University of Reading

Towards the assimilation of atmospheric CO2 concentration data in a land surface model using adjoint-free variational methods

Article

Published Version

Creative Commons: Attribution 4.0 (CC-BY)

Open Access

Beylat, S. ORCID: https://orcid.org/0009-0006-6315-4527, Raoult, N. ORCID: https://orcid.org/0000-0003-2907-9456, Bacour, C. ORCID: https://orcid.org/0000-0002-1913-3722, Douglas, N. ORCID: https://orcid.org/0000-0002-3404-8761, Quaife, T. ORCID: https://orcid.org/0000-0001-6896-4613, Bastrikov, V., Rayner, P. J. and Peylin, P. (2025) Towards the assimilation of atmospheric CO2 concentration data in a land surface model using adjoint-free variational methods. Geoscientific Model Development, 18 (20). pp. 7501-7527. ISSN 1991-9603 doi: 10.5194/gmd-18-7501-2025 Available at https://centaur.reading.ac.uk/125426/

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

To link to this article DOI: http://dx.doi.org/10.5194/gmd-18-7501-2025

Publisher: European Geosciences Union

University of Reading

All outputs in CentAUR are protected by Intellectual Property Rights law, including copyright law. Copyright and IPR is retained by the creators or other copyright holders. Terms and conditions for use of this material are defined in the End User Agreement.

www.reading.ac.uk/centaur

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Geosci. Model Dev., 18, 7501–7527, 2025
https://doi.org/10.5194/gmd-18-7501-2025
© Author(s) 2025. This work is distributed under
the Creative Commons Attribution 4.0 License.

Geoscientific Model Development
Open Access
EGU

Towards the assimilation of atmospheric CO_2 concentration data in a land surface model using adjoint-free variational methods

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

[1]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
[2]School of Geography, Earth and Atmospheric Sciences, University of Melbourne, Parkville, 3010 Victoria, Australia
[3]European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, UK
[4]National Centre for Earth Observation, Department of Meteorology, University of Reading, Reading, UK
[5]Science Partners, Paris, France

Correspondence: Simon Beylat (simon.beylat@lsce.ipsl.fr)

Received: 10 January 2025 – Discussion started: 30 January 2025
Revised: 24 July 2025 – Accepted: 31 July 2025 – Published: 21 October 2025

Abstract. A comprehensive understanding and 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_2 concentration data for parameter calibration of the ORganizing Carbon and Hydrology In Dynamic EcosystEms (ORCHIDEE) land surface model using an EnVarDA method, an adjoint-free ensemble-variational data assimilation method. By circumventing the challenges associated with developing and maintaining tangent linear and adjoint models, the EnVarDA method offers a very promising alternative. Using synthetic observations generated through a twin experiment, we demonstrate the ability of EnVarDA to assimilate atmospheric CO_2 concentrations for model parameter calibration. We then compare the results to a VarDA method that uses finite differences to estimate tangent linear and adjoint models, which reveals that EnVarDA is superior in terms of computational efficiency, fit to the observations, and parameter recovery.

1 Introduction

Since the link between the increase in atmospheric CO_2 concentrations 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 mechanisms 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_2 concentration data have long been considered a rich source of information to understand the global carbon cycle and characterise the spatio-temporal variation in natural CO_2 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, 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_2 inversions have been used to estimate natural CO_2 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_2 concentrations based on a CO_2 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_2 surface flux scenario as prior information, atmospheric inversions statistically adjust the surface CO_2 fluxes, minimising the differences between observed and modelled concentrations. This statistical optimisation generally assumes that the corrections to CO_2 surface fluxes are isotropic in time and space. This suggests that errors in surface fluxes are only correlated in space by the 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 net surface fluxes, with no direct information on the underlying components (i.e. photosynthesis uptake, ecosystem respiration release, fire release). Consequently, this approach is also not suitable for making 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 soil 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 parameterisations 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_2 concentration data, which offer 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). Most of the methods used for parameter calibration are derived from Bayesian formulations of inverse problems and are defined here as variational data assimilation (VarDA) methods. The VarDA method is inspired by the four-dimensional variational (4DVar) 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 has also been employed in atmospheric inversions to correct surface CO_2 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 typically based on a least-squares criterion. This cost function calculates two terms: (i) an observation term that computes the difference between observations and model outputs and (ii) a background term that incorporates prior knowledge of the state. The computation of both terms is performed in space and time. We define the VarDA method here, as our focus is not on directly optimising the prior state. Instead, we concentrate on time-invariant parameters used in the parameterisation that define 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 differs from the conventional 4DVar framework, which generally uses sequential cycles with shorter assimilation windows. In order to minimise this cost function, the 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 models (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 must be made to ensure 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., 2025); however, for complex community models such as ORganizing Carbon and Hydrology In Dynamic EcosystEms (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 order to use the 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 the 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 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 – for instance emulator-based methods (Williamson et al., 2013; Couvreux et al., 2021) – 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 models (Williamson et al., 2017; Hourdin et al., 2021; King et al., 2024). In these methods, the model is 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 a 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_2 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 Institut Pierre-Simon-Laplace (IPSL) 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_2 concentrations. We demonstrate the potential of EnVarDA using synthetic observation data according to different criteria: (i) the differences between synthetic observations and simulations of atmospheric CO_2 concentrations, (ii) the spatial distribution of carbon fluxes and their subcomponents, and (iii) the recovery of the true parameters used to generate the synthetic observations. We also compare the performance of EnVarDA using these criteria with that of VarDA with finite differences. Section 2 presents the methods, the models, the data, and the experiments. Results are shown in Sect. 3, with discussions and conclusions in Sects. 4 and 5, respectively.

2 Method

2.1 Models and datasets

2.1.1 ORCHIDEE land surface model

ORganizing Carbon and Hydrology In Dynamic EcosystEms (ORCHIDEE; 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 min and a slow one that simulates carbon allocation in plant reservoirs, soil carbon dynamics, and litter decomposition every day. In this study, we used 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 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 and 90 m (Wang et al., 2016). The hydrological 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 with three categories. The set of equations governing the soil organic matter (SOM) pools and their temporal evolution has analytical solutions 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 C_3 plants and Collatz et al. (1991) for C_4 plants. The ORCHIDEE LSM uses different types

of vegetation grouped into plant functional types (PFTs) with similar structural characteristics. It distinguishes between 14 vegetation PFT classes, described 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 an LC-to-PFT cross-walking approach (Poulter et al., 2015) (see https://orchidas.lsce.ipsl.fr/dev/lccci/, last access: 24 July 2025).

In this study, ORCHIDEE is run offline using 3 h 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° latitude × 3.75° longitude). 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_2 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 are 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_2 concentration and LUC 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 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 the Van Leer scheme (Van Leer, 1977), the deep convection is parameterised 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° (latitude) × 3.75° (longitude) and 19 sigma-pressure layers up to 3 hPa. The calculated winds (u and v) used to drive LMDZ are provided by 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 in 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_2 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 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_2 concentration data using the NOAA Earth System Laboratory's collaborative product (GLOBALVIEW-CO2, 2013). They model average monthly concentrations at 53 stations over the period of 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

Figure 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_2 fluxes

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

- The oceanic flux component was derived from a neural network model which estimated the spatial and temporal variations in CO_2 fluxes between the air and the sea (Peylin et al., 2016).
- The global maps of biomass burning emissions are taken from the Global Fire Emissions Database version 3 (Randerson et al., 2013).

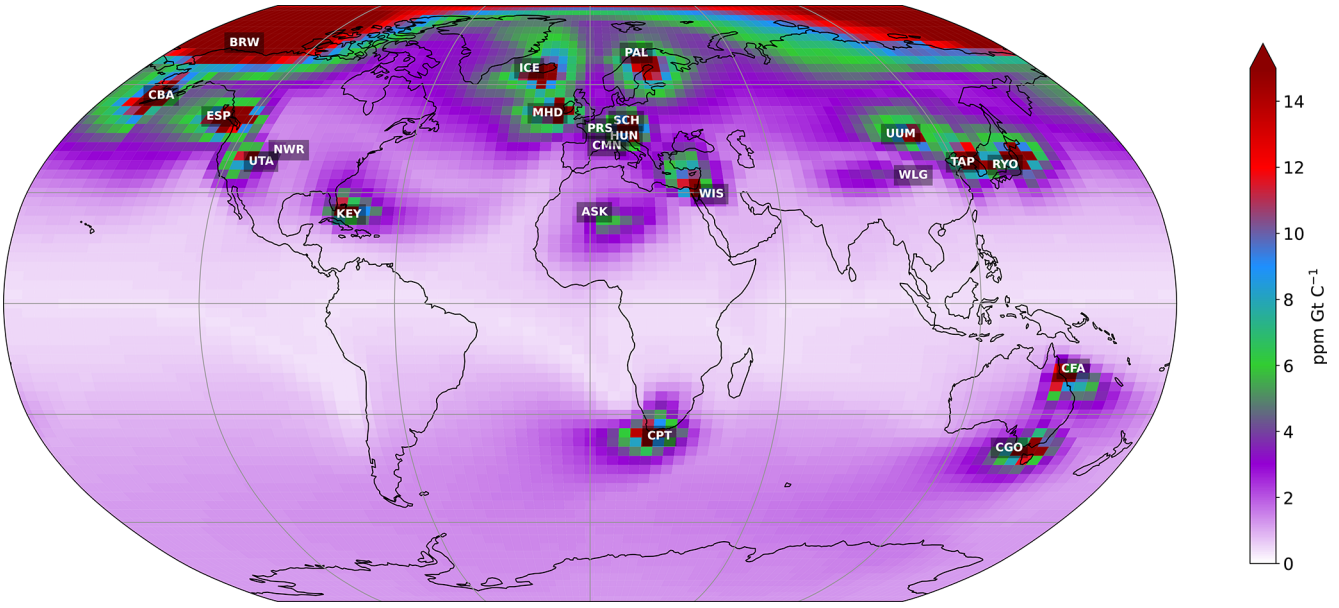

Figure 1. Monthly mean sensitivity map of atmospheric CO_2 concentrations to land carbon fluxes at the 21 stations considered over the 2000–2001 period. The average sensitivity map is obtained by deriving, for each atmospheric station and each of the 24 months, the map of the average daily sensitivity of the atmospheric concentration of CO_2 to surface carbon flux (in ppm GtC^{-1}) over the last 6 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_2, 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 details of the stations at https://gml.noaa.gov/dv/site/index.php, last access: 5 June 2025).

– The fossil fuel CO_2 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.

