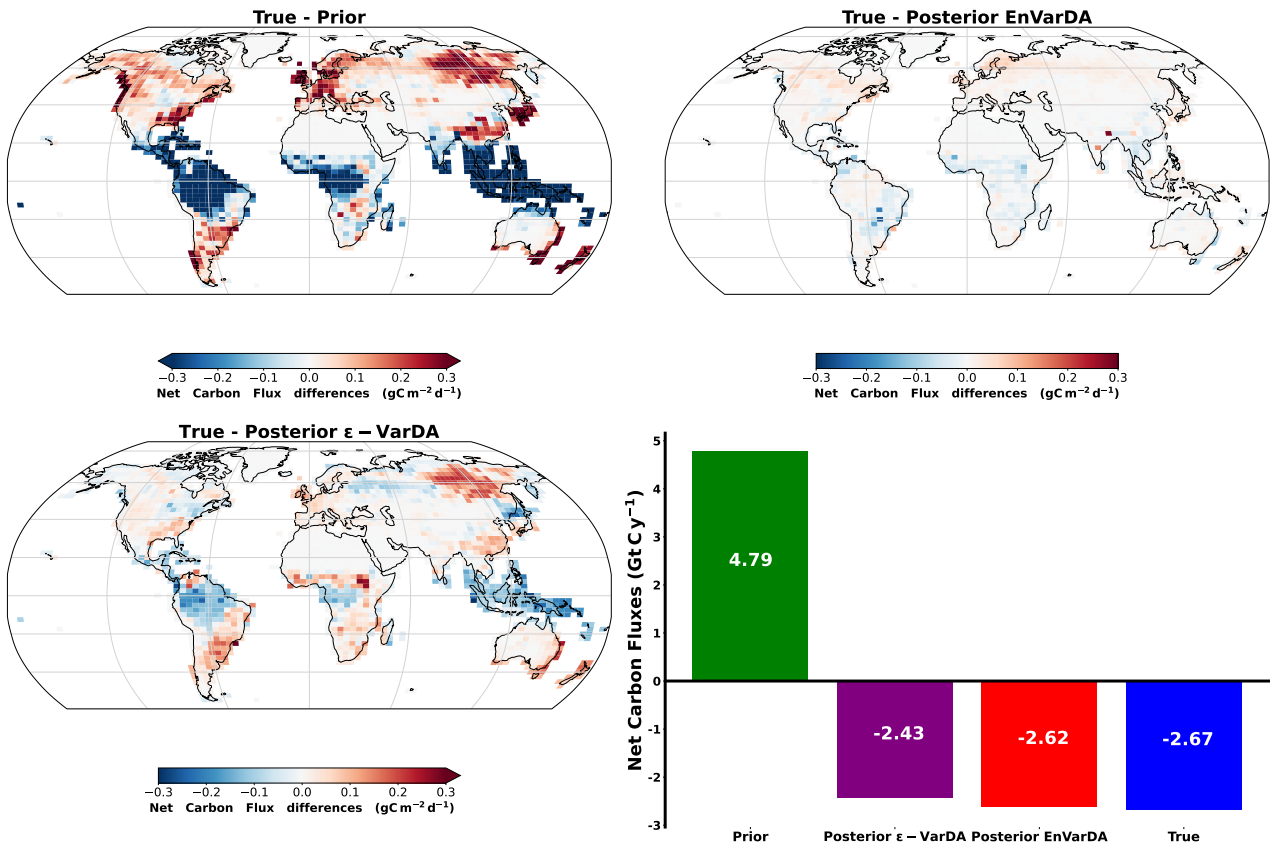
In Section 3.2.1, we found that the 4DEnVarEnVarDA method outperforms the  $\epsilon$ -4DVar-VarDA method, both in terms of RMSD reduction and MAD score, and with a smaller number of model simulations for the simple case. The 4DEnVarEnVarDA method reduces the RMSD by 97%, and almost recovers the "true" parameters, whereas the  $\epsilon$ -4DVar-VarDA method reduces the RMSD by 82% and seems to converge into a local minimum. In addition, 4DEnVarEnVarDA requires three times fewer simulations. Other configurations presented in Section 3.1 show that the 4DEnVarEnVarDA, using 50 members, leads to similar RMSD reduction as  $\epsilon$ -4DVar-VarDA (see Table 1). However, this 4DEnVarEnVarDA configuration still gives a better MAD



**Figure 6.** Results in parameter space for the complex case: the prior parameter values are represented by the green triangles and the posterior parameter values after optimization are represented by the purple + symbol for the  $\epsilon-4DVar-VarDA$  method and the red x symbol for the  $4DEnVar-EnVarDA$  method. The blue circles represent the values used to produce the assimilated 'synthetic' observations. The green error bar represents the prior uncertainty, which is also equal to the standard deviation of the prior ensemble used for  $4DEnVar-EnVarDA$ . The red error bar represents the standard deviation of the posterior ensemble obtained by  $4DEnVar-EnVarDA$ , which can be interpreted as the posterior uncertainty. The MAD (or the absolute differences for Q10) scores shown are calculated for each parameter independently.

score of 0.44 giving a reduction of the prior MAD by 85% - this shows that  $4DEnVar-EnVarDA$  method is less influenced by local minima than the  $\epsilon-4DVar-VarDA$  method. We can also note that using the  $\epsilon-4DVar-VarDA$  method results in a *posterior* parameter values that either i) remain close to the *a priori* values or that ii) increase the distance from the value of the "true" parameters. The first case can be explained by the lower sensitivity of the parameters concerned. The sensitivity of the  $V_{cmax}$

parameter depends on the associated PFT. Not all PFTs have the same impact on NBP fluxes, as they do not have the same spatial distribution or the same proportion (see Table A1). This is the case for the parameters of PFT TeBE, BoNS, BoND, CropsC<sub>4</sub> and TrNC<sub>3</sub> which have a proportion equal to or less than 3%, and are therefore less influential on global NBP fluxes. The second case can be explained by self-compensation due to the equifinality of the problem. Indeed, as some parameters are not properly calibrated, others compensate and may not converge towards the “true” minimum. It may concern the parameter of PFT NC<sub>4</sub>. The  $\epsilon$ -4DVar-EnVarDA method seems to be less affected by these problems and is therefore a promising solution. Furthermore, Fig. 3 shows a significant decrease in the standard deviation of the posterior ensemble. This allows us to identify which parameters and therefore which PFTs appear more sensitive. In this case, it seems that the results for the TrNC<sub>3</sub> and CropsC<sub>4</sub> PFTs are the most uncertain.



**Figure 7.** Spatial differences in net land carbon fluxes between "synthetic" fluxes and the prior/posterior estimate of the two methods alongside their mean annual global net carbon flux for the complex case. (Negative values are carbon uptakes and positive values are carbon emissions)

In Section 3.2.2, we saw that the  $\epsilon$ -4DVar-EnVarDA method is able to calibrate 57 parameters and reduces the mean RMSD by 94.3%, which is slightly better than the  $\epsilon$ -4DVar-VarDA method with a mean RMSD reduction of 92.5%. It is worth noting

that the mean RMSD reduction may be better in the complex case than in the simple case . This is mainly due to the fact that the *a priori* error is much larger in the complex case. Nevertheless the mean RMSD score remains bigger than in the simple case (0.17 ppm and 0.24 ppm for the complex case and 0.01 ppm and 0.1ppm for the simple case for ~~4DEnVar-EnVarDA~~ and  ~~$\epsilon$ -4DVar-VarDA~~ respectively) . Furthermore the MSD score is better for the EnVarDA method (0.04 ppm<sup>2</sup> using the EnVarDA method and to 0.08 ppm<sup>2</sup> using the VarDA method) and the MSD decomposition (Geman et al., 1992; Hodson et al., 2021) highlights that EnVarDA better reduces the error variance, whereas the squared bias reduction is slightly better for the VarDA method. However, squared bias values below 0.01 ppm<sup>2</sup> are negligible. While the mean RSMD reduction and MSD scores are similar for the complex case, the MAD scores in parameter space are different. In fact, the ~~4DEnVar-EnVarDA~~ method is closer to the “true” parameters by reducing the normalised MAD by 53%, whereas the  ~~$\epsilon$ -4DVar-VarDA~~ method remains very close to the *a priori* position. The posterior ensemble generated for the ~~4DEnVar-EnVarDA~~ also shows a reduction in uncertainty for all parameters. This uncertainty reduction is not equal for all parameters - a maximum reduction can be seen for the Q10 parameter(reducing the standard deviation by 94 %) and the lowest for the less sensitive `mmaint.resp` parameter (with a 14% reduction for the NC4 PFT). Both methods are capable of recovering the “true” Q10 parameter since it is the most sensitive parameter. The  ~~$\epsilon$ -4DVar-VarDA~~ method seems to have difficulties in the calibration of the parameter `mmaint.resp` and `LAImax` showing reductions in MAD scores that are less than 10%. Considering Fig. A2, we can see that some PFTs give a partial derivative that do not completely converge with an  $\epsilon$  of  $10^{-2}$  (for example the PFT CropsC<sub>4</sub>), so it is likely that the  $\epsilon$  for these parameters are underestimated. Other PFTs seem to give a partial derivative that do completely converge with an  $\epsilon$  of  $10^{-2}$  (for example the PFT TrBE), but remain close to their *a priori* value, so it is likely that the sensitivity of these parameters is low. The other parameters are therefore self-compensating and this may partly explain the poorer performance of this method in terms of MAD score which are always better for the ~~4DEnVar-EnVarDA~~ method. The self-compensating effect can be illustrated in Fig. 7. The posterior spatial distributions of net carbon flux obtained from the two methods exhibit notable differences. It appears that the  ~~$\epsilon$ -4DVar-VarDA~~ method obtains a different spatial structure of the net carbon fluxes. Indeed, the carbon fluxes absent from one region can be reallocated to another, resulting in only minor variation in atmospheric CO<sub>2</sub> concentration. We believe that the different spatial structure obtained by  ~~$\epsilon$ -4DVar-VarDA~~ against the synthetic net carbon flux is likely to be explained by the fact that the two PFTs TrBE and BoND are not well monitored, creating a dipole in the Amazonian and Siberian regions to compensate for the erroneous carbon flux in other regions. It is therefore notable that the ~~4DEnVar-EnVarDA~~ method demonstrates superior performance, as it is more aligned with the ‘synthetic’ net carbon fluxes both spatially and globally than the  ~~$\epsilon$ -4DVar-VarDA~~ method. Fig. A3 shows the differences in spatial distribution of gross primary production (GPP) between the "synthetic" fluxes and the prior/posterior estimate of the two methods, as well as their global yearly budget. We can see that GPP obtained with the ~~4DEnVar-EnVarDA~~ method is slightly better than the  ~~$\epsilon$ -4DVar-VarDA~~ method for the global budget and better matches the spatial distribution of the synthetic flux. The  ~~$\epsilon$ -4DVar-VarDA~~ method appears to compensate for the lack of change between the prior and posterior GPP across most of the Northern Hemisphere. However, the ~~4DEnVar-EnVarDA~~ method outperforms the  ~~$\epsilon$ -4DVar-VarDA~~ method also in terms of computational cost: the ~~4DEnVar-EnVarDA~~ method only needs 300 simulations, whereas the  ~~$\epsilon$ -4DVar-VarDA~~ method needs 1450, which means a reduction in computing time of almost five times. This experiment of calibrating