2.2 Data assimilation framework

2.2.1 A Bayesian setup

First, let us define a general Bayesian framework, mainly following Tarantola (1987, 2005), which accounts for both model/observation error and an a priori background error. Taking the approach of Kennedy and O'Hagan (2001) for an observational constraint $\boldsymbol{y}$, let

$$\boldsymbol{y} = \boldsymbol{\mathcal{Y}} + \boldsymbol{e}, \tag{1}$$

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

$$\boldsymbol{y} = \boldsymbol{\mathcal{Y}} + \boldsymbol{e} = \boldsymbol{\mathcal{H}}\left(\boldsymbol{x}^*\right) + \boldsymbol{\eta} + \boldsymbol{e}, \tag{2}$$

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

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

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

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

2.2.2 The VarDA method

Standard VarDA

In this section, we present the VarDA method. Maximising the probability in Eq. (4) is equivalent to minimising the following function, usually referred to as the VarDA *cost function*:

$$J(\boldsymbol{x}) = \frac{1}{2}\sum_{t}^{N_t}(\boldsymbol{\mathcal{H}}_t(\boldsymbol{x})-\boldsymbol{y}_t)^T\mathbf{R}_t^{-1}(\boldsymbol{\mathcal{H}}_t(\boldsymbol{x})-\boldsymbol{y}_t) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_b)^T\mathbf{B}^{-1}(\boldsymbol{x}-\boldsymbol{x}_b), \quad (5)$$

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

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

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

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \left(\nabla^2 J(\boldsymbol{x}_i)\right)^{-1} \times \nabla J(\boldsymbol{x}_i). \quad (7)$$

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

$$\begin{aligned}\nabla J(\boldsymbol{x}_i) &= \mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\boldsymbol{x}_i-\boldsymbol{y})+\mathbf{B}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_b),\\ \nabla^2 J(\boldsymbol{x}_i) &= \mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\mathbf{B}^{-1}.\end{aligned} \quad (8)$$

We can update Eq. (7) using Eq. (8):

$$\boldsymbol{x}_{i+1} = \boldsymbol{x}_i - \left[\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H}+\mathbf{B}^{-1}\right]^{-1}\left[\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\boldsymbol{x}_i-\boldsymbol{y}) + \mathbf{B}^{-1}(\boldsymbol{x}_i-\boldsymbol{x}_b)\right]. \quad (9)$$

Here, the notation $\boldsymbol{\mathcal{H}}$ becomes $\mathbf{H}$ because it does not represent the use of the direct operator $\boldsymbol{\mathcal{H}}$. 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 VarDA variant: $\boldsymbol{\epsilon}$-VarDA

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

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

where $\Delta\boldsymbol{x}$ represents a small change in $\boldsymbol{x}$. 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 $\Delta\boldsymbol{x}$. A selection of $\Delta\boldsymbol{x}$ that is too small may lead to $\mathbf{H}$ being insensitive to the parameter vector; i.e. $\boldsymbol{\mathcal{H}}(\boldsymbol{x}+\Delta\boldsymbol{x})-\boldsymbol{\mathcal{H}}(\boldsymbol{x}) = \mathbf{H}\Delta\boldsymbol{x} \approx \mathbf{0}$. This leads to the term corresponding to the difference between the observation and the output's operator ($\mathbf{H}\boldsymbol{x}_i-\boldsymbol{y}$) becoming negligible in Eq. (8) and hence resulting in an ineffective minimisation. By contrast, if the choice of $\Delta\boldsymbol{x}$ is too large, the result gives inaccurate tangent linear and adjoint models that lose their local vision around $\boldsymbol{x}$. This results in a large loss of information and therefore a much less accurate minimisation. In our case, we define $\boldsymbol{\epsilon}$ such that $\Delta\boldsymbol{x} = \boldsymbol{x}_{\text{range}} \times \boldsymbol{\epsilon}$, where $\boldsymbol{x}_{\text{range}} = \boldsymbol{x}_{\max} - \boldsymbol{x}_{\min}$, and we refer to this method as $\boldsymbol{\epsilon}$-VarDA. Due to this approximation, $\boldsymbol{\epsilon}$-VarDA is therefore not entirely equivalent to standard VarDA.

2.2.3 The EnVarDA method

From VarDA to EnVarDA

We present here an implementation of VarDA that we do not use in this study but that is important for understanding the EnVarDA method. This implementation is presented in several studies (Courtier et al., 1994; Gilbert and Lemaréchal, 1989; Liu et al., 2008; Bannister, 2016; 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 $\boldsymbol{w}$ to ensure that the 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

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

where $\boldsymbol{x}_a$ 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):

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

and its gradient:

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

EnVarDA

The EnVarDA method described in Liu et al. (2008) and Pinnington et al. (2020) incorporates an aspect of the ensemble Kalman filter (EnKF) in order to avoid the calculation of tangent linear or adjoint models necessary for 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). The perturbation matrix is defined as:

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

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

$$\mathbf{B} \approx \mathbf{X}'_B \mathbf{X}'^T_B. \quad (16)$$

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

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

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

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

and the gradient in Eq. (14) becomes

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

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

$$\mathbf{HX}'_b \approx \frac{1}{\sqrt{N-1}} (\boldsymbol{\mathcal{H}}(\boldsymbol{x}_1) - \boldsymbol{\mathcal{H}}(\boldsymbol{x}_b); \boldsymbol{\mathcal{H}}(\boldsymbol{x}_2) - \boldsymbol{\mathcal{H}}(\boldsymbol{x}_b), \ldots, \boldsymbol{\mathcal{H}}(\boldsymbol{x}_N) - \boldsymbol{\mathcal{H}}(\boldsymbol{x}_b)). \quad (20)$$

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

$$\nabla J(\boldsymbol{w}) = \left(\mathbf{HX}'_b\right)^T \mathbf{R}^{-1} \left(\mathbf{HX}'_b \boldsymbol{w} + \boldsymbol{\mathcal{H}}(\boldsymbol{x}_b) - \boldsymbol{y}\right) + \boldsymbol{w}. \quad (21)$$

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

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

2.2.4 Implementation into ORCHIDAS

The ORCHIDEE data assimilation system (ORCHIDAS) is a system designed 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 the water and energy budget, methane, and nitrogen (see the full list of studies published at https://orchidas.lsce.ipsl.fr/publications.php, last access: 24 July 2025).

This system has long used VarDA as described in Sect. 2.2.2, but it also allows for 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 EnVarDA method as described in the EnVarDA section.

2.3 Experiment design

2.3.1 Twin experiment description

To test the data assimilation methods presented in Sect. 2.2, we conducted a so-called *twin experiment* to evaluate their efficiency in calibrating parameters involved in calculating

net biome productivity (NBP) fluxes in the ORCHIDEE LSM. This experimental framework reduces the complexities associated with model–data errors, focusing on the performance of the assimilation methods. The known "true" parameters, which are the default parameter values of the ORCHIDEE model, are used to generate the synthetic observations. New values of a priori parameters are manually generated, ensuring physically meaningful values that differ from the true parameters, both of which are 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_2 concentrations from the 21 continental stations are assimilated simultaneously over a 2-year window (2000–2001) to monitor spatial and temporal variations in carbon fluxes, as shown in Fig. 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 at the global scale using the ORCHIDEE LSM 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, calculated as the difference between emission fluxes (heterotrophic and autotrophic respiration and disturbance fluxes due to land-use change) and sink fluxes (primarily due to photosynthesis). The concentrations given by the surface fluxes (the simulated NBP fluxes, along with other fluxes described in Sect. 2.1.4) are transported using pre-calculated transport fields of the LMDZ model over the 2000–2001 period. We then extract atmospheric CO_2 concentrations at 21 continental atmospheric stations, shown in Fig. 1, which are highly sensitive to continental carbon fluxes, providing significant constraints on the parameters. This process enabled the generation of synthetic observations of monthly average atmospheric CO_2 concentrations at these 21 stations over the 2-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 the simulation. This means that there is at least one solution where the model can perfectly match the synthetic observation.

2.3.3 Simplified case

First, we focus on a simplified case involving the calibration of only one PFT-dependent parameter: Vcmax, which controls the maximum rate of carboxylation limited by RuBisCO activity at 25 °C. This parameter was chosen because its impact on the atmospheric 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 Vcmax for the 14 PFTs, resulting in the calibration of 14 parameters. This simplified case is very useful for performing several 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 biogeophysical processes. The parameters selected have already been optimised in previous data assimilation studies using atmospheric CO_2 concentrations (Peylin et al., 2016; Bacour et al., 2023). In addition to Vcmax, we choose the following:

- the PFT-dependent parameter specific leaf area (SLA), which 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 $m_{\text{maint.resp}}$, which defines the slope of the maintenance respiration coefficient that controls autotrophic respiration;
- the PFT-dependent parameter LAI_{max}, which controls the maximum leaf area index for carbon allocation (once the LAI reaches LAI_{max}, no carbon is allocated to the leaf, which impacts the vegetation biomass and therefore acts on both photosynthesis and respiration).

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

2.3.5 Error covariance matrices

To implement the two data assimilation methods, ϵ-VarDA and EnVarDA, we define two error covariance matrices: **R** and **B**. These matrices are configured to be diagonal, as we assimilate synthetic observations, and are 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 CO_2 observations.

R matrix

The **R** matrix represents the model structural and observation errors. In our 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) and

in the transport model (i.e. coarse spatial resolution and wind biases) and measurement errors are discarded. Indeed, since the synthetic observations are generated by a simulation, as detailed in Sect. 2.3.2, there exists at least one solution where all observations can be matched perfectly. For this reason, we use a simplified **R** matrix 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 **B** matrix represents the background errors associated with prior knowledge of the parameters. We set the error to 30 % of the parameter range for the simple case and 20 % for the complex case (as we use larger parameter ranges for this case). The background errors of each parameter can be seen in Fig. 3 for the simple case and in Fig. 6 and Table A2 for the complex case.

2.3.6 Tuning ϵ for gradient calculation

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

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

where ϵ defines $\Delta\boldsymbol{x}$ as explained in Sect. 2.2.2. By changing ϵ we change $\Delta\boldsymbol{x}$, and we can attempt to find the value of ϵ for which the derivative becomes stable. Figure 2 shows the sensitivity of ϵ, ranging from 10^{-8} to 10^{-2}, on the calculated partial derivative of each Vcmax. We see that the partial derivative of Vcmax is unstable with an ϵ below 10^{-3} for all PFTs. Therefore, we need a value of ϵ greater than 10^{-3} to ensure correct gradient calculation with respect to the Vcmax parameter. Table A3 shows the values of the mean of the partial derivatives for all parameters and PFTs using an ϵ that allows for a stable derivative. This also allows us to check the consistency of the derivation calculation. For example, the increase in Vcmax leads to an increase in the photosynthetic capacity and subsequently in the carbon uptake by vegetation. This leads to a reduction in atmospheric CO_2 concentrations. We can see in Table A3 that the values obtained for Vcmax are negative, which is the expected response. The same ϵ test was carried out for the other four parameters used in the complex case, and the results are shown in Fig. A2 and Table A3:

- The partial derivative of SLA diverges with an ϵ below 10^{-3} for all PFTs. SLA has the same impact that Vcmax has on atmospheric CO_2 concentrations, so the negative mean values obtained are expected.
- The partial derivative of Q_{10} does not diverge for any values of ϵ. As expected, the mean value of its derivation is negative. Increasing Q_{10} 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 $m_{\text{maint.resp}}$ diverges with different ϵ values depending on the PFT, ranging from 10^{-5} for PFT TrBE to 10^{-2} for PFT CropsC_4. This may be due to different distributions and proportions of PFTs (see Table A1). However, the mean values at 10^{-2} are all positive. $m_{\text{maint.resp}}$ has an impact on autotrophic respiration; increasing this parameter increases vegetation respiration and therefore the atmospheric CO_2 concentration.
- The partial derivative of LAI_{max} diverges with an ϵ below 10^{-2} for all PFTs. Determining the sign of the mean values of the partial derivative of this parameter is not trivial here. LAI_{max} influences vegetation biomass and therefore photosynthesis and respiration. All PFTs give negative mean values for their partial derivative; only the PFT TrBR gives a positive mean value.

2.3.7 Defining the impact of the configuration

For both methods, ϵ-VarDA and EnVarDA, the configuration used plays an important role in the quality of the minimisation of the associated cost function and thus the calibration of the parameter. Whether it is the choice of ϵ in ϵ-VarDA or the number of members used to generate the ensemble in EnVarDA, it is up to the user to make a choice, which can only be subjective. To assess the impact of these choices, we launch the twin experiment using different configurations:

- For the simple case, we use five different values of ϵ for ϵ-VarDA based on the sensitivity test presented in Sect. 2.3.6 and five different ensemble sizes in EnVarDA.
- For the complex case, we use five different ensemble sizes in EnVarDA and one different value of ϵ in ϵ-VarDA.

For the complex case using ϵ-VarDA, ϵ is selected relative to the results in the simple case and Fig. A2. Re-tuning ϵ for each parameter requires too many simulations and is therefore not feasible for the complex case.

For each minimisation, the limited memory Broyden–Fletcher–Goldfarb–Shanno algorithm with bound constraints (L-BFGS-B) is used (Byrd et al., 1995). For ϵ-VarDA, we set the 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 to the solution obtained). For EnVarDA, no maximum iteration limit

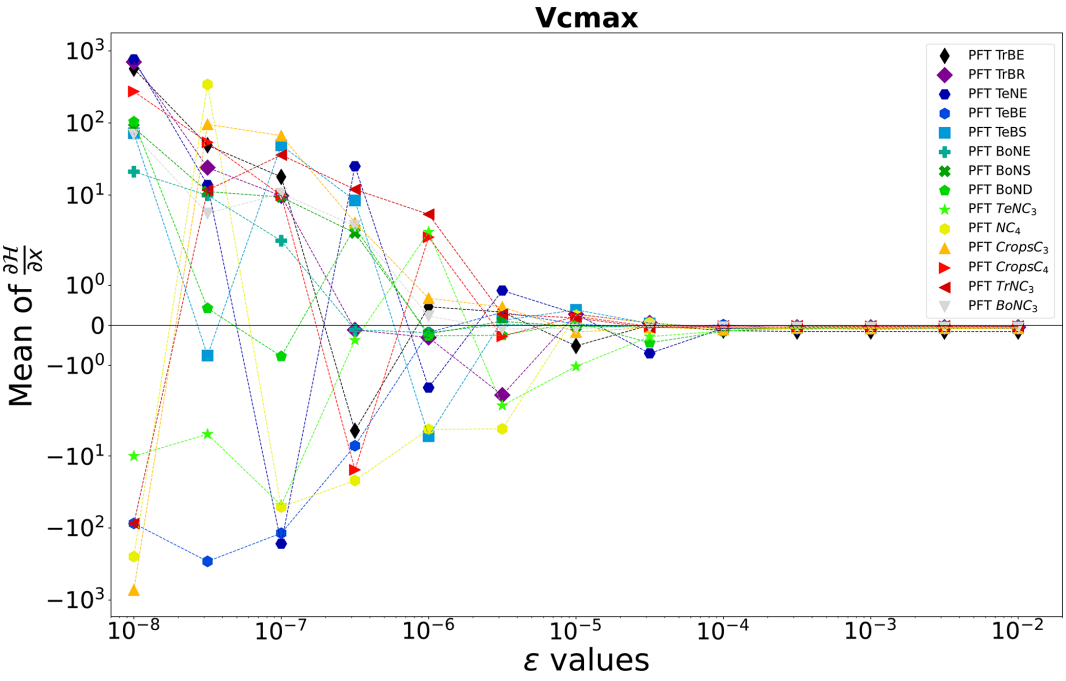

Figure 2. ϵ test: spatial and temporal average of the partial derivative of $\mathcal{H}$ as a function of ϵ. The partial derivative of $\mathcal{H}$ is calculated with respect to the parameter Vcmax for each PFT. It is calculated on the concentration space using every station over 2 years. The mean 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 ϵ.

is chosen, since an iteration does not require further simulation of the model (all required information is contained 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

The results in terms of (1) the mean reduction in the root mean square difference (RMSD) calculated between the pseudo-observation and the simulation over the 2 years of the assimilation window for the 21 atmospheric stations, (2) the mean absolute differences (MAD) in parameter space, and (3) the computational demand of each experiment using the simple case are summarised in Table 1. We see that for the ϵ-VarDA method, the best results are obtained with an ϵ 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 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 Sect. 3.2.1.

For the complex case, results are presented in Table 2. We see that the best results for the EnVarDA method are obtained with an ensemble of 300 members, giving a mean RMSD

Table 1. Mean RMSD reduction score between the synthetic observations and posterior simulations of the atmospheric CO_2 concentration at 21 atmospheric stations, the mean absolute difference (MAD) score computed between the true parameter values used to generate the synthetic observations and the posterior parameters, and the number of simulations used for each configuration of ϵ-VarDA and EnVarDA for the simple case.

ϵ-VarDA	ϵ	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
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

Table 2. Mean RMSD reduction score between synthetic observations and posterior simulations of the atmospheric CO_2 concentration at 21 atmospheric stations for each configuration of the EnVarDA method for the complex case.

EnVarDA	Ensemble	Mean RMSD reduction
	100	81 %
	200	89.8 %
	300	94.4 %
	350	94.0 %
	400	90.9 %

reduction of 94.4 %. This configuration is considered for the complex case in Sect. 3.2.2 using the EnVarDA method.

3.2 Comparing ϵ-VarDA and EnVarDA

3.2.1 Simple case

Figures 3 and 4 compare the results obtained for the ϵ-VarDA and EnVarDA methods using the configurations chosen in Sect. 3.1. Figure 3 shows that the parameter values obtained by the EnVarDA method are 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 EnVarDA method is able to almost recover the true parameters. The parameter values obtained by the ϵ-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 Vcmax values of PFTs $TeNC_3$, TeNE, and TrBE are close to the true value of the parameters, whereas PFTs $BoNC_3$, BoNE, TeBS, TeBE, and TrBR give values that are between the prior and the true value; the Vcmax values of other PFTs have either maintained or increased the distance between the prior and the true values. This shows that the ϵ-VarDA method falls into a local minimum and is therefore unable to recover the true parameters. Figure 4 shows the different RMSD scores between the synthetic observations and prior/posterior simulations for each of the 21 atmospheric stations. The average reduction in RMSD for the ϵ-VarDA method is 82 %, with a mean RMSD of 0.1 ppm. The largest reduction in RMSD is for the German station, Schauinsland (SCH) (87 %), and the lowest is for the Australian station Cape Grim (CGO) (49.8 %). Comparatively, the average reduction in RMSD for the EnVarDA method reaches 97 %, with a mean RMSD of 0.01 ppm across all stations. The highest reduction in RMSD is for the Chinese station, Waliguan (WLG) (99 %), and the lowest is for the Australian station Cape Cleveland (CFA) (92.7 %). We see that the EnVarDA method outperforms the ϵ-VarDA method: the EnVarDA method has the best fit to the assimilated synthetic observations 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 EnVarDA method using an ensemble of 300 members and the ϵ-VarDA method using an ϵ of 5×10^{-2} for all parameters. We stop the ϵ-VarDA method after 25 iterations, which already represents 1450 model simulations, as it shows no significant improvement in the minimisation of its cost function. We find that EnVarDA gives a mean reduction in RMSD of 94.3 % across all stations, with a maximum reduction in RMSD at the South African station, Cape Point (CPT) (98.8 %), and a minimum RMSD reduction at the Finland station, Pallas (PAL) (85 %). The ϵ-VarDA method gives a mean reduction in RMSD of 92.5 % across all stations, with a maximum reduction in 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 to 0.17 ppm after assimilation for EnVarDA and to 0.24 ppm for ϵ-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 EnVarDA and ϵ-VarDA, with a p value of 0.046. This confirms that the average posterior RMSD obtained by EnVarDA is significantly lower than the posterior RMSD obtained by ϵ-VarDA, with a confidence level of 95 %. We computed the mean squared difference (MSD) between the synthetic observations concatenated across all stations, the prior simulation, and 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 Appendix A. The prior MSD is 11.49 ppm^2 and is reduced to 0.04 ppm^2 using the EnVarDA method and to 0.08 ppm^2 using the ϵ-VarDA method. The decomposition of the prior MSD indicates a squared bias of 4.96 ppm^2 and an error variance equal to 6.53 ppm^2. The same decomposition for the posterior simulations yields a squared bias of 0.006 ppm^2 and an error variance equal to 0.03 ppm^2 for the EnVarDA method and a squared bias of 0.002 ppm^2 and an error variance equal to 0.07 ppm^2 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 EnVarDA, a 53 % reduction in this score is obtained, giving a normalised MAD score of 0.08. The ϵ-VarDA method gives a reduction in the normalised MAD of 15 %, giving a normalised MAD score of 0.14. Figure 6 shows the prior, true, and posterior parameter values obtained using both methods. For each parameter, we calculate the MAD score independently. The EnVarDA method gives a MAD reduction of 44.7 % for Vcmax, 78.2 % for

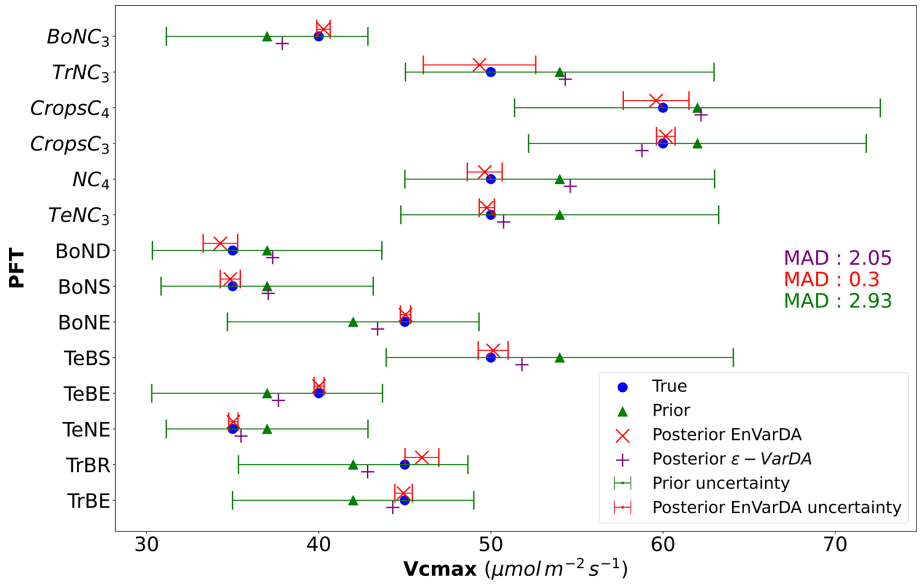

Figure 3. Results in parameter space for the simple case: the prior parameter values are represented by the green triangles, and the posterior parameter values after optimisation are represented by the purple + symbol for the ϵ-VarDA method and the red × symbol for the EnVarDA method. The blue circles 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 EnVarDA. The red error bar represents the standard deviation of the posterior ensemble obtained by 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 colour code (green score using the prior parameter, purple score using the posterior parameter of ϵ-VarDA, and red score using the posterior parameter of EnVarDA).