a large number of parameters represents a more realistic case, even if we consider a very low model/observation error. The results demonstrate the good performance of 4DEnVar-EnVarDA, which, even in a ‘perfect’ model situation, i.e. a model that can perfectly simulate observations, can assimilate observations while being less impacted by local minima. However, this may not be the case when using actual observations and introducing more complex modelling/observation errors. The use of 4DEnVar-EnVarDA here therefore demonstrates its ability to calibrate many parameters with fewer model simulations. In this experiment, one simulation took 11 minutes (wall times) on average, using 20 CPUs of a computer server (using Intel Xeon Gold 5115 processor). Neglecting the other computational times, using the  $\epsilon$ -4DVar-VarDA for the complex case represents more than 265 hours of computation, where it only represents 55 hours for the 4DEnVar-EnVarDA. Such a reduction cannot be ignored since a simulation in this experiment represents a short (only 2 years), low-cost model configuration - low ORCHIDEE spatial resolution and use of pre-calculated LMDZ transport fields.

In this twin experiment, both methods have to deal with the inherent equifinality of atmospheric concentration assimilation. This equifinality occurs when parameters compensate for each other, resulting in either an incorrect spatial distribution of NBP or inaccurate estimates of subcomponents such as GPP and total ecosystem respiration (TER), but still allowing for a match with observations. Although both methods considered in this study successfully recovered the global budgets for NBP and GPP, the  $\epsilon$ -4DVar-VarDA method did not obtain the correct spatial distributions of NBP and GPP (see Fig. 7 and Fig. A3). This is not the case for the 4DEnVar-EnVarDA method, which better recovered the ‘true’ spatial distributions of NBP and GPP. We believe that this equifinality could increase the number of local minima, further disrupting the performance of the  $\epsilon$ -4DVar-VarDA method. We also believe that the ensemble nature of the 4DEnVar-EnVarDA method provides a more comprehensive view of the parameter space, making it less sensitive to local minima and therefore to equifinality issues.

The poorer performance of the  $\epsilon$ -4DVar-VarDA method is likely related to inaccurate determination of  $\epsilon$ , which results in inaccurate estimates of the tangent linear and adjoint models. The 4DEnVar-EnVarDA method avoids the development and maintenance of tangent linear and adjoint models, and ensures a fully functional assimilation method that does not require the use of finite differences. But the performance of the 4DEnVar-EnVarDA method seems dependent on the generated ensemble. As shown in Tab. 2, slightly lower performance is observed with larger ensembles, indicating that a bigger ensemble does not necessarily yield better results. This could be due to the increased dimensionality of the problem, making the iterative minimization more challenging. Additionally, we generated a new ensemble for each experiment, which provides different information about the parameter space and can lead to different optimal values. This shows the importance of the prior ensemble generated. Nevertheless, the reduction in RMSD remains satisfactory, with a reduction of more than 90%. It seems that the subjective choice of the 4DEnVar-EnVarDA set-up, i.e. the size and distribution of the ensemble, is less critical than the subjective choice of  $\epsilon$  used in the  $\epsilon$ -4DVar-VarDA, which must be determined independently for each of the parameters and given assimilated data-streams (with the associated number of model simulations). Moreover, like the tangent linear and adjoint models, this  $\epsilon$  must be re-tuned for a different model as the sensitivity of the parameter can be different. Indeed, other studies using different versions of the ORCHIDEE LSM used different  $\epsilon$  values (Santaren et al., 2007; Kuppel et al., 2012; Peylin et al., 2016; Bacour et al., 2023).

600 The results obtained here for the  $\epsilon$ -4DVar-VarDA are not equivalent to a standard 4DVar-VarDA using a tangent linear and adjoint model. Therefore, we can draw no conclusions on the comparison between the 4DVar-VarDA and a standard 4DVar-VarDA methods as was highlighted in Liu et al. (2008). A potential - but hard to implement - way to improve the  $\epsilon$ -4DVar-VarDA may be to have a dynamic  $\epsilon$  that becomes more refined as the methods converge. Nevertheless, even with the "perfect"  $\epsilon$ , we cannot guarantee that the  $\epsilon$ -4DVar-VarDA method would be less computationally expensive. The assimilation of atmospheric CO<sub>2</sub> concentration data using 4DVar-VarDA has been implemented with a tangent linear model, as in Castro-Morales et al. (2019), or an adjoint model, as in Scholze et al. (2007). In these cases, the tangent linear or adjoint model was developed alongside the forward model. However, the  $\epsilon$ -4DVar-VarDA method was used in experiments where obtaining the tangent linear or adjoint model proved too difficult, such as in Peylin et al. (2016); Bacour et al. (2023). Although  $\epsilon$ -4DVar-VarDA is not equivalent to standard 4DVar-VarDA, a comparison of 4DVar-EnVarDA with  $\epsilon$ -4DVar-VarDA demonstrates 610 the strong performance of 4DVar-EnVarDA, making it a promising candidate for this application.

## 4.2 Challenges and perspectives

This study relies on twin experiments, which eliminate the complexities associated with model/observation errors, and allows us to focus on the performance of two assimilation methods. This experiment highlights the superiority of the 4DVar-EnVarDA method to assimilate atmospheric CO<sub>2</sub> concentration data. However, the assimilation of real observations is not 615 straightforward. The use of real data must be followed by characterisation of the model/observation errors. Indeed, the matrix  $\mathbf{R}$  must reflect modelling/observation errors at each site, which would introduce spatial heterogeneity, as each station may have different modelling errors, mainly structural errors from both the transport model and fluxes given by the ORCHIDEE LSM, or measurement problems. A good characterisation of the matrix  $\mathbf{R}$  is of paramount importance, as it can have a considerable impact on the results obtained. If the model/observation errors are incorrect, the 4DVar-EnVarDA method can give infeasible 620 ~~a-posterior~~ posterior parameter values, i.e. outside the imposed parameter boundaries (and therefore give non-physical parameter values). Furthermore, even with feasible ~~a-posterior~~ posterior parameter values, the parameters obtained may be beyond the assumption of linearity made by the use of linear combinations in Eq. 18 and therefore do not improve the associated simulation. Nevertheless, several techniques seem promising for managing these limitations. The inclusion of a weight factor in the background term, as is done in (Raoult et al., 2016), and a better definition of the error covariance matrix  $\mathbf{B}$  may provide 625 a solution. Some of these challenges are not specific to the 4DVar-EnVarDA method and are common to the 4DVar-VarDA method (Raoult et al., 2024b). These challenges are therefore the subject of active research to improve the assimilation of real observational data.

The assimilation of real observations of atmospheric concentrations may also increase the equifinality mentioned in Section 4.1 for several reasons, such as: i) Incorrect initial conditions of the carbon pools, which can impact respiration. ii) Wrong 630 estimates of other flux components, such as ocean or fossil fuel components. iii) Structural errors in either the land surface model or the transport model. The issue of incorrect initial conditions can be addressed by starting the simulation a couple of years before the assimilation window. This allows for the correction of the initial carbon pool and better accounts for the effects of the new parameter values on the carbon pool. To handle other components, such as ocean components, the same

assimilation can be repeated using different estimates of the ocean flux. Ideally, an ocean model could be included in the optimization to calibrate both land and ocean components, as is done in atmospheric inversion. The advantage of the ~~4DEnVar~~ EnVarDA method is that it only requires forward simulations. Therefore, no code adaptations are needed, making it easier to use different transport models. This should help detect and address structural errors. The equifinality can also be reduced by assimilating multiple data streams simultaneously, as done in Peylin et al. (2016) and Bacour et al. (2023), to calibrate both GPP and NBP at the same time.

This study acts as a proof-of-concept for the assimilation of atmospheric CO<sub>2</sub> concentration data using adjoint-free methods. The next steps for the future would be to use real observations, which come with other technical and scientific problems (e.g. quantifying the model/observation error). Future studies should focus on the assimilation of more recent and more spatially distributed atmospheric CO<sub>2</sub> concentration data - e.g. satellite XCO<sub>2</sub> product, using ~~4DEnVar~~ EnVarDA. To do so, a more recent version of LMDZ and/or ORCHIDEE should be used. Those studies will focus on the processes involved in the carbon cycle to improve their parameterisation and/or to detect any missing processes in the model. As the ~~4DEnVar~~ EnVarDA method only requires forward simulations of the models, it is easy to change the model (either the LSM or the atmospheric transport model). Furthermore, the method is easy to parallelise as each element of the ensemble is independent. Once built, no further call of the model is necessary (except in the analysis step), which allows us to explore different configurations, e.g. in the construction of the error matrix **R** or the weighting of background terms - both of which play a key role in the assimilation of real observations - without additional computational cost.