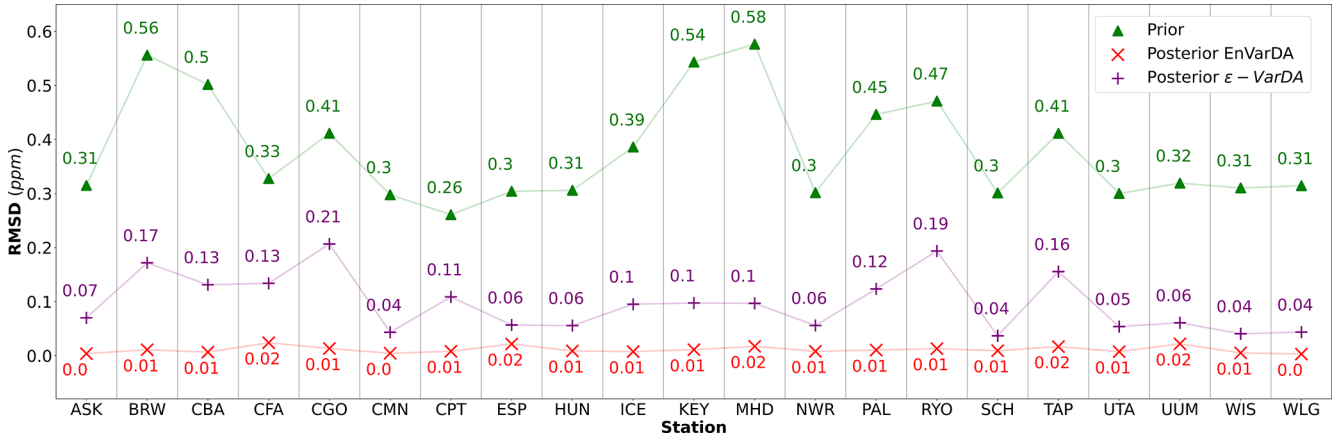

Figure 4. The root mean squared difference (RMSD) scores between synthetic observations and prior simulations for the simple case for each of the 21 atmospheric stations are represented by green triangles. The RMSD scores between synthetic observations and posterior simulations given by the ϵ-VarDA (EnVarDA) method are represented by the purple + symbol (red × symbol).

SLA, 36.3 % for LAI_{max}, and 54.2 % for $m_{maint.resp}$ and a reduction in the absolute difference (AD) of 98.8 % for Q_{10}. The ϵ-VarDA method gives a MAD reduction of 11.3 % for Vcmax, 32.7 % for SLA, 9.6 % for LAI_{max}, and 4.2 % for $m_{maint.resp}$ and a reduction in the AD of 97.5 % for Q_{10}.

Figure 7 illustrates the spatial disparities in net land carbon fluxes between the synthetic fluxes and the prior/posterior estimation of the two methodologies, in addition to their mean annual global net carbon flux. The EnVarDA method achieved a mean annual global net flux of $-2.62\,\mathrm{Gt\,C\,yr^{-1}}$, with a difference of $0.05\,\mathrm{Gt\,C\,yr^{-1}}$ compared to the synthetic fluxes. Spatial differences were limited to an absolute maximum of $0.28\,\mathrm{g\,C\,m^{-2}\,d^{-1}}$, with an absolute mean of $0.009\,\mathrm{g\,C\,m^{-2}\,d^{-1}}$. In contrast, the ϵ-VarDA method produced a mean annual global net flux of $-2.43\,\mathrm{Gt\,C\,yr^{-1}}$, with a difference of $0.24\,\mathrm{Gt\,C\,yr^{-1}}$ relative to the synthetic

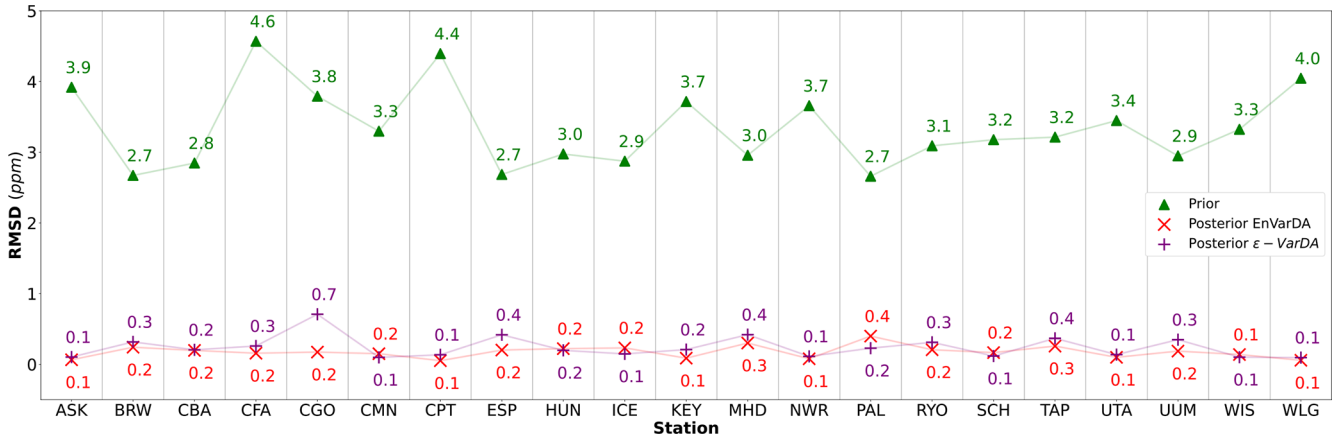

Figure 5. The root mean squared difference (RMSD) scores between synthetic observations and prior simulations for each of the 21 atmospheric stations for the complex case are represented by green triangles. The RMSD scores between synthetic observations and posterior simulations given by the ϵ-VarDA (EnVarDA) method are represented by the purple + symbol (red × symbol).

fluxes. Spatial differences for this method reached an absolute maximum of $0.6\,\mathrm{g\,C\,m^{-2}\,d^{-1}}$, with an absolute mean of $0.031\,\mathrm{g\,C\,m^{-2}\,d^{-1}}$. 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 EnVarDA 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 ϵ-VarDA method has correlation coefficients of 0.98 in time and 0.84 in space.

4 Discussion

4.1 Experiments

In Sect. 3.2.1, we found that the EnVarDA method outperforms the ϵ-VarDA method in terms of both RMSD reduction and MAD score, while requiring a smaller number of model simulations for the simple case. The EnVarDA method reduces the RMSD by 97 % and almost recovers the true parameters, whereas the ϵ-VarDA method reduces the RMSD by 82 % and seems to converge into a local minimum. In addition, EnVarDA requires 3 times fewer simulations. The other configurations presented in Sect. 3.1 show that EnVarDA using 50 members leads to similar RMSD reduction to ϵ-VarDA (see Table 1). However, this EnVarDA configuration still gives a better MAD score of 0.44, giving a reduction in the prior MAD by 85 % – this shows that the EnVarDA method is less influenced by local minima than the ϵ-VarDA method. We can also note that using the ϵ-VarDA method results in a posterior parameter values that either (i) remain close to the a priori values or (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 Vcmax 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 PFTs TeBE, BoNS, BoND, $\mathrm{CropsC_4}$, and $\mathrm{TrNC_3}$, which have a proportion equal to or less than 3% and are therefore less influential in 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 $\mathrm{NC_4}$. The 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 $\mathrm{TrNC_3}$ and $\mathrm{CropsC_4}$ PFTs are the most uncertain.

In Sect. 3.2.2, we saw that the EnVarDA method is able to calibrate 57 parameters and reduces the mean RMSD by 94.3 %, which is slightly better than the ϵ-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 and 0.24 ppm for the complex case and 0.01 and 0.1 ppm for the simple case for EnVarDA and ϵ-VarDA, respectively). Furthermore, the MSD score is better for the EnVarDA method ($0.04\,\mathrm{ppm^2}$ using the EnVarDA method and $0.08\,\mathrm{ppm^2}$ 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 ϵ-VarDA method. However, squared bias values below $0.01\,\mathrm{ppm^2}$ are negligible. While the mean RSMD re-

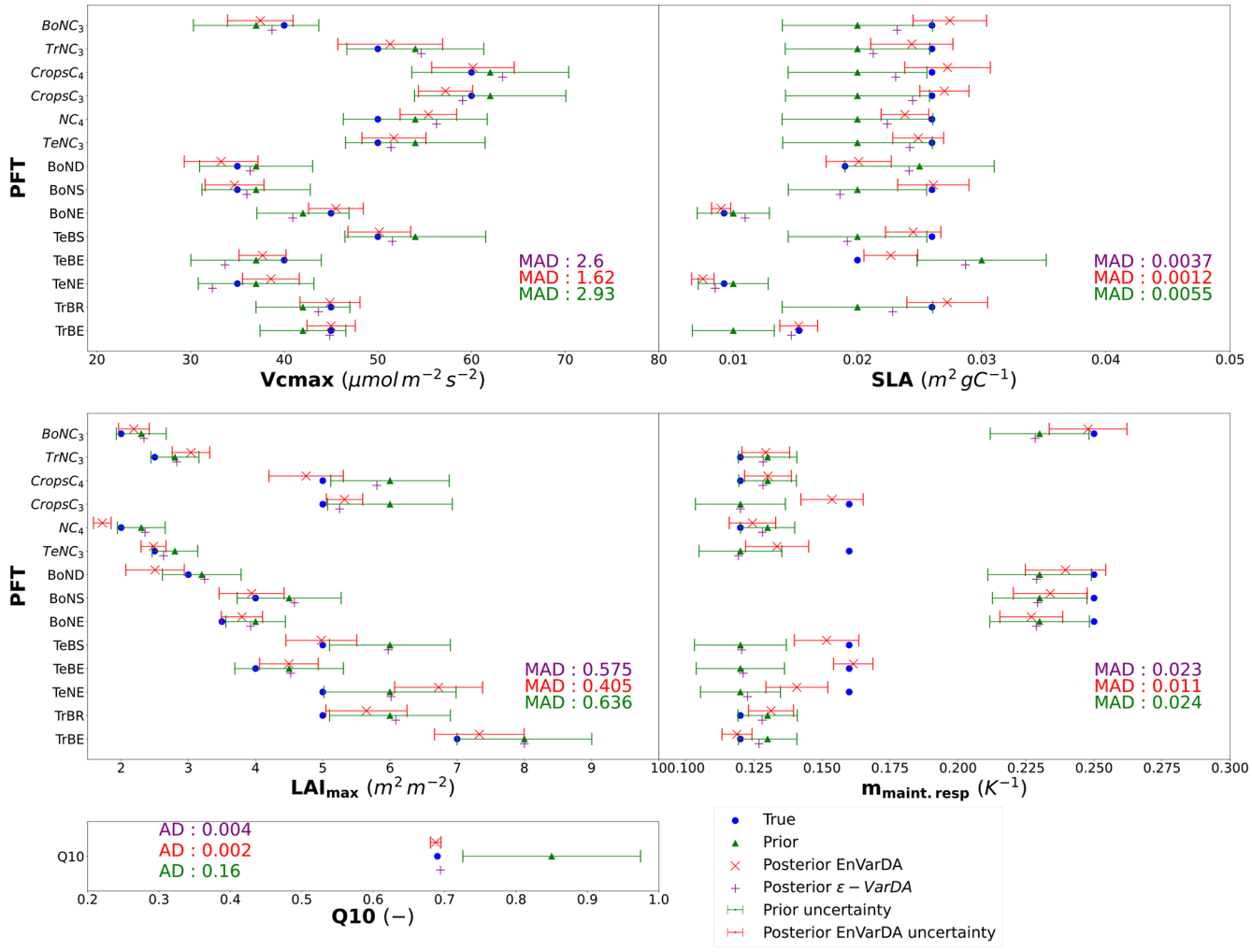

Figure 6. Results in parameter space for the complex case: the prior parameter values are represented by the green triangles, and the posterior parameter values after optimisation are represented by the purple + symbol for the ϵ-VarDA method and the red × symbol for the EnVarDA method. The blue circles 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 EnVarDA. The red error bar represents the standard deviation of the posterior ensemble obtained by EnVarDA, which can be interpreted as the posterior uncertainty. The MAD (or the absolute differences for Q_{10}) scores shown are calculated for each parameter independently.

duction and MSD scores are similar for the complex case, the MAD scores in parameter space are different. In fact, the EnVarDA method is closer to the true parameters by reducing the normalised MAD by 53 %, whereas the ϵ-VarDA method remains very close to the a priori position. The posterior ensemble generated for 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 Q_{10} parameter (reducing the standard deviation by 94 %) and the lowest for the less sensitive $m_{\text{maint.resp}}$ parameter (with a 14 % reduction for the NC_4 PFT). Both methods are capable of recovering the true Q_{10} parameter since it is the most sensitive parameter. The ϵ-VarDA method seems to have difficulties in calibrating the parameters $m_{\text{maint.resp}}$ and LAI_{max}, showing reductions in MAD scores that are less than 10 %. Considering Fig. A2, we can see that some PFTs give partial derivatives that do not completely converge, with an ϵ of 10^{-2} (for example, the PFT CropsC_4), so it is likely that ϵ for these parameters is underestimated. Other PFTs seem to give partial derivatives that do completely converge, with an ϵ 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 the MAD score, which is always better for the EnVarDA method. The self-compensating effect is illustrated in Fig. 7. The posterior spatial distributions of net carbon flux obtained from the two methods exhibit notable differences. It appears that the ϵ-VarDA method obtains a different spatial

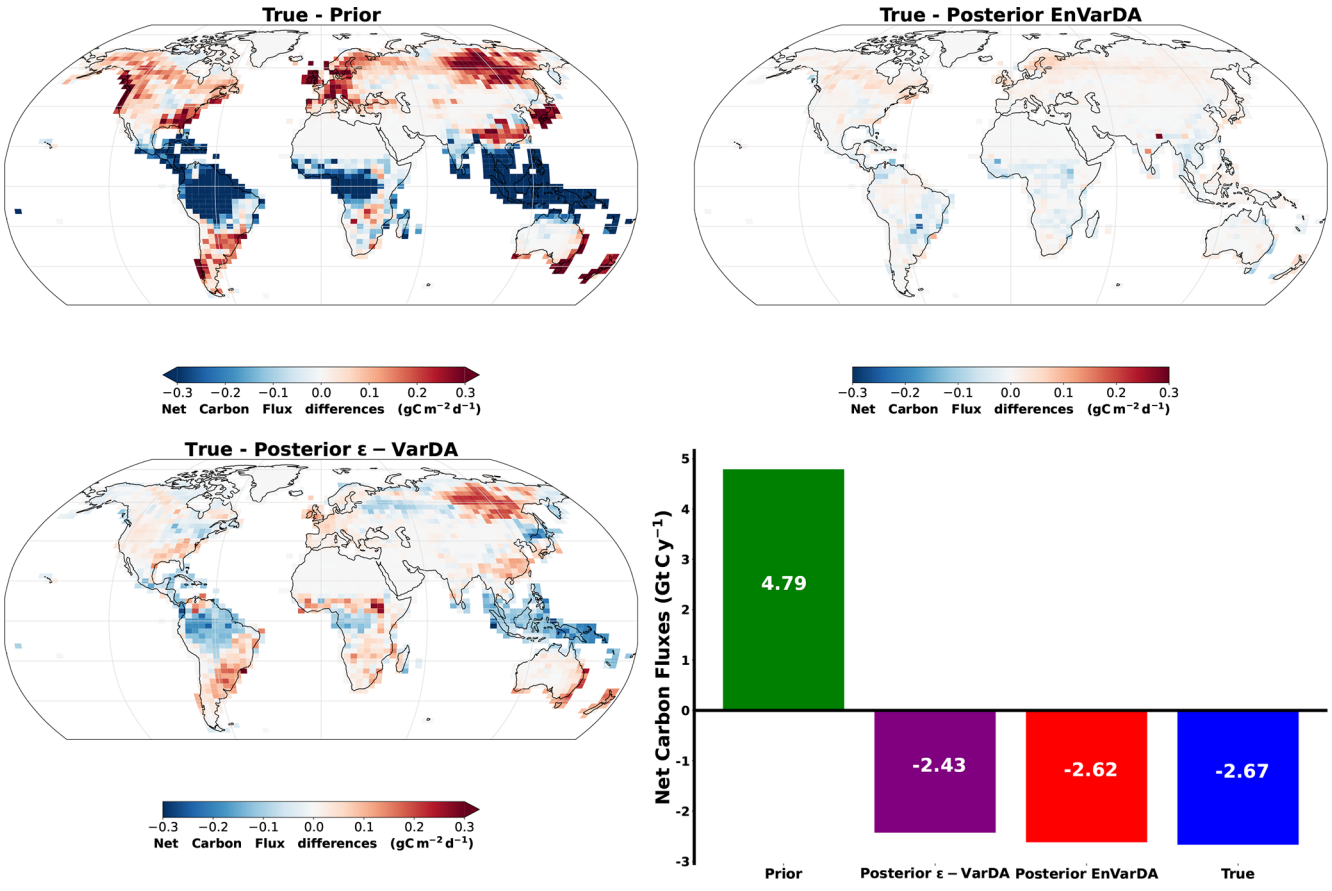

Figure 7. Spatial differences in net land carbon fluxes between synthetic fluxes and the prior/posterior estimate of the two methods alongside their mean annual global net carbon flux for the complex case. (Negative values are carbon uptakes, and positive values are carbon emissions.)

structure of the net carbon fluxes. Indeed, the carbon fluxes absent from one region can be reallocated to another, resulting in only minor variations in atmospheric CO_2 concentrations. We believe that the different spatial structure obtained by ϵ-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 EnVarDA method demonstrates superior performance, as it is more aligned with the synthetic net carbon fluxes both spatially and globally compared to the ϵ-VarDA method. Figure A3 shows the differences in the 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 EnVarDA method is slightly better than that of the ϵ-VarDA method for the global budget and better matches the spatial distribution of the synthetic flux. The ϵ-VarDA method appears to compensate for the lack of change between the prior and posterior GPP across most of the Northern Hemisphere. However, the EnVarDA method also outperforms the ϵ-VarDA method in terms of computational cost: the EnVarDA method only needs 300 simulations, whereas the ϵ-VarDA method needs 1450, which means a reduction in computing time of almost 5 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 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 EnVarDA here therefore demonstrates its ability to calibrate many parameters with fewer model simulations. In this experiment, one simulation took 11 min on average (wall time), using 20 CPUs of a computer server (with an Intel Xeon Gold 5115 processor). Neglecting the other computational times, using ϵ-VarDA for the complex case represents more than 265 h of computation compared with only 55 h for EnVarDA. Such a reduction cannot be ignored, especially since a simulation in this experiment represents a short (only 2 years), low-cost model configuration – low

ORCHIDEE spatial resolution and the 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 ϵ-VarDA method did not obtain the correct spatial distributions of NBP and GPP (see Figs. 7 and A3). This is not the case for the 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 ϵ-VarDA method. We also believe that the ensemble nature of the 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 ϵ-VarDA method is likely related to inaccurate determination of ϵ, which results in inaccurate estimates of the tangent linear and adjoint models. The 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. However, the performance of the EnVarDA method seems dependent on the generated ensemble. As shown in Table 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 minimisation 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 EnVarDA setup, i.e. the size and distribution of the ensemble, is less critical than the subjective choice of ϵ used in ϵ-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 ϵ 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 ϵ values (Santaren et al., 2007; Kuppel et al., 2012; Peylin et al., 2016; Bacour et al., 2023).

The results obtained here for ϵ-VarDA are not equivalent to a standard VarDA method using tangent linear and adjoint models. Therefore, we can draw no conclusions on the comparison between the EnVarDA method and a standard VarDA method, as highlighted in Liu et al. (2008). A potential – but hard to implement – way to improve ϵ-VarDA may be to have a dynamic ϵ that becomes more refined as the methods converge. Nevertheless, even with the "perfect" ϵ, we cannot guarantee that the ϵ-VarDA method would be less computationally expensive. The assimilation of atmospheric CO_2 concentration data using VarDA has been implemented with a tangent linear model, as in Castro-Morales et al. (2019), and an adjoint model, as in Scholze et al. (2007). In these cases, the tangent linear and adjoint models were developed alongside the forward model. However, the ϵ-VarDA method was used in experiments where obtaining the tangent linear or adjoint model proved too difficult, such as in Peylin et al. (2016) and Bacour et al. (2023). Although ϵ-VarDA is not equivalent to standard VarDA, a comparison of EnVarDA with ϵ-VarDA demonstrates the strong performance of 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 allow us to focus on the performance of two assimilation methods. This experiment highlights the superiority of the EnVarDA method in assimilating atmospheric CO_2 concentration data. However, the assimilation of real observations is not 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 the 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 EnVarDA method can give infeasible posterior parameter values, i.e. outside the imposed parameter boundaries, and therefore give non-physical parameter values. Furthermore, even with feasible 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 a solution. Some of these challenges are not specific to the EnVarDA method and are common to the 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 Sect. 4.1 for several reasons, such as (i) incorrect initial conditions of the carbon pools, which can impact respiration;

(ii) wrong estimates of other flux components, such as ocean or fossil fuel components; and (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 optimisation to calibrate both land and ocean components, as is done in atmospheric inversion. The advantage of the 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 is 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_2 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_2 concentration data – e.g. satellite XCO_2 products – using 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 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 $\mathbf{R}$ 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 EnVarDA is therefore an excellent opportunity, as it can be implemented quickly and requires no model modification.

Moreover, the EnVarDA method has been used to assimilate several types of data using either a simple carbon model (Douglas et al., 2025) or more complex LSMs such 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 have shown that the EnVarDA method has good potential for calibrating ORCHIDEE parameters by assimilating atmospheric CO_2 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 EnVarDA effectively recovered the true parameter values, whereas the $\boldsymbol{\epsilon}$-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 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, EnVarDA is computationally less demanding and does not require the development or maintenance of tangent linear and adjoint models, 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.