Despite extensive research on the automatic generation of tangent linear and adjoint models - either using new languages or differential software - it remains an enormous challenge to acquire and maintain tangent linear and adjoint models for complex and continuously evolving models. However, it is still a key priority to understand structural errors, to quantify uncertainties and to refine future predictions via parameter calibration. The use of adjoint-free data assimilation methods such as ~~4DEnVar~~ EnVarDA is therefore an excellent opportunity, as it can be implemented quickly and requires no model modification.

Moreover, the ~~4DEnVar~~ EnVarDA method was used to assimilate several types of data using either simple carbon model (Douglas et al., 2025) or more complex LSM as the JULES LSM (Pinnington et al., 2020, 2021; Cooper et al., 2021). This new application in the ORCHIDEE LSM shows that this method is model-independent. By adding different observation terms (one term per data flux) to the cost function, the method should be able to perform multi-flux data assimilation, which would help to reduce the equifinality problem.

## 5 Conclusions

We showed that the ~~4DEnVar~~ EnVarDA method has good potential for calibrating ORCHIDEE parameters assimilating atmospheric CO<sub>2</sub> concentration data and using the LMDZ atmospheric transport model. The method was tested on a so-called *twin experiment* using two different cases: 1) a simple case where ~~4DEnVar~~ EnVarDA effectively recovered the “true” parameter values, whereas the ~~ε-4DVar~~ VarDA method, despite reducing the RMSD, failed to recover the “true” parameters; and 2) a complex case where both methods achieved up to a 90% reduction in RMSD, with ~~4DEnVar~~ EnVarDA showing slightly

better performance, including a lower MAD score in parameter space, indicating greater efficiency in parameter recovery and an improved alignment with 'synthetic' net carbon fluxes, both spatially and globally. Additionally, ~~4DEnVar~~ EnVarDA is computationally less demanding and does not require the development or maintenance of tangent linear and adjoint models, 670 facilitating the use of updated model versions without modification. By successfully applying this method to the ORCHIDEE model with a pre-calculated LMDZ transport model, we illustrated its adaptability, making it well-suited for other land surface models, whether coupled with atmospheric transport models or not.

*Code and data availability.* The source code for the ORCHIDEE version used in this model is freely available online at <https://doi.org/10.14768/c68bc728-da71-4383-84df-dcde31d9c006> ORCHIDEE (2025). The ORCHIDAS EnVarDA code and data used in 675 this paper are available from a Zenodo repository at <https://doi.org/10.5281/zenodo.14609416> Beylat (2025).

## Appendix A: Metrics calculation

The RMSD and MAD used in these studies are calculated as follows:

$$RMSD = \sqrt{\frac{1}{N_t} \sum_{t=0}^{N_t} (\mathcal{H}(x_*)_t - y_t)^2}, \quad (A1)$$

$$680 \quad MAD = \frac{1}{n_{param}} \sum_{i=0}^{n_{param}} |x_{*i} - x_{truei}|, \quad (A2)$$

where  $x_*$  can be either  $x_b$  or  $x_a$ . The Pearson correlation coefficients were computed using the Numpy Python library with the `corrcoef` function. The paired t-tests were computed using the `stats.ttest_rel` function from the Scipy library. We use the decomposition of MSD into bias and variance that was proposed by Geman et al. (1992) and presented by Hodson et al. (2021):

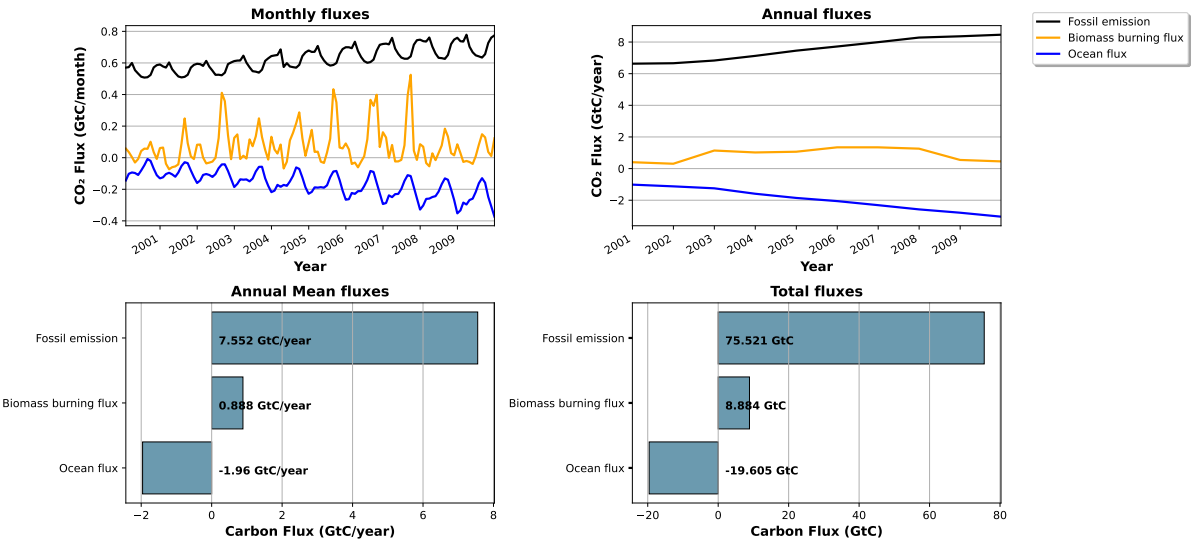
$$685 \quad \underline{e = \mathcal{H}(x) - y} \quad (A3)$$

$$\underline{MSD(e) = \mathbb{E}[e]} \quad (A4)$$

$$\underline{MSD(e) = (\mathbb{E}[e^2] - \mathbb{E}[e]^2) + \mathbb{E}[e]^2} \quad (A5)$$

$$690 \quad \underline{MSD(e) = Var(e) + \mathbb{E}[e]^2} \quad (A6)$$

$$MSD(e) = Var(e) + Bias(e)^2$$
(A7)