Appendix A: Metric calculations

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

$$\mathrm{RMSD} = \sqrt{\frac{1}{N_t}\sum_{t=0}^{N_t}\left(\boldsymbol{\mathcal{H}}(\boldsymbol{x}_*)_t - \boldsymbol{y}_t\right)^2}, \tag{A1}$$

$$\mathrm{MAD} = \frac{1}{n_{\mathrm{param}}}\sum_{i=0}^{n_{\mathrm{param}}} |\boldsymbol{x}_{*i} - \boldsymbol{x}_{\mathbf{true}i}|, \tag{A2}$$

where $\boldsymbol{x}_*$ can be either $\boldsymbol{x}_b$ or $\boldsymbol{x}_a$. 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):

$$e = \mathcal{H}(x) - y, \tag{A3}$$

$$\mathrm{MSD}(e) = \mathbb{E}[e], \tag{A4}$$

$$\mathrm{MSD}(e) = \left(\mathbb{E}\left[e^2\right] - \mathbb{E}[e]^2\right) + \mathbb{E}[e]^2, \tag{A5}$$

$$\mathrm{MSD}(e) = \mathrm{Var}(e) + \mathbb{E}[e]^2, \tag{A6}$$

$$\mathrm{MSD}(e) = \mathrm{Var}(e) + \mathrm{Bias}(e)^2. \tag{A7}$$

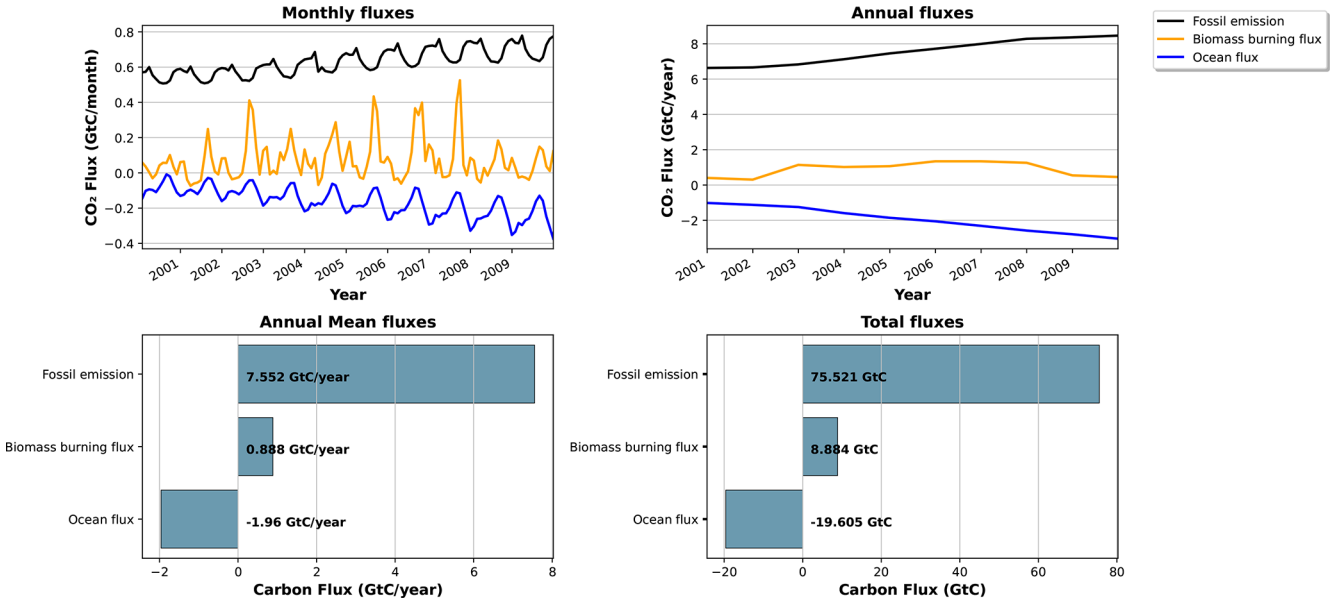

Figure A1. Ocean fluxes, biomass burning carbon fluxes, and fossil emissions (2000–2009).

Table A1. Plant functional types (PFTs) in ORCHIDEE with their abbreviations used in this study and their respective global cover fraction (with the remaining portion being bar soil).

PFT	Abbreviation	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 (C_3)	$TeNC_3$	6.43 %
Natural grassland (C_4)	NC_4	8.73 %
Crops (C_3)	$CropsC_3$	11.4 %
Crops (C_4)	$CropsC_4$	3.04 %
Tropical natural grassland (C_3)	$TrNC_3$	1.58 %
Boreal natural grassland (C_3)	$BoNC_3$	10.6 %

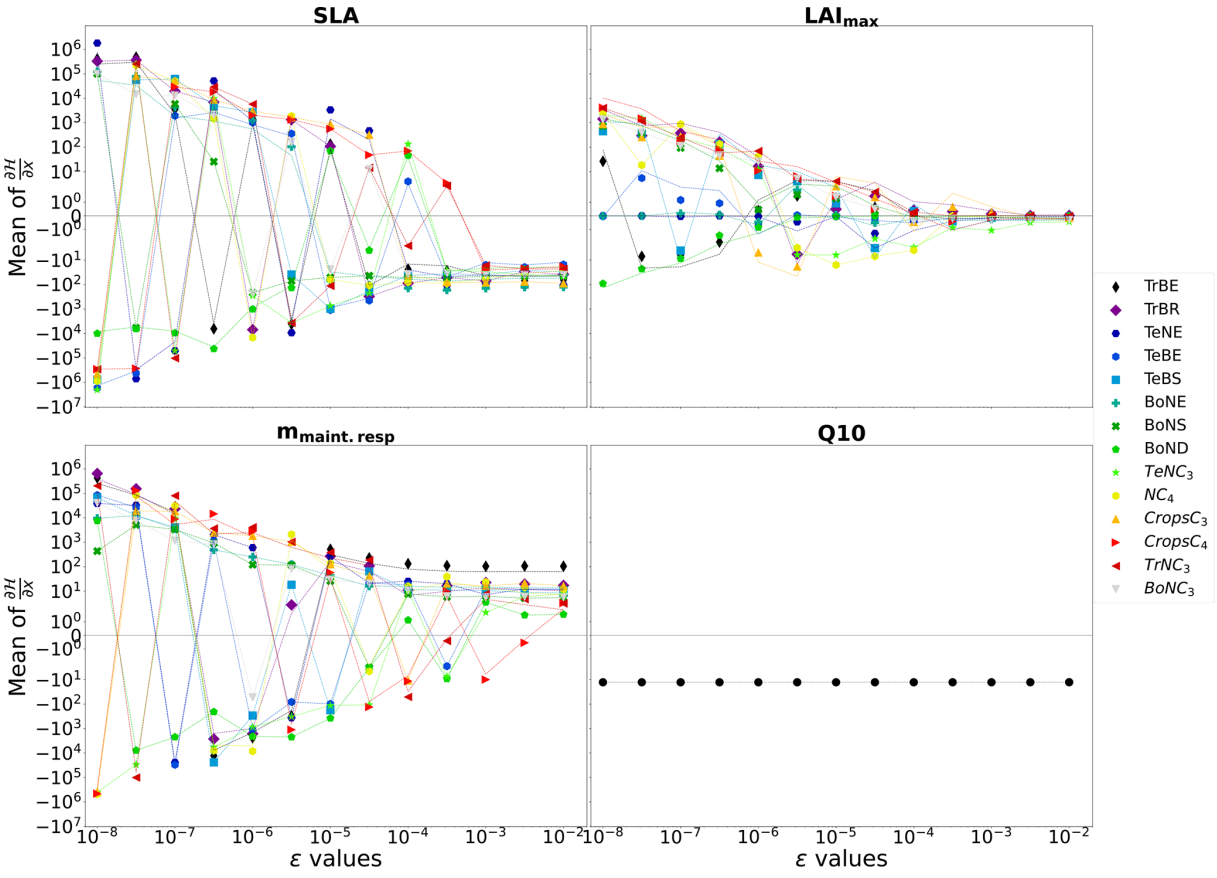

Figure A2. ϵ test: spatial and temporal average of the partial derivative as a function of ϵ. The partial derivative of the $\mathcal{H}$ model is calculated with respect to the parameters used in the complex case for each PFT. It is calculated on the concentration space using every station over 2 years. 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 ϵ.

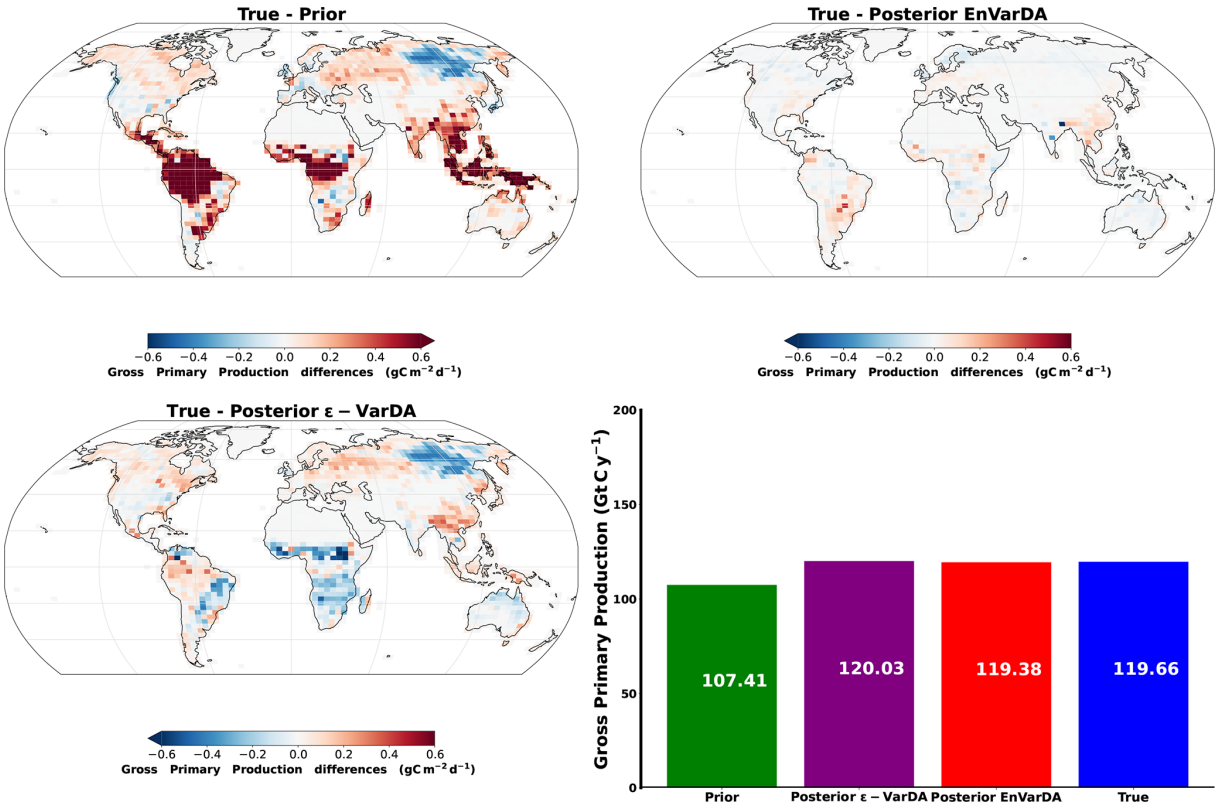

Figure A3. Spatial differences in gross primary production (GPP) fluxes between synthetic fluxes and the prior/posterior estimate of the two methods alongside their mean annual global GPP for the complex case.

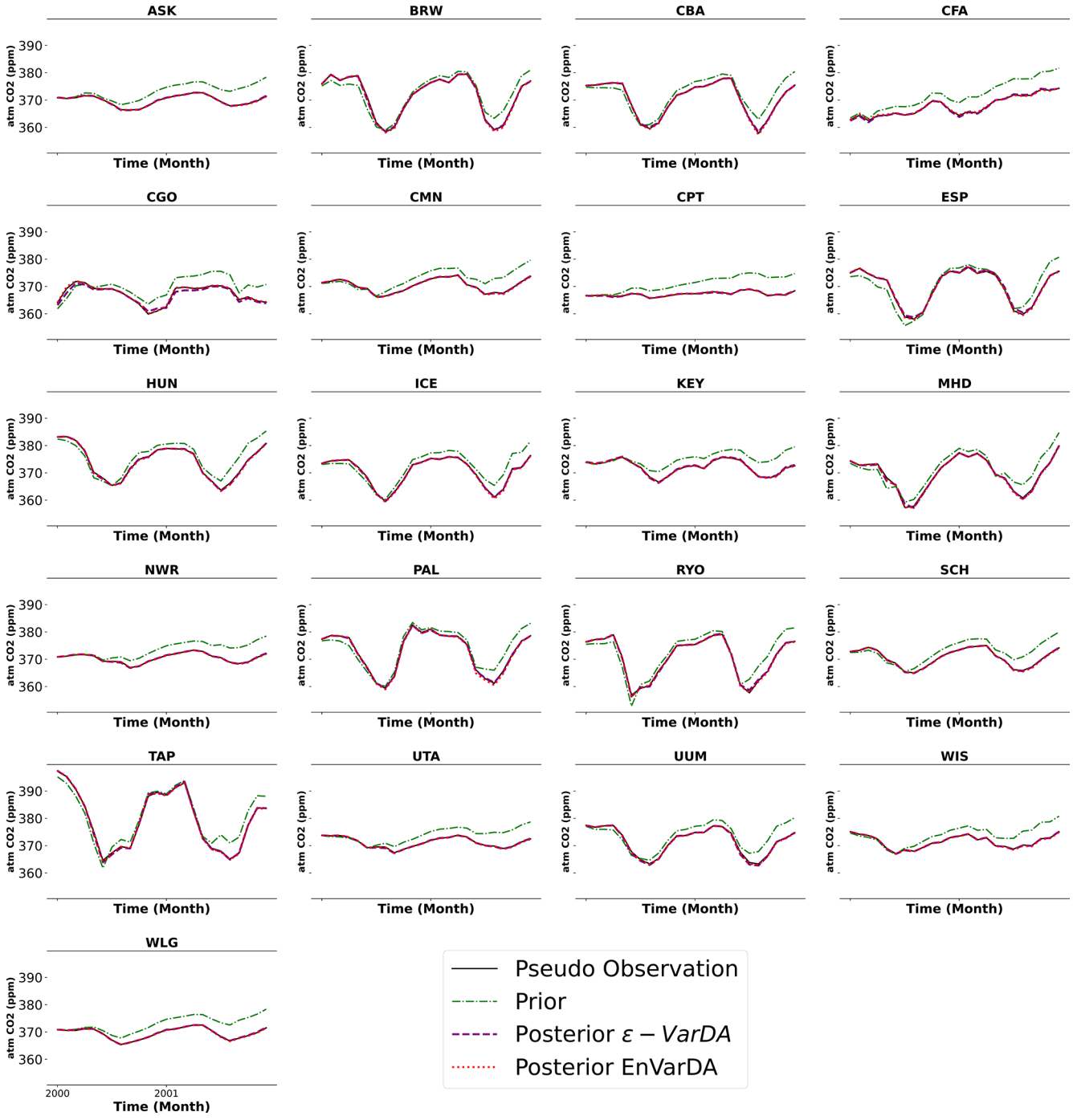

Figure A4. Time series of pseudo-observations (in black) of atmospheric CO_2 concentrations at 21 stations over the assimilated period, as well as the a priori (in green) and a posteriori time series from ϵ-VarDA (in purple) and EnVarDA (in red) for the complex case.

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

Parameters	Vcmax			SLA			LAI_{max}			$m_{maint.resp}$				Q_{10}		
	Prior	Errors	True	Prior	Errors	True	Prior	Errors	True	Prior	Errors	True		Prior	Errors	True
PFT													Global	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	00926	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				
$TeNC_3$	54	8	50	0.02	0.0074	0.026	2.8	0.4	2.5	0.12	0.02	0.16				
NC_4	54	8	50	0.02	0.0074	0.026	2.3	0.4	2	0.13	0.012	0.12				
$CropsC_3$	62	8.4	60	0.02	0.0074	0.026	6	1	5	0.12	0.02	0.16				
$CropsC_4$	62	8.4	60	0.02	0.0074	0.026	6	1	5	0.13	0.012	0.12				
$TrNC_3$	54	8	50	0.02	0.0074	0.026	2.8	0.4	2.5	0.13	0.012	0.12				
$BoNC_3$	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 parameters for each PFT computed using ϵ, allowing for 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	Vcmax	SLA	LAI_{max}	$m_{maint.resp}$		Q_{10}
ϵ	10^{-3}	10^{-3}	10^{-2}	10^{-2}		10^{-3}
PFT					Global	−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		
$TeNC_3$	−0.053	−44.9	−0.43	7.9		
NC_4	−0.072	−27.0	−0.22	12.0		
$CropsC_3$	−0.072	−89.3	−0.15	16.4		
$CropsC_4$	−0.023	−18.3	−0.07	3.3		
$TrNC_3$	−0.017	−23.0	−0.01	3.1		
$BoNC_3$	−0.056	−40.7	−0.28	5.7		

Code and data availability. The source code for the ORCHIDEE version used in this paper 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 this paper are available from a Zenodo repository at https://doi.org/10.5281/zenodo.14609416 (Beylat, 2025).

Author contributions. SB and PP conceived the study. VB implemented the coupling of ORCHIDEE and LMDZ, originally developed by PP. SB implemented the EnVarDA method, and VB implemented the ϵ-VarDA method. SB carried out and analysed the DA experiments. NR, PR, PP, and CB provided expertise on the assimilation of atmospheric CO_2 concentrations and the ϵ-VarDA method. ND and TQ provided expertise on the EnVarDA method. SB prepared the article with contributions from all co-authors.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibil-

ity lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The authors would like to thank the three reviewers for their constructive comments, Catherine Ottle for her feedback, and the LSCE IT cluster team.

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

Review statement. This paper was edited by Marko Scholze and reviewed by three anonymous referees.

References

Asch, M., Bocquet, M., and Nodet, M.: Data Assimilation: Methods, Algorithms, and Applications, vol. 28, https://doi.org/10.1137/1.9781611974546, 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, J. Geophys. Res.-Biogeo., 124, 3281–3306, 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, 1089–1111, 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_2, Tellus B, 58, 359–365, https://doi.org/10.1111/j.1600-0889.2006.00218.x, 2006.

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, Geosci. Model Dev., 15, 1913–1929, https://doi.org/10.5194/gmd-15-1913-2022, 2022.

Bannister, R. N.: A review of operational methods of variational and ensemble-variational data assimilation, Q. J. Roy. Metorol. Soc, 143, 607–633, https://doi.org/10.1002/qj.2982, 2016.

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), Geosci. Model Dev., 11, 4739–4754, 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_2 fluxes estimated from GOSAT retrievals of total column CO_2, Atmo. Chem. Phys., 13, 8695–8717, https://doi.org/10.5194/acp-13-8695-2013, 2013.

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., Groot Zwaaftink, C. D., Haussaire, J.-M., Henne, S., Houweling, S., Karstens, U., Kutsch, W. L., Luijkx, I. T., Monteil, G., Palmer, P. I., van Peet, J. C. A., Peters, W., Peylin, P., Potier, E., Rödenbeck, C., Saunois, M., Scholze, M., Tsuruta, A., and Zhao, Y.: The Community Inversion Framework v1.0: a unified system for atmospheric inversion studies, Geosci. Model Dev., 14, 5331–5354, https://doi.org/10.5194/gmd-14-5331-2021, 2021.

Beylat, S.: 4DEnVar ORCHIDEE: v1.0.0, Zenodo [code and data set], 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., Guez, Lionel, E., Guilyardi, E., Guimberteau, M., Hauglustaine, D., Hourdin, F., Idelkadi, A., Joussaume, S., Kageyama, M., Khodri, M., Krinner, G., Lebas, N., Levavasseur, G., Lévy, C., Li, L., Lott, 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 IPSL-CM6A-LR Climate Model, J. Adv. Model. Earth Syst., 12, e2019MS002010, https://doi.org/10.1029/2019MS002010, 2020.

Bousquet, P., Peylin, P., Ciais, P., Quéré, C. L., Friedlingstein, P., and Tans, P. P.: Regional Changes in Carbon Dioxide Fluxes of Land and Oceans Since 1980, Science, 290, 1342–1346, 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 J. Scient. Comput., 16, 1190–1208, https://doi.org/10.1137/0916069, 1995.

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, 3009–3032, 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 CO_2 sources and sinks from satellite observations: Method and application to TOVS data, J. Geophys. Res.-Atmos., 110, https://doi.org/10.1029/2005JD006390, 2005.

Chevallier, F., Bréon, F.-M., and Rayner, P. J.: Contribution of the Orbiting Carbon Observatory to the estimation of CO_2 sources and sinks: Theoretical study in a variational data assimilation framework, J. Geophys. Res.-Atmos., 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 CO_2 flux estimates from in situ and spaceborne measurements of atmospheric CO_2, Geophys. Res. Lett., 41, 1065–1070, https://doi.org/10.1002/2013GL058772, 2014.

Collatz, G., Ball, J., Grivet, C., and Berry, J. A.: Physiological and environmental regulation of stomatal conductance, photosynthesis and transpiration: a model that includes a laminar boundary layer, Agr. Forest Meteorol., 54, 107–136, https://doi.org/10.1016/0168-1923(91)90002-8, 1991.

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, Hydrol. Earth Syst. Sci., 25, 2445–2458, https://doi.org/10.5194/hess-25-2445-2021, 2021.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. Roy. Meteorol. Soc., 120, 1367–1387, https://doi.org/10.1002/qj.49712051912, 1994.

Couvreux, 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, J. Adv. Model. Earth Syst., 13, e2020MS002217, 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, Adv. Stat. Climatol. Meteorol. Oceanogr., 6, 223–244, https://doi.org/10.5194/ascmo-6-223-2020, 2020.