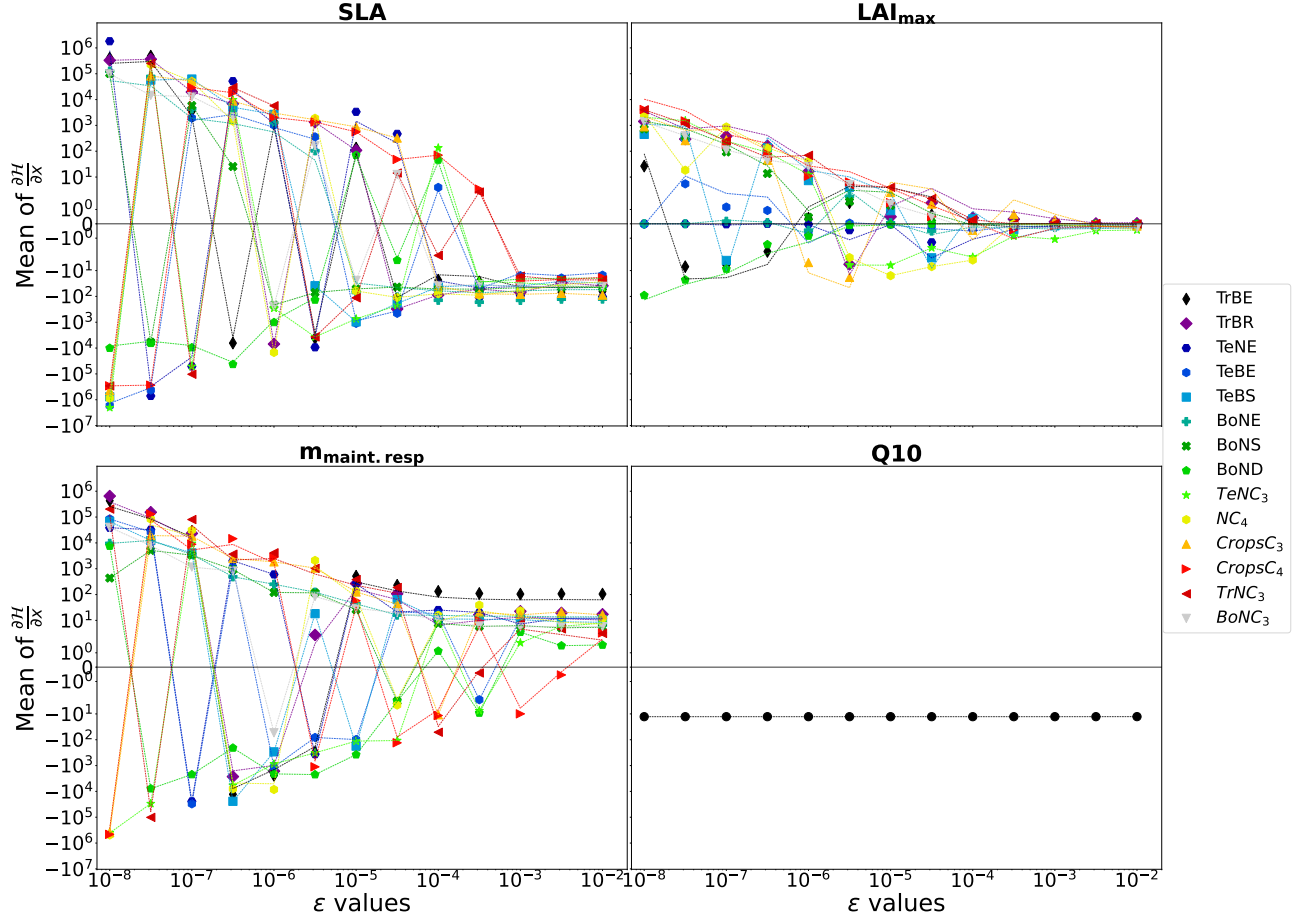


**Figure A1.** Ocean, Biomass burning carbon fluxes and Fossil emission (2000-2009).

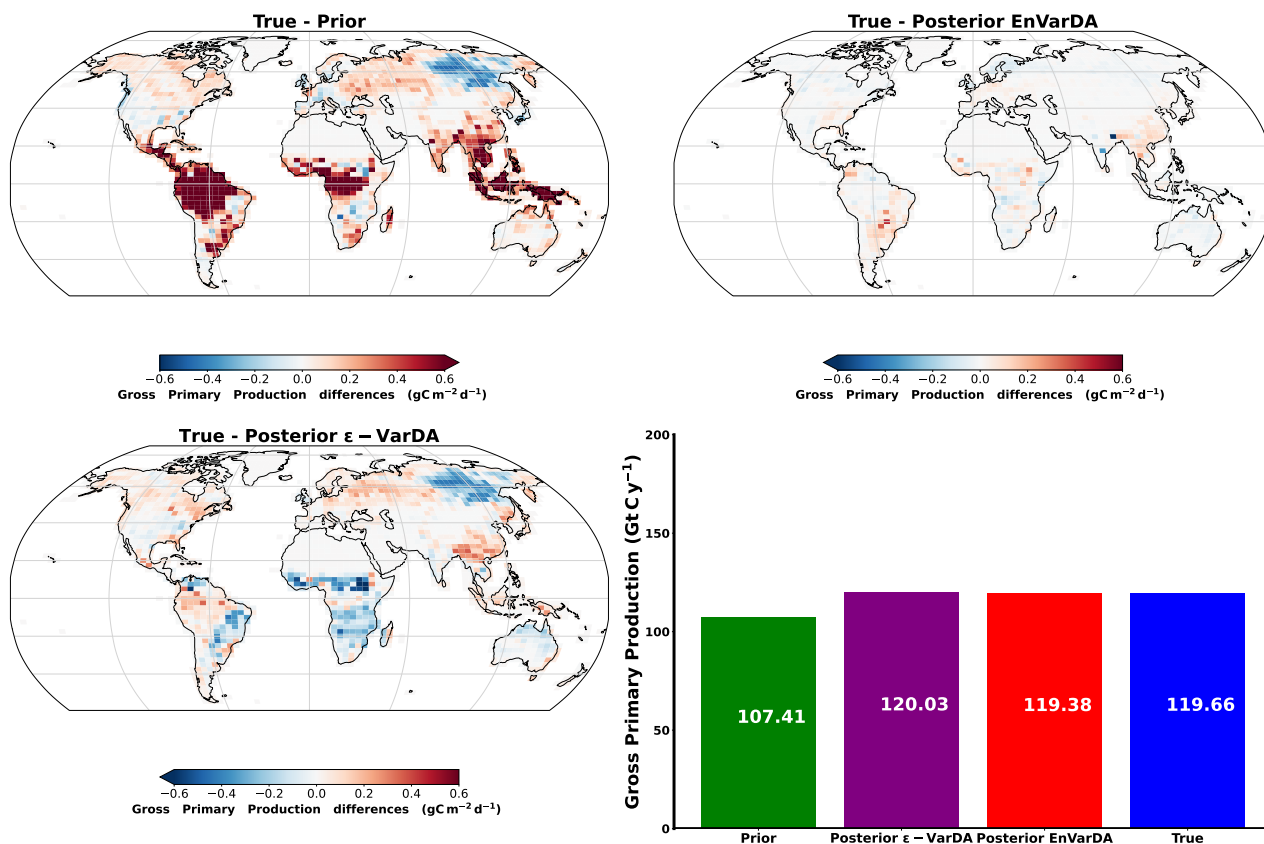
*Author contributions.* SB and PP conceived on the study. VB implemented the coupling of ORCHIDEE and LMDZ, originally developed by PP. SB implemented the EnVarDA method and VB implemented the  $\epsilon$ -VarDA method. SB performed and analysed the DA experiments. NR, PR, PP, and CB provided expertise on the assimilation of Atmospheric CO<sub>2</sub> concentration and the  $\epsilon$ -VarDA method. ND and TQ provided expertise on the EnVarDA method. SB prepared the article with contributions from all co-authors.

*Competing interests.* The authors declare that they have no competing interests.

*Acknowledgements.* This work has been supported by a scholarship from CNRS under the Melbourne–CNRS joint doctoral programme.



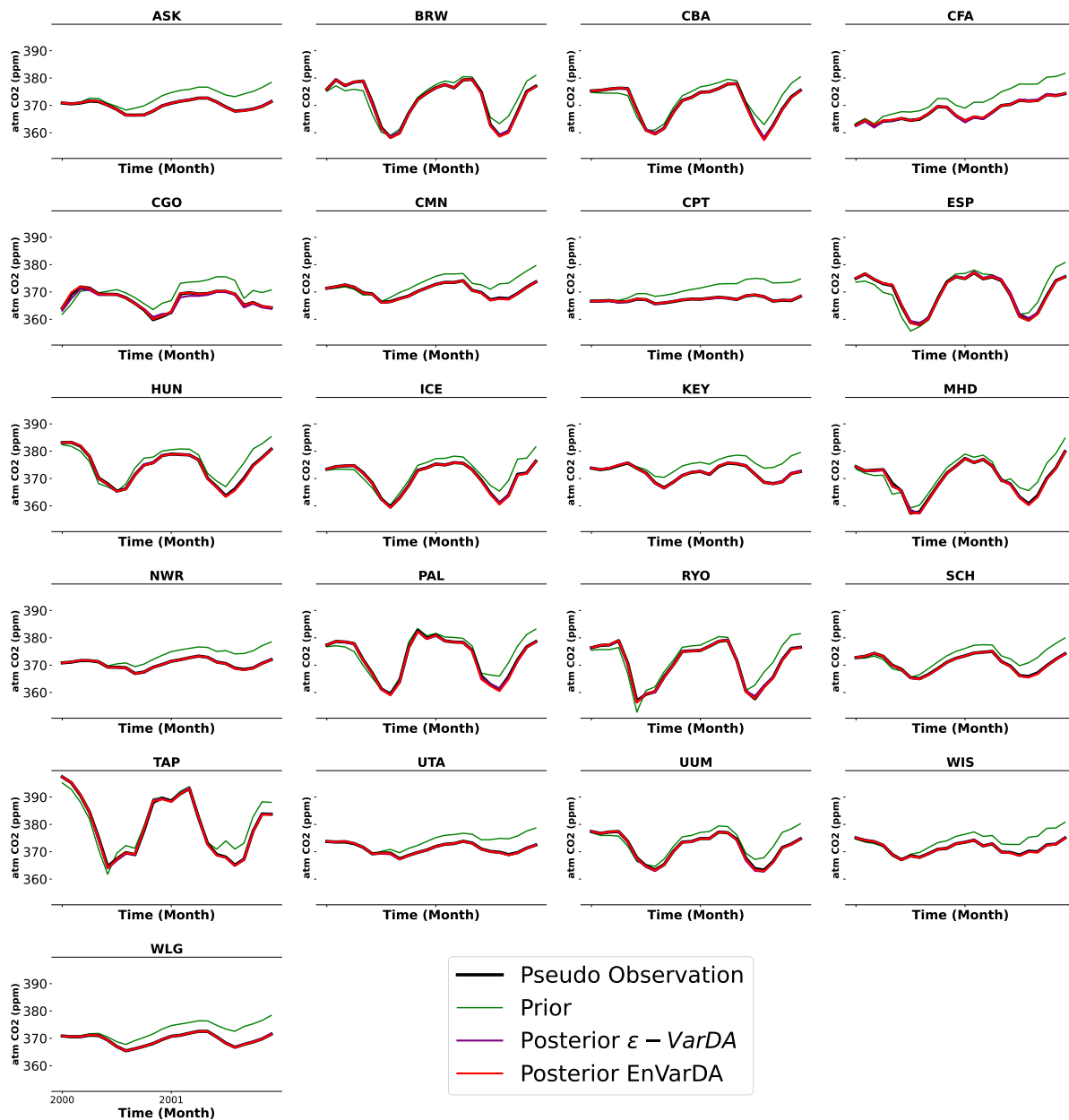
**Figure A2.**  $\epsilon$ -test: spatial and temporal average of partial derivative as a function of  $\epsilon$ . The partial derivative of the  $\mathcal{H}$  model is calculated with respect to the parameters use in complex case for each PFT. It is calculated on the concentration space using every station over 2 years. The mean of the partial derivative is then calculated over space and time in order to visualise the local derivative. The derivative of  $\mathcal{H}$  is calculated for several  $\epsilon$ .



**Figure A3.** Spatial differences in gross primary production (GPP) fluxes between "synthetic" fluxes and the prior/posterior estimate of the two methods alongside their mean annual global GPP for the complex case.

## 700 References

- Asch, M., Bocquet, M., and Nodet, M.: Data Assimilation: Methods, Algorithms, and Applications, vol. 28, 2016.
- Bacour, C., Maignan, F., MacBean, N., Porcar-Castell, A., Flexas, J., Frankenberg, C., Peylin, P., Chevallier, F., Vuichard, N., and Bastrikov, V.: Improving Estimates of Gross Primary Productivity by Assimilating Solar-Induced Fluorescence Satellite Retrievals in a Terrestrial Biosphere Model Using a Process-Based SIF Model, *Journal of Geophysical Research: Biogeosciences*, 124, <https://doi.org/10.1029/2019JG005040>, 2019.
- Bacour, C., Macbean, N., Chevallier, F., Léonard, S., Koffi, E. N., and Peylin, P.: Assimilation of multiple datasets results in large differences in regional-to global-scale NEE and GPP budgets simulated by a terrestrial biosphere model, *Biogeosciences*, 20, <https://doi.org/10.5194/bg-20-1089-2023>, 2023.
- Baker, D. F., Doney, S. C., and Schimel, D. S.: Variational data assimilation for atmospheric CO<sub>2</sub>, in: *Tellus, Series B: Chemical and Physical Meteorology*, vol. 58, <https://doi.org/10.1111/j.1600-0889.2006.00218.x>, 2006.



**Figure A4.** Time series of pseudo-observations (in black) of atmospheric CO<sub>2</sub> concentrations at 21 stations over the assimilated period, as well as the a priori (in green) and a posteriori time series from  $\epsilon - \text{VarDA}$  (in purple) and  $\text{EnVarDA}$  (in red) for the complex case.

Baker, E., Harper, A. B., Williamson, D., and Challenor, P.: Emulation of high-resolution land surface models using sparse Gaussian processes with application to JULES, *Geoscientific Model Development*, 15, <https://doi.org/10.5194/gmd-15-1913-2022>, 2022.

**Table A1.** Plant functional types (PFT) in ORCHIDEE and acronyms used in this study and respective Global Cover Fraction (with the remaining portion being bar soil) .

| PFT                              | Acronym             | Proportion |
|----------------------------------|---------------------|------------|
| Tropical Broadleaf Evergreen     | TrBE                | 10.3%      |
| Tropical Broadleaf Raingreen     | TrBR                | 5.92%      |
| Temperate Needleleaf Evergreen   | TeNE                | 4.51%      |
| Temperate Broadleaf Evergreen    | TeBE                | 2.31%      |
| Temperate Broadleaf Summergreen  | TeBS                | 5.09%      |
| Boreal Needleleaf Evergreen      | BoNE                | 4.72%      |
| Boreal Broadleaf Summergreen     | BoNS                | 2.29%      |
| Boreal Needleleaf Deciduous      | BoND                | 2.59%      |
| Temperate Natural Grassland (C3) | TeNC <sub>3</sub>   | 6.43%      |
| Natural Grassland (C4)           | NC <sub>4</sub>     | 8.73%      |
| Crops (C3)                       | CropsC <sub>3</sub> | 11.4%      |
| Crops (C4)                       | CropsC <sub>4</sub> | 3.04%      |
| Tropical Natural Grassland (C3)  | TrNC <sub>3</sub>   | 1.58%      |
| Boreal Natural Grassland (C3)    | BoNC <sub>3</sub>   | 10.6%      |

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, <https://doi.org/10.1002/qj.2982>, 2017.

715 Bastrikov, V., Macbean, N., Bacour, C., Santaren, D., Kuppel, S., and Peylin, P.: Land surface model parameter optimisation using in situ flux data: Comparison of gradient-based versus random search algorithms (a case study using ORCHIDEE v1.9.5.2), *Geoscientific Model Development*, 11, <https://doi.org/10.5194/gmd-11-4739-2018>, 2018.

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, <https://doi.org/10.5194/acp-13-8695-2013>, 2013.

720 Berchet, A., Sollum, E., Thompson, R. L., Pison, I., Thanwerdas, J., Broquet, G., Chevallier, F., Aalto, T., Berchet, A., Bergamaschi, P., Brunner, D., Engelen, R., Fortems-Cheiney, A., Gerbig, C., Zwaafink, C. D. G., Haussaire, J. M., Henne, S., Houweling, S., Karstens, U., Kutsch, W. L., Luijkx, I. T., Monteil, G., Palmer, P. I., Peet, J. C. V., Peters, W., Peylin, P., Potier, E., Rödenbeck, C., Saunio, M., Scholze, M., Tsuruta, A., and Zhao, Y.: The Community Inversion Framework v1.0: A unified system for atmospheric inversion studies, *Geoscientific Model Development*, 14, <https://doi.org/10.5194/gmd-14-5331-2021>, 2021.

725 Beylat, S.: 4DEnVar ORCHIDEE: v1.0.0, <https://doi.org/10.5281/zenodo.14609416>, 2025.

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D’Andrea, F., Davini, P., de Lavergne, C., Denvil, S., Deshayes, J., Devilliers, M., Ducharne, A., Dufresne, J. L., Dupont, E., Éthé, C., Fairhead, L., Falletti, L., Flavoni, S., Foujols, M. A., Gardoll, S., Gastineau, G., Ghattas, J., Grandpeix, J. Y., Guenet, B., Lionel, E. G., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levvasseur, G., Lévy, C., Li, L., Lott,

730

**Table A2.** Prior value of each parameter and their 'true' values (default values in the ORCHIDEE LSM model)

|                     | Parameters |       |        | $V_{cmax}$ |       |        | $SLA$   |       |        | $LAI_{max}$ |       |        | $m_{maint.resp}$ |       |        | Q10  |        |       |        |      |
|---------------------|------------|-------|--------|------------|-------|--------|---------|-------|--------|-------------|-------|--------|------------------|-------|--------|------|--------|-------|--------|------|
|                     |            | Prior | Errors | True       | Prior | Errors | True    | Prior | Errors | True        | Prior | Errors | True             | Prior | Errors | True | Global | Prior | Errors | True |
| PFT                 |            |       |        |            |       |        |         |       |        |             |       |        |                  |       |        |      |        | 0.85  | 0.16   | 0.69 |
| TrBE                |            | 42    | 6      | 45         | 0.01  | 0.0046 | 0.0153  | 8     | 1.2    | 7           | 0.13  | 0.012  | 0.12             |       |        |      |        |       |        |      |
| TrBR                |            | 42    | 6      | 45         | 0.02  | 0.0074 | 0.026   | 6     | 1      | 5           | 0.13  | 0.012  | 0.12             |       |        |      |        |       |        |      |
| TeNE                |            | 37    | 6.4    | 35         | 0.01  | 0.0032 | 0.0926  | 6     | 1      | 5           | 0.12  | 0.02   | 0.16             |       |        |      |        |       |        |      |
| TeBE                |            | 37    | 8      | 40         | 0.03  | 0.006  | 0.02    | 4.5   | 0.8    | 4           | 0.12  | 0.02   | 0.16             |       |        |      |        |       |        |      |
| TeBS                |            | 54    | 8      | 50         | 0.02  | 0.0074 | 0.026   | 6     | 1      | 5           | 0.12  | 0.02   | 0.16             |       |        |      |        |       |        |      |
| BoNE                |            | 42    | 6      | 45         | 0.01  | 0.0032 | 0.00926 | 4     | 0.6    | 3.5         | 0.23  | 0.02   | 0.25             |       |        |      |        |       |        |      |
| BoNS                |            | 37    | 6.4    | 35         | 0.02  | 0.0074 | 0.026   | 4.5   | 0.8    | 4           | 0.23  | 0.02   | 0.25             |       |        |      |        |       |        |      |
| BoND                |            | 37    | 6.4    | 35         | 0.025 | 0.0062 | 0.019   | 3.2   | 0.6    | 3           | 0.23  | 0.02   | 0.25             |       |        |      |        |       |        |      |
| TeNC <sub>3</sub>   |            | 54    | 8      | 50         | 0.02  | 0.0074 | 0.026   | 2.8   | 0.4    | 2.5         | 0.12  | 0.02   | 0.16             |       |        |      |        |       |        |      |
| NC <sub>4</sub>     |            | 54    | 8      | 50         | 0.02  | 0.0074 | 0.026   | 2.3   | 0.4    | 2           | 0.13  | 0.012  | 0.12             |       |        |      |        |       |        |      |
| CropsC <sub>3</sub> |            | 62    | 8.4    | 60         | 0.02  | 0.0074 | 0.026   | 6     | 1      | 5           | 0.12  | 0.02   | 0.16             |       |        |      |        |       |        |      |
| CropsC <sub>4</sub> |            | 62    | 8.4    | 60         | 0.02  | 0.0074 | 0.026   | 6     | 1      | 5           | 0.13  | 0.012  | 0.12             |       |        |      |        |       |        |      |
| TrNC <sub>3</sub>   |            | 54    | 8      | 50         | 0.02  | 0.0074 | 0.026   | 2.8   | 0.4    | 2.5         | 0.13  | 0.012  | 0.12             |       |        |      |        |       |        |      |
| BoNC <sub>3</sub>   |            | 37    | 8      | 40         | 0.02  | 0.0074 | 0.026   | 2.3   | 0.4    | 2           | 0.23  | 0.02   | 0.25             |       |        |      |        |       |        |      |

**Table A3.** Spatial and temporal average of the partial derivative for all parameter for each PFT computed using  $\epsilon$  allowing a stable derivation. The partial derivative is calculated on the concentration space using every station over 2 years. The mean of the partial derivative is then calculated over space and time

|                     | Parameters | $V_{cmax}$ | $SLA$     | $LAI_{max}$    | $m_{maint.resp}$ | Q10                 |
|---------------------|------------|------------|-----------|----------------|------------------|---------------------|
|                     | $\epsilon$ | $10^{-3}$  | $10^{-3}$ | $10^{-2}$      | $10^{-2}$        | $10^{-3}$           |
| <b>PFT</b>          |            |            |           |                |                  | <b>Global</b> -12.7 |
| TrBE                |            | -0.153     | -63.7     | -0.07          | 104.7            |                     |
| TrBR                |            | -0.042     | -38.7     | 0.03           | 17.2             |                     |
| TeNE                |            | -0.077     | -103.3    | -0.02          | 12.1             |                     |
| TeBE                |            | -0.06      | -15.1     | -0.13          | 11.4             |                     |
| TeBS                |            | -0.04      | -30.6     | -0.06          | 8.5              |                     |
| BoNE                |            | -0.057     | -126.2    | -0.19          | 13.8             |                     |
| BoNS                |            | -0.065     | -44.6     | -0.07          | 5.4              |                     |
| BoND                |            | -0.023     | -23.6     | $-5 * 10^{-5}$ | 1.5              |                     |
| TeNC <sub>3</sub>   |            | -0.053     | -44.9     | -0.43          | 7.9              |                     |
| NC <sub>4</sub>     |            | -0.072     | -27.0     | -0.22          | 12.0             |                     |
| CropsC <sub>3</sub> |            | -0.072     | -89.3     | -0.15          | 16.4             |                     |
| CropsC <sub>4</sub> |            | -0.023     | -18.3     | -0.07          | 3.3              |                     |
| TrNC <sub>3</sub>   |            | -0.017     | -23.0     | -0.01          | 3.1              |                     |
| BoNC <sub>3</sub>   |            | -0.056     | -40.7     | -0.28          | 5.7              |                     |