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. M., 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, L., 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, Q. J. Roy. Meteorol. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

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, J. Geophys. Res.-Atmos., 107, ACL 3-1–ACL 3-19, https://doi.org/10.1029/2001JD000634, 2002.

Douglas, N., Quaife, T., and Bannister, R.: Exploring a hybrid ensemble–variational data assimilation technique (4DEnVar) with a simple ecosystem carbon model, Environ. Model. Softw., 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., Fichefet, 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 IPSL-CM5 Earth System Model: from CMIP3 to CMIP5, Clim. Dynam., 40, 2123–2165, https://doi.org/10.1007/s00382-012-1636-1, 2013.

Enting, I. G.: Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press, 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, J. Geophys. Res.-Oceans, 99, 10143–10162, https://doi.org/10.1029/94JC00572, 1994.

Farquhar, G. D., von Caemmerer, S., and Berry, J. A.: A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species, Planta, 149, 78–90, https://doi.org/10.1007/BF00386231, 1980.

Fisher, R. A. and Koven, C. D.: Perspectives on the Future of Land Surface Models and the Challenges of Representing Complex Terrestrial Systems, J. Adv. Model. Earth Syst., 12, e2018MS001453, https://doi.org/10.1029/2018MS001453, 2020.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., 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., 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., Gürses, O., 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., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, 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 Syst. Sci. Data, 15, 5301–5369, 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 Comput., 4, 1–58, 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, 54–57, https://doi.org/10.1002/pamm.200310014, 2003.

Gilbert, J. C. and Lemaréchal, C.: Some numerical experiments with variable-storage quasi-newton algorithms, Math. Program., 45, 407–435, https://doi.org/10.1007/bf01589113, 1989.

GLOBALVIEW-CO2: GLOBALVIEW-CO2: Cooperative Atmospheric Data Integration Project–Carbon Dioxide, NASA's Earthdata portal [data set], https://doi.org/10.3334/ORNLDAAC/1111, 2013.

Gurney, K. R., Law, R. M., Denning, A. S., Rayner, P. J., Baker, D., Bousquet, P., Bruhwiler, L., Chen, Y.-H., Ciais, P., Fan, S., Fung, I. Y., Gloor, M., Heimann, M., Higuchi, K., John, J., Maki, T., Maksyutov, S., Masarie, K., Peylin, P., Prather, M., Pak, B. C., Randerson, J., Sarmiento, J., Taguchi, S., Takahashi, T., and Yuen, C.-W.: Towards robust regional estimates of CO_2 sources and sinks using atmospheric transport models, Nature, 415, 626–630, https://doi.org/10.1038/415626a, 2002.

Hodson, T. O., Over, T. M., and Foks, S. S.: Mean Squared Error, Deconstructed, J. Adv. Model. Earth Syst., 13, e2021MS002681, 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: Test of Various Formulations in a General Circulation Model, Mon. Weather Rev., 127, 822–837, 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, Q. J. Roy. Meteorol. Soc., 132, 567–583, 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, Q. J. Roy. Meteorol. Soc., 132, 585–603, 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, B. Am. Meteorol. Soc., 98, 589–602, 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, J. Adv. Model. Earth Syst., 13, e2020MS002225, 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, Sci. Adv., 9, eadf2758, https://doi.org/10.1126/sciadv.adf2758, 2023.

IPCC: Human Influence on the Climate System, Cambridge Universty Press, 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_2 in the 1980s, J. Geophys. Res.-Atmos., 104, 18555–18581, 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, J. Geophys. Res.-Atmos., 104, 18535–18553, 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 Biogeochem. Cy., 16, 14-1–14-16, 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_2 into a terrestrial vegetation model and interactive mission benefit analysis, Biogeosciences, 9, 3173–3184, https://doi.org/10.5194/bg-9-3173-2012, 2012.

Kennedy, M. C. and O'Hagan, A.: Bayesian calibration of computer models, J. Roy. Stat. Soc. Ser. B, 63, 425–464, 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, J. Adv. Model. Earth Syst., 16, e2023MS004163, https://doi.org/10.1029/2023MS004163, 2024.

Knorr, W. and Heimann, M.: Impact of drought stress and other factors on seasonal land biosphere CO_2 exchange studied through an atmospheric tracer transport model, Tellus B, 47, 471–489, 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., Smallman, T. L., Steele-Dunne, S. C., Vreugdenhil, M., Green, T., Zaehle, S., Aurela, M., Bouvet, A., Bueechi, E., Dorigo, W., El-Madany, T. S., Migliavacca, M., Honkanen, M., Kerr, Y. H., Kontu, A., Lemmetyinen, J., Lindqvist, H., Mialon, A., Miinalainen, T., Pique, G., Ojasalo, A., Quegan, S., Rayner, P. J., 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, Geosci. Model Dev., 18, 2137–2159, https://doi.org/10.5194/gmd-18-2137-2025, 2025.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C.: A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system, Global Biogeochem. Cy., 19, 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, 3757–3776, https://doi.org/10.5194/bg-9-3757-2012, 2012.

Kuppel, S., Chevallier, F., and Peylin, P.: Quantifying the model structural error in carbon cycle data assimilation systems, Geosci. Model Dev., 6, 45–55, 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, Environ. Model. Softw., 26, 1759–1763, https://doi.org/10.1016/j.envsoft.2011.05.016, 2011.

Liu, C., Xiao, Q., and Wang, B.: An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test, Mon. Weather Rev., 136, 3363–3373, https://doi.org/10.1175/2008MWR2312.1, 2008.

Liu, J., Baskaran, L., Bowman, K., Schimel, D., Bloom, A. A., Parazoo, N. C., Oda, T., Carroll, D., Menemenlis, D., Joiner, J., Commane, R., Daube, B., Gatti, L. V., McKain, K., Miller, J., Stephens, B. B., Sweeney, C., and Wofsy, S.: Carbon Monitoring System Flux Net Biosphere Exchange 2020 (CMS-Flux NBE 2020), Earth Syst. Sci. Data, 13, 299–330, https://doi.org/10.5194/essd-13-299-2021, 2021.

Locatelli, R., Bousquet, P., Hourdin, F., Saunois, M., Cozic, A., Couvreux, F., Grandpeix, J.-Y., Lefebvre, M.-P., Rio, C., Bergamaschi, P., Chambers, S. D., Karstens, U., Kazan, V., van der Laan, S., Meijer, H. A. J., 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, Geosci. Model Dev., 8, 129–150, https://doi.org/10.5194/gmd-8-129-2015, 2015.

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., Levavasseur, 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, J. Adv. Model. Earth Syst., 12, e2019MS001940, https://doi.org/10.1029/2019MS001940, 2020.

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, 7185–7208, 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, Geosci. Model Dev., 9, 3569–3588, 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 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 Biogeochem. Cy., 36, e2021GB007177, https://doi.org/10.1029/2021GB007177, 2022.

McNeall, D., Robertson, E., and Wiltshire, A.: Constraining the carbon cycle in JULES-ES-1.0, Geosci. Model Dev., 17, 1059–1089, https://doi.org/10.5194/gmd-17-1059-2024, 2024.

ORCHIDEE: ORCHIDEE V22 r7878 gmd 2025 4DEnVar, IPSL Data Catalog [code], https://doi.org/10.14768/c68bc728-da71-4383-84df-dcde31d9c006, 2025.

Peylin, P., Bousquet, P., Le Quéré, C., Sitch, S., Friedlingstein, P., McKinley, G., Gruber, N., Rayner, P., and Ciais, P.: Multiple constraints on regional CO_2 flux variations over land and oceans, Global Biogeochem. Cy., 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., van der Laan-Luijkx, I. T., and Zhang, X.: Global atmospheric carbon budget: results from an ensemble of atmospheric CO_2 inversions, Biogeosciences, 10, 6699–6720, 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, Geosci. Model Dev., 9, 3321–3346, 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, Geosci. Model Dev., 13, 55–69, https://doi.org/10.5194/gmd-13-55-2020, 2020.

Pinnington, E., Amezcua, 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, Hydrol. Earth Syst. Sci., 25, 1617–1641, 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, Geophys. J. Int., 167, 495–503, 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, Geosci. Model Dev., 8, 2315–2328, 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, NASA [data set], https://doi.org/10.3334/ORNLDAAC/1191, 2013.

Raoult, N., Beylat, S., Salter, J. M., Hourdin, F., Bastrikov, V., Ottle, C., and Peylin, P.: Exploring the potential of history matching for land surface model calibration, Geosci. Model Dev., 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, Geosci. Model Dev., 9, 2833–2852, https://doi.org/10.5194/gmd-9-2833-2016, 2016.

Rayner, P. J.: The current state of carbon-cycle data assimilation, Curr. Opin. Environ. Sustain., 2, 289–296, 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_2, $\delta^{13}C$ and O_2/N_2 observations, Tellus B, 51, 213–232, 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 Biogeochem. Cy., 19, https://doi.org/10.1029/2004GB002254, 2005.

Rayner, P. J., Michalak, A. M., and Chevallier, F.: Fundamentals of data assimilation applied to biogeochemistry, Atmos. Chem. Phys., 19, 13911–13932, 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_2 transport, Geosci. Model Dev., 11, 4489–4513, 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 Biogeochem. Cy., 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, 7137–7158, 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, J. Geophys. Res.-Atmos., 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 Zaehle, S.: Constraining a land-surface model with multiple observations by application of the MPI-Carbon Cycle Data Assimilation System V1.0, Geosci. Model Dev., 9, 2999–3026, 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 Zaehle, S.: Trends and Drivers of Terrestrial Sources and Sinks of Carbon Dioxide: An Overview of the TRENDY Project, Global Biogeochem. Cy., 38, e2024GB008102, https://doi.org/10.1029/2024GB008102, 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, Q. J. Roy. Meteorol. Soc., 113, 1311–1328, https://doi.org/10.1002/qj.49711347812, 1987.

Tarantola, A.: Inverse problem theory: methods for data fitting and model parameter estimation, SIAM, 1987.

Tarantola, A.: Inverse Problem Theory and Methods for Model Parameter Estimation, SIAM, https://doi.org/10.1137/1.9780898717921, 2005.

Tiedtke, M.: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models, Mon. Weather Rev., 117, 1779–1800, https://doi.org/10.1175/1520-0493(1989)117<1779:ACMFSF>2.0.CO;2, 1989.

Van Leer, B.: Towards the ultimate conservative difference scheme. IV. A new approach to numerical convection, J. Comput. Phys., 23, 276–299, https://doi.org/10.1016/0021-9991(77)90095-X, 1977.

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, Geosci. Model Dev., 9, 363–381, 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, Geosci. Model Dev., 14, 7659–7672, 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, Clim. Dynam., 41, 1703–1729, 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, Geosci. Model Dev., 10, 1789–1816, 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 J. Life Sci., 57, 27–38, 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 Biogeochem. Cy., 26, https://doi.org/10.1029/2011GB004185, 2012.

Zobler, L.: Global Soil Types, 1-Degree Grid (Zobler), NASA [data set], https://doi.org/10.3334/ORNLDAAC/418, 1999.