F., Lurton, T., Luyssaert, S., Madec, G., Madeleine, J. B., Maignan, F., Marchand, M., Marti, O., Mellul, L., Meurdesoif, Y., Mignot, J., Musat, I., Ottlé, C., Peylin, P., Planton, Y., Polcher, J., Rio, C., Rochetin, N., Rousset, C., Sepulchre, P., Sima, A., Swingedouw, D., Thiéblemont, R., Traore, A. K., Vancoppenolle, M., Vial, J., Vialard, J., Viovy, N., and Vuichard, N.: Presentation and Evaluation of the

735 IPSL-CM6A-LR Climate Model, Journal of Advances in Modeling Earth Systems, 12, <https://doi.org/10.1029/2019MS002010>, 2020.

Bousquet, P., Peylin, P., Ciais, P., Quere, C. L., Friedlingstein, P., and Tans, P. P.: Regional changes in carbon dioxide fluxes of land and oceans since 1980, Science, 290, <https://doi.org/10.1126/science.290.5495.1342>, 2000.

Byrd, R. H., Lu, P., Nocedal, J., and Zhu, C.: A Limited Memory Algorithm for Bound Constrained Optimization, SIAM Journal on Scientific Computing, 16, <https://doi.org/10.1137/0916069>, 1995.

740 Castro-Morales, K., Schürmann, G., Köstler, C., Rödenbeck, C., Heimann, M., and Zaehle, S.: Three decades of simulated global terrestrial carbon fluxes from a data assimilation system confronted with different periods of observations, Biogeosciences, 16, <https://doi.org/10.5194/bg-16-3009-2019>, 2019.

Chevallier, F., Fisher, M., Peylin, P., Serrar, S., Bousquet, P., Bréon, F. M., Chédin, A., and Ciais, P.: Inferring CO2 sources and sinks from satellite observations: Method and application to TOVS data, Journal of Geophysical Research Atmospheres, 110, <https://doi.org/10.1029/2005JD006390>, 2005.

745

- Chevallier, F., Bréon, F. M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO<sub>2</sub> sources and sinks: Theoretical study in a variational data assimilation framework, *Journal of Geophysical Research Atmospheres*, 112, <https://doi.org/10.1029/2006JD007375>, 2007.
- Chevallier, F., Palmer, P. I., Feng, L., Boesch, H., O'Dell, C. W., and Bousquet, P.: Toward robust and consistent regional  
750 CO<sub>2</sub> flux estimates from in situ and spaceborne measurements of atmospheric CO<sub>2</sub>, *Geophysical Research Letters*, 41, <https://doi.org/10.1002/2013GL058772>, 2014.
- Collatz, G. J., Ball, J. T., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, *Agricultural and Forest Meteorology*, 54, [https://doi.org/10.1016/0168-1923\(91\)90002-8](https://doi.org/10.1016/0168-1923(91)90002-8), 1991.
- 755 Cooper, E., Blyth, E., Cooper, H., Ellis, R., Pinnington, E., and Dadson, S. J.: Using data assimilation to optimize pedotransfer functions using field-scale in situ soil moisture observations, *Hydrology and Earth System Sciences*, 25, <https://doi.org/10.5194/hess-25-2445-2021>, 2021.
- Courtier, P., Thépaut, J., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, *Quarterly Journal of the Royal Meteorological Society*, 120, <https://doi.org/10.1002/qj.49712051912>, 1994.
- 760 Couvreur, F., Hourdin, F., Williamson, D., Roehrig, R., Volodina, V., Villefranque, N., Rio, C., Audouin, O., Salter, J., Bazile, E., Brient, F., Favot, F., Honnert, R., Lefebvre, M. P., Madeleine, J. B., Rodier, Q., and Xu, W.: Process-Based Climate Model Development Harnessing Machine Learning: I. A Calibration Tool for Parameterization Improvement, *Journal of Advances in Modeling Earth Systems*, 13, <https://doi.org/10.1029/2020MS002217>, 2021.
- Dagon, K., Sanderson, B. M., Fisher, R. A., and Lawrence, D. M.: A machine learning approach to emulation and biophysical parameter estimation with the Community Land Model, version 5, *Advances in Statistical Climatology, Meteorology and Oceanography*, 6, <https://doi.org/10.5194/ascmo-6-223-2020>, 2020.
- 765 de Rosnay, P., Polcher, J., Bruen, M., and Laval, K.: Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes, *Journal of Geophysical Research: Atmospheres*, 107, <https://doi.org/10.1029/2001jd000634>, 2002.
- 770 Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, I., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J. J., Park, B. K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Quarterly Journal of the Royal Meteorological Society*, 137, <https://doi.org/10.1002/qj.828>, 2011.
- 775 Douglas, N., Quaife, T., and Bannister, R.: Exploring a hybrid ensemble-variational data assimilation technique (4D-EnVar) with a simple ecosystem carbon model, *Environmental Modelling & Software*, 186, 106361, <https://doi.org/10.1016/j.envsoft.2025.106361>, 2025.
- Dufresne, J. L., Foujols, M. A., Denvil, S., Caubel, A., Marti, O., Aumont, O., Balkanski, Y., Bekki, S., Bellenger, H., Benshila, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., de Noblet, N., Duvel, J. P., Ethé, C., Fairhead, L., Fichet, T., Flavoni, S., Friedlingstein, P., Grandpeix, J. Y., Guez, L., Guilyardi, E., Hauglustaine, D., Hourdin, F., Idelkadi, A., Ghattas, J., Joussaume, S., Kageyama, M., Krinner, G., Labetoulle, S., Lahellec, A., Lefebvre, M. P., Lefevre, F., Levy, C., Li, Z. X., Lloyd, J., Lott, F., Madec, G., Mancip, M., Marchand, M., Masson, S., Meurdesoif, Y., Mignot, J., Musat, I., Parouty, S., Polcher, J., Rio,

C., Schulz, M., Swingedouw, D., Szopa, S., Talandier, C., Terray, P., Viovy, N., and Vuichard, N.: Climate change projections using the  
785 IPSL-CM5 Earth System Model: From CMIP3 to CMIP5, *Climate Dynamics*, 40, <https://doi.org/10.1007/s00382-012-1636-1>, 2013.

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, <https://doi.org/10.1017/cbo9780511535741>, 2002.

Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics,  
*Journal of Geophysical Research*, 99, <https://doi.org/10.1029/94jc00572>, 1994.

Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO<sub>2</sub> assimilation in  
790 leaves of C<sub>3</sub> species, *Planta*, 149, 1980.

Fisher, R. A. and Koven, C. D.: Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial  
Systems, <https://doi.org/10.1029/2018MS001453>, 2020.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C., Hauck, J., Landschützer, P., Quéré, C. L., Luijkx, I. T., Peters,  
G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L.,  
795 Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.  
T. T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas,  
J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Özgür Gürses, Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C.,  
Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarníková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D.,  
Goldewijk, K. K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot,  
800 N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S. I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar,  
A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck,  
C., Rosan, T. M., Schwinger, J., Séférian, R., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C.,  
Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe,  
M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, *Earth*  
805 *System Science Data*, 15, <https://doi.org/10.5194/essd-15-5301-2023>, 2023.

Geman, S., Bienenstock, E., and Doursat, R.: Neural Networks and the Bias/Variance Dilemma, *Neural Computation*, 4,  
<https://doi.org/10.1162/neco.1992.4.1.1>, 1992.

Giering, R. and Kaminski, T.: Applying TAF to generate efficient derivative code of Fortran 77-95 programs, *PAMM*, 2,  
<https://doi.org/10.1002/pamm.200310014>, 2003.

810 Gilbert, J. C. and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-Newton algorithms, *Mathematical Programming*,  
45, <https://doi.org/10.1007/BF01589113>, 1989.

GLOBALVIEW-CO2: GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project–Carbon Dioxide, NOAA/CMDL, 2013.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y. H., Clais, 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.,  
815 Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C. W.: Towards robust regional estimates of CO2 sources and sinks using atmospheric  
transport models, *Nature*, 415, <https://doi.org/10.1038/415626a>, 2002.

Hodson, T. O., Over, T. M., and Foks, S. S.: Mean Squared Error, Deconstructed, *Journal of Advances in Modeling Earth Systems*, 13,  
<https://doi.org/10.1029/2021MS002681>, 2021.

Hourdin, F. and Armengaud, A.: The use of finite-volume methods for atmospheric advection of trace species. Part I:  
820 Test of various formulations in a general circulation model, *Monthly Weather Review*, 127, [https://doi.org/10.1175/1520-0493\(1999\)127<0822:TUOFVM>2.0.CO;2](https://doi.org/10.1175/1520-0493(1999)127<0822:TUOFVM>2.0.CO;2), 1999.

- Hourdin, F. and Talagrand, O.: Eulerian backtracking of atmospheric tracers. I: Adjoint derivation and parametrization of subgrid-scale transport, *Quarterly Journal of the Royal Meteorological Society*, 132, <https://doi.org/10.1256/qj.03.198.A>, 2006.
- Hourdin, F., Talagrand, O., and Idelkadi, A.: Eulerian backtracking of atmospheric tracers. II: Numerical aspects, *Quarterly Journal of the Royal Meteorological Society*, 132, <https://doi.org/10.1256/qj.03.198.B>, 2006.
- Hourdin, F., Mauritsen, T., Gettelman, A., Golaz, J. C., Balaji, V., Duan, Q., Folini, D., Ji, D., Klocke, D., Qian, Y., Rauser, F., Rio, C., Tomassini, L., Watanabe, M., and Williamson, D.: The art and science of climate model tuning, *Bulletin of the American Meteorological Society*, 98, <https://doi.org/10.1175/BAMS-D-15-00135.1>, 2017.
- Hourdin, F., Williamson, D., Rio, C., Couvreux, F., Roehrig, R., Villefranque, N., Musat, I., Fairhead, L., Diallo, F. B., and Volodina, V.: Process-Based Climate Model Development Harnessing Machine Learning: II. Model Calibration From Single Column to Global, *Journal of Advances in Modeling Earth Systems*, 13, <https://doi.org/10.1029/2020MS002225>, 2021.
- Hourdin, F., Ferster, B., Deshayes, J., Mignot, J., Musat, I., and Williamson, D.: Toward machine-assisted tuning avoiding the underestimation of uncertainty in climate change projections, *Science Advances*, 9, <https://doi.org/10.1126/sciadv.adf2758>, 2023.
- IPCC: Human Influence on the Climate System, <https://doi.org/10.1017/9781009157896.005>, 2023.
- Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three-dimensional global inverse model of the atmospheric transport 2. Inversion of the transport of CO<sub>2</sub> in the 1980s, *Journal of Geophysical Research Atmospheres*, 104, <https://doi.org/10.1029/1999JD900146>, 1999a.
- Kaminski, T., Heimann, M., and Giering, R.: A coarse grid three-dimensional global inverse model of the atmospheric transport 1. Adjoint model and Jacobian matrix, *Journal of Geophysical Research Atmospheres*, 104, <https://doi.org/10.1029/1999JD900147>, 1999b.
- Kaminski, T., Knorr, W., Rayner, P. J., and Heimann, M.: Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Global Biogeochemical Cycles*, 16, <https://doi.org/10.1029/2001gb001463>, 2002.
- Kaminski, T., Knorr, W., Scholze, M., Gobron, N., Pinty, B., Giering, R., and Mathieu, P. P.: Consistent assimilation of MERIS FAPAR and atmospheric CO<sub>2</sub> into a terrestrial vegetation model and interactive mission benefit analysis, *Biogeosciences*, 9, <https://doi.org/10.5194/bg-9-3173-2012>, 2012.
- Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 63, <https://doi.org/10.1111/1467-9868.00294>, 2001.
- King, R. C., Mansfield, L. A., and Sheshadri, A.: Bayesian History Matching Applied to the Calibration of a Gravity Wave Parameterization, *Journal of Advances in Modeling Earth Systems*, 16, e2023MS004163, <https://doi.org/https://doi.org/10.1029/2023MS004163>, e2023MS004163 2023MS004163, 2024.
- Knorr, W. and Heimann, M.: Impact of drought stress and other factors on seasonal land biosphere CO<sub>2</sub> exchange studied through an atmospheric tracer transport model, *Tellus B*, 47, <https://doi.org/10.1034/j.1600-0889.47.issue4.7.x>, 1995.
- Knorr, W., Williams, M., Thum, T., Kaminski, T., Voßbeck, M., Scholze, M., Quaife, T., Smallmann, L., Steele-Dunne, S., Vreugdenhil, M., Green, T., Zähle, S., Aurela, M., Bouvet, A., Bueechi, E., Dorigo, W., El-Madany, T., Migliavacca, M., Honkanen, M., Kerr, Y., Kontu, A., Lemmetyinen, J., Lindqvist, H., Mialon, A., Miinalainen, T., Pique, G., Ojasalo, A., Quegan, S., Rayner, P., Reyes-Muñoz, P., Rodríguez-Fernández, N., Schwank, M., Verrelst, J., Zhu, S., Schüttemeyer, D., and Drusch, M.: A comprehensive land surface vegetation model for multi-stream data assimilation, *D&B v1.0, EGU sphere*, 2024, 1–40, <https://doi.org/10.5194/egusphere-2024-1534>, 2024.
- 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, <https://doi.org/10.1029/2003GB002199>, 2005.
- Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F., and Richardson, A. D.: Constraining a global ecosystem model with multi-site eddy-covariance data, *Biogeosciences*, 9, <https://doi.org/10.5194/bg-9-3757-2012>, 2012.

- 860 Kuppel, S., Chevallier, F., and Peylin, P.: Quantifying the model structural error in carbon cycle data assimilation systems, *Geoscientific Model Development*, 6, <https://doi.org/10.5194/gmd-6-45-2013>, 2013.
- Lardy, R., Bellocchi, G., and Soussana, J. F.: A new method to determine soil organic carbon equilibrium, *Environmental Modelling and Software*, 26, <https://doi.org/10.1016/j.envsoft.2011.05.016>, 2011.
- Leer, B. V.: Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, *Journal of Computational*  
865 *Physics*, 23, [https://doi.org/10.1016/0021-9991\(77\)90095-X](https://doi.org/10.1016/0021-9991(77)90095-X), 1977.
- Liu, C., Xiao, Q., and Wang, B.: An ensemble-based four-dimensional variational data assimilation scheme. Part I: Technical formulation and preliminary test, *Monthly Weather Review*, 136, <https://doi.org/10.1175/2008MWR2312.1>, 2008.
- Liu, J., Baskaran, L., Bowman, K., Schimel, D., Bloom, A. A., Parazoo, C. N., Oda, T., Carroll, D., Menemenlis, D., Joiner, J., Commane, R., Daube, B., Gatti, V. L., McKain, K., Miller, J., Stephens, B. B., Sweeney, C., and Wofsy, S.: Carbon Monitoring System Flux Net  
870 *Biosphere Exchange 2020 (CMS-Flux NBE 2020)*, *Earth System Science Data*, 13, <https://doi.org/10.5194/essd-13-299-2021>, 2021.
- Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreur, F., Grandpeix, J. Y., Lefebvre, M. P., Rio, C., Bergamaschi, P., Chambers, S. D., Karstens, U., Kazan, V., Laan, S. V. D., Meijer, H. A., Moncrieff, J., Ramonet, M., Scheeren, H. A., Schlosser, C., Schmidt, M., Vermeulen, A., and Williams, A. G.: Atmospheric transport and chemistry of trace gases in LMDz5B: Evaluation and implications for inverse modelling, *Geoscientific Model Development*, 8, <https://doi.org/10.5194/gmd-8-129-2015>, 2015.
- 875 Lurton, T., Balkanski, Y., Bastrikov, V., Bekki, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Contoux, C., Cozic, A., Cugnet, D., Dufresne, J. L., Éthé, C., Foujols, M. A., Ghattas, J., Hauglustaine, D., Hu, R. M., Kageyama, M., Khodri, M., Lebas, N., Levvasseur, G., Marchand, M., Ottlé, C., Peylin, P., Sima, A., Szopa, S., Thiéblemont, R., Vuichard, N., and Boucher, O.: Implementation of the CMIP6 Forcing Data in the IPSL-CM6A-LR Model, *Journal of Advances in Modeling Earth Systems*, 12, <https://doi.org/10.1029/2019MS001940>, 2020.
- 880 MacBean, N., Maignan, F., Peylin, P., Bacour, C., Bréon, F. M., and Ciais, P.: Using satellite data to improve the leaf phenology of a global terrestrial biosphere model, *Biogeosciences*, 12, <https://doi.org/10.5194/bg-12-7185-2015>, 2015.
- MacBean, N., Peylin, P., Chevallier, F., Scholze, M., and Schürmann, G.: Consistent assimilation of multiple data streams in a carbon cycle data assimilation system, *Geoscientific Model Development*, 9, <https://doi.org/10.5194/gmd-9-3569-2016>, 2016.
- MacBean, N., Bacour, C., Raoult, N., Bastrikov, V., Koffi, E. N., Kuppel, S., Maignan, F., Ottlé, C., Peaucelle, M., Santaren, D., and  
885 Peylin, P.: Quantifying and Reducing Uncertainty in Global Carbon Cycle Predictions: Lessons and Perspectives From 15 Years of Data Assimilation Studies With the ORCHIDEE Terrestrial Biosphere Model, *Global Biogeochemical Cycles*, 36, e2021GB007177, <https://doi.org/https://doi.org/10.1029/2021GB007177>, e2021GB007177 2021GB007177, 2022.
- McNeill, D., Robertson, E., and Wiltshire, A.: Constraining the carbon cycle in JULES-ES-1.0, *Geoscientific Model Development*, 17, <https://doi.org/10.5194/gmd-17-1059-2024>, 2024.
- 890 ORCHIDEE: ORCHIDEE V22 r7878 gmd 2025 4DEnVar, <https://doi.org/10.14768/c68bc728-da71-4383-84df-dcde31d9c006>, 2025.
- Peylin, P., Bousquet, P., Quéré, C. L., Sitch, S., Friedlingstein, P., McKinley, G., Gruber, N., Rayner, P., and Ciais, P.: Multiple constraints on regional CO<sub>2</sub> flux variations over land and oceans, *Global Biogeochemical Cycles*, 19, <https://doi.org/10.1029/2003GB002214>, 2005.
- 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., Laan-Luijkx, I. T. V. D., and Zhang, X.: Global atmospheric carbon budget: Results from an ensemble of atmospheric CO<sub>2</sub> inversions,  
895 *Biogeosciences*, 10, <https://doi.org/10.5194/bg-10-6699-2013>, 2013.

- Peylin, P., Bacour, C., MacBean, N., Leonard, S., Rayner, P., Kuppel, S., Koffi, E., Kane, A., Maignan, F., Chevallier, F., Ciais, P., and Prunet, P.: A new stepwise carbon cycle data assimilation system using multiple data streams to constrain the simulated land surface carbon cycle, *Geoscientific Model Development*, 9, <https://doi.org/10.5194/gmd-9-3321-2016>, 2016.
- Pinnington, E., Quaife, T., Lawless, A., Williams, K., Arkebauer, T., and Scoby, D.: The Land Variational Ensemble Data Assimilation Framework: LAVENDAR v1.0.0, *Geoscientific Model Development*, 13, <https://doi.org/10.5194/gmd-13-55-2020>, 2020.
- Pinnington, E., Amezcu, J., Cooper, E., Dadson, S., Ellis, R., Peng, J., Robinson, E., Morrison, R., Osborne, S., and Quaife, T.: Improving soil moisture prediction of a high-resolution land surface model by parameterising pedotransfer functions through assimilation of SMAP satellite data, *Hydrology and Earth System Sciences*, 25, <https://doi.org/10.5194/hess-25-1617-2021>, 2021.
- Plessix, R. E.: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, <https://doi.org/10.1111/j.1365-246X.2006.02978.x>, 2006.
- Poulter, B., MacBean, N., Hartley, A., Khlystova, I., Arino, O., Betts, R., Bontemps, S., Boettcher, M., Brockmann, C., Defourny, P., Hagemann, S., Herold, M., Kirches, G., Lamarche, C., Lederer, D., Ottlé, C., Peters, M., and Peylin, P.: Plant functional type classification for earth system models: Results from the European Space Agency's Land Cover Climate Change Initiative, *Geoscientific Model Development*, 8, <https://doi.org/10.5194/gmd-8-2315-2015>, 2015.
- Randerson, J., van der Werf, G., Giglio, L., Collatz, G., and Kasibhatla, P.: Global Fire Emissions Database, Version 3.1, <https://doi.org/10.3334/ORNLDAAAC/1191>, 2013.
- Raoult, N., Beylat, S., Salter, J. M., Hourdin, F., Bastrikov, V., Ottlé, C., and Peylin, P.: Exploring the potential of history matching for land surface model calibration, *Geoscientific Model Development*, 17, 5779–5801, <https://doi.org/10.5194/gmd-17-5779-2024>, 2024a.
- Raoult, N., Douglas, N., MacBean, N., Kolassa, J., Quaife, T., Roberts, A. G., Fisher, R. A., Fer, I., Bacour, C., Dagon, K., Hawkins, L., Carvalhais, N., Cooper, E., Dietze, M., Gentine, P., Kaminski, T., Kennedy, D., Liddy, H. M., Moore, D., Peylin, P., Pinnington, E., Sanderson, B. M., Scholze, M., Seiler, C., Smallman, T. L., Vergopolan, N., Viskari, T., Williams, M., and Zobitz, J.: Parameter Estimation in Land Surface Models: Challenges and Opportunities with Data Assimilation and Machine Learning, ESS Open Archive, <https://doi.org/10.22541/essoar.172838640.01153603/v1>, 2024b.
- Raoult, N. M., Jupp, T. E., Cox, P. M., and Luke, C. M.: Land-surface parameter optimisation using data assimilation techniques: The adJULES system V1.0, *Geoscientific Model Development*, 9, <https://doi.org/10.5194/gmd-9-2833-2016>, 2016.
- Rayner, P. J.: The current state of carbon-cycle data assimilation, <https://doi.org/10.1016/j.cosust.2010.05.005>, 2010.
- Rayner, P. J., Enting, I. G., Francey, R. J., and Langenfelds, R.: Reconstructing the recent carbon cycle from atmospheric CO<sub>2</sub>,  $\delta^{13}\text{C}$  and O<sub>2</sub>/N<sub>2</sub> observations, *Tellus, Series B: Chemical and Physical Meteorology*, 51, <https://doi.org/10.1034/j.1600-0889.1999.t01-1-00008.x>, 1999.
- Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R., and Widmann, H.: Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochemical Cycles*, 19, <https://doi.org/10.1029/2004GB002254>, 2005.
- Rayner, P. J., Michalak, A. M., and Chevallier, F.: Fundamentals of data assimilation applied to biogeochemistry, *Atmospheric Chemistry and Physics*, 19, <https://doi.org/10.5194/acp-19-13911-2019>, 2019.
- Remaud, M., Chevallier, F., Cozic, A., Lin, X., and Bousquet, P.: On the impact of recent developments of the LMDz atmospheric general circulation model on the simulation of CO<sub>2</sub> transport, *Geoscientific Model Development*, 11, <https://doi.org/10.5194/gmd-11-4489-2018>, 2018.
- Santaren, D., Peylin, P., Viovy, N., and Ciais, P.: Optimizing a process-based ecosystem model with eddy-covariance flux measurements: A pine forest in southern France, *Global Biogeochemical Cycles*, 21, <https://doi.org/10.1029/2006GB002834>, 2007.

- Santaren, D., Peylin, P., Bacour, C., Ciais, P., and Longdoz, B.: Ecosystem model optimization using in situ flux observations: Benefit of Monte Carlo versus variational schemes and analyses of the year-to-year model performances, *Biogeosciences*, 11, <https://doi.org/10.5194/bg-11-7137-2014>, 2014.
- Scholze, M., Kaminski, T., Rayner, P., Knorr, W., and Giering, R.: Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, *Journal of Geophysical Research Atmospheres*, 112, <https://doi.org/10.1029/2007JD008642>, 2007.
- Schürmann, G. J., Kaminski, T., Köstler, C., Carvalhais, N., Voßbeck, M., Kattge, J., Giering, R., Rödenbeck, C., Heimann, M., and Zaehe, S.: Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1.0, *Geoscientific Model Development*, 9, <https://doi.org/10.5194/gmd-9-2999-2016>, 2016.
- Sitch, S., O’Sullivan, M., Robertson, E., Friedlingstein, P., Albergel, C., Anthoni, P., Arneth, A., Arora, V. K., Bastos, A., Bastrikov, V., Bellouin, N., Canadell, J. G., Chini, L., Ciais, P., Falk, S., Harris, I., Hurtt, G., Ito, A., Jain, A. K., Jones, M. W., Joos, F., Kato, E., Kennedy, D., Klein Goldewijk, K., Kluzek, E., Knauer, J., Lawrence, P. J., Lombardozzi, D., Melton, J. R., Nabel, J. E. M. S., Pan, N., Peylin, P., Pongratz, J., Poulter, B., Rosan, T. M., Sun, Q., Tian, H., Walker, A. P., Weber, U., Yuan, W., Yue, X., and Zaehe, S.: Trends and Drivers of Terrestrial Sources and Sinks of Carbon Dioxide: An Overview of the TRENDY Project, *Global Biogeochemical Cycles*, 38, e2024GB008102, <https://doi.org/https://doi.org/10.1029/2024GB008102>, e2024GB008102 2024GB008102, 2024.
- Student: The probable error of a mean, *Biometrika*, 6, 1–25, <https://doi.org/10.1093/biomet/6.1.1>, 1908.
- Talagrand, O. and Courtier, P.: Variational Assimilation of Meteorological Observations With the Adjoint Vorticity Equation. I: Theory, *Quarterly Journal of the Royal Meteorological Society*, 113, <https://doi.org/10.1002/qj.49711347812>, 1987.
- Tarantola, A.: Inverse problem theory: methods for data fitting and model parameter estimation., *Inverse problem theory: methods for data fitting and model parameter estimation.*, [https://doi.org/10.1016/0031-9201\(89\)90124-6](https://doi.org/10.1016/0031-9201(89)90124-6), 1987.
- Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, <https://doi.org/10.1137/1.9780898717921>, 2005.
- Tiedtke, M.: A comprehensive mass flux scheme for cumulus parameterization in large-scale models, *Monthly Weather Review*, 117, [https://doi.org/10.1175/1520-0493\(1989\)117<1779:ACMFSF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2), 1989.
- Wang, F., Cheruy, F., and Dufresne, J. L.: The improvement of soil thermodynamics and its effects on land surface meteorology in the IPSL climate model, *Geoscientific Model Development*, 9, <https://doi.org/10.5194/gmd-9-363-2016>, 2016.
- Watson-Parris, D., Williams, A., Deaconu, L., and Stier, P.: Model calibration using ESEm v1.1.0-an open, scalable Earth system emulator, *Geoscientific Model Development*, 14, <https://doi.org/10.5194/gmd-14-7659-2021>, 2021.
- Williamson, D., Goldstein, M., Allison, L., Blaker, A., Challenor, P., Jackson, L., and Yamazaki, K.: History matching for exploring and reducing climate model parameter space using observations and a large perturbed physics ensemble, *Climate Dynamics*, 41, <https://doi.org/10.1007/s00382-013-1896-4>, 2013.
- Williamson, D. B., Blaker, A. T., and Sinha, B.: Tuning without over-tuning: Parametric uncertainty quantification for the NEMO ocean model, *Geoscientific Model Development*, 10, <https://doi.org/10.5194/gmd-10-1789-2017>, 2017.
- Yin, X. and Struik, P. C.: C3 and C4 photosynthesis models: An overview from the perspective of crop modelling, *NJAS - Wageningen Journal of Life Sciences*, 57, <https://doi.org/10.1016/j.njas.2009.07.001>, 2009.
- Ziehn, T., Scholze, M., and Knorr, W.: On the capability of Monte Carlo and adjoint inversion techniques to derive posterior parameter uncertainties in terrestrial ecosystem models, *Global Biogeochemical Cycles*, 26, <https://doi.org/10.1029/2011gb004185>, 2012.
- Zobler, L.: Global Soil Types, 1-Degree Grid (Zobler), <https://doi.org/10.3334/ORNLDAAAC/418>, 1999.