University of Reading

Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using Four Dimensional Variational data assimilation

Article

Accepted Version

Creative Commons: Attribution-Noncommercial-No Derivative Works 4.0

Pinnington, E. M., Casella, E., Dance, S. L. ORCID: https://orcid.org/0000-0003-1690-3338, Lawless, A. S. ORCID: https://orcid.org/0000-0002-3016-6568, Morison, J. I. L., Nichols, N. K. ORCID: https://orcid.org/0000-0003-1133-5220, Wilkinson, M. and Quaife, T. L. ORCID: https://orcid.org/0000-0001-6896-4613 (2016) Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using Four Dimensional Variational data assimilation. Agricultural and Forest Meteorology, 228-229. pp. 299-314. ISSN 0168-1923 doi: 10.1016/j.agrformet.2016.07.006 Available at https://centaur.reading.ac.uk/66177/

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/10.1016/j.agrformet.2016.07.006

University of Reading

Publisher: Elsevier

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

Investigating the role of prior and observation error correlations in improving a model forecast of forest carbon balance using four dimensional variational data assimilation

Ewan M. Pinnington[a,*], Eric Casella[c], Sarah L. Dance[a,b], Amos S. Lawless[a,b,d], James I. L. Morison[c], Nancy K. Nichols[a,b,d], Matthew Wilkinson[c], Tristan L. Quaife[a,d]

[a]*Department of Meteorology, University of Reading, Reading, UK*

[b]*Department of Mathematics and Statistics, University of Reading, Reading, UK*

[c]*Centre for Sustainable Forestry and Climate Change, Forest Research, Alice Holt, Farnham, UK*

[d]*National Centre for Earth Observation, University of Reading, Reading, UK*

Abstract

Efforts to implement variational data assimilation routines with functional ecology models and land surface models have been limited, with sequential and Markov chain Monte Carlo data assimilation methods being prevalent. When data assimilation has been used with models of carbon balance, prior or "background" errors (in the initial state and parameter values) and observation errors have largely been treated as independent and uncorrelated. Correlations between background errors have long been known to be a key aspect of data assimilation in numerical weather prediction. More recently, it has been shown that accounting for correlated observation errors in the assimilation algorithm can considerably improve data assimilation results and forecasts. In this paper we implement a Four-Dimensional Variational data assimilation (4D-Var) scheme with a simple model of forest carbon balance, for joint parameter and state estimation and assimilate daily observations of Net Ecosystem CO_2 Exchange (NEE) taken at the Alice Holt forest CO_2 flux site in Hampshire, UK. We then investigate the effect of specifying correlations between parameter and state variables in background error statistics and the effect of specifying correlations in time between observation errors. The idea of including these correlations in time is new and has not been previously explored in carbon balance model data assimilation. In data assimilation, background and observation error statistics are often described by the background error covariance matrix and the observation error covariance matrix. We outline novel methods for creating correlated versions of these matrices, using a set of previously postulated dynamical constraints to include correlations in the background error statistics and a Gaussian correlation function to include time correlations in the observation error statistics. The methods used in this paper will allow the inclusion of time correlations between many different observation types in the assimilation algorithm, meaning that previously neglected information can be accounted for. In our experiments we assimilate a single year of NEE observations and then run a forecast for the next 14 years. We compare the results using our new correlated background and observation error covariance matrices and those using diagonal covariance matrices. We find that using the new correlated matrices reduces the root mean square error in the 14 year forecast of daily NEE by 44% decreasing from 4.22 $\mathrm{gCm^{-2}day^{-1}}$ to 2.38 $\mathrm{gCm^{-2}day^{-1}}$.

*Corresponding author
Email address: e.m.pinnington@pgr.reading.ac.uk (Ewan M. Pinnington)

Keywords: Data assimilation, Carbon cycle, DALEC, Eddy covariance, Error covariance, Parameter estimation

1. Introduction

The land surface and oceans are responsible for removing around half of all human emitted carbon-dioxide from the atmosphere and therefore mediate the effect of anthropogenic induced climate change. Terrestrial ecosystem carbon uptake is the least understood process in the global carbon cycle (Ciais et al., 2014). It is therefore vital that we improve understanding of the carbon uptake of terrestrial ecosystems and their response to climate change in order to better constrain predictions of future carbon budgets. Observations of the Net Ecosystem Exchange (NEE) of CO_2 between terrestrial ecosystems and the atmosphere are now routinely made at flux tower sites world-wide, at sub-hourly resolution and covering multiple years (Baldocchi, 2008), providing a valuable resource for carbon balance model validation and data assimilation.

Data assimilation is the process of combining a mathematical model with observations in order to improve the estimate of the state of a system. Data assimilation has successfully been used in many applications to significantly improve model state and forecasts. Perhaps the most important application has been in numerical weather prediction where data assimilation has contributed to the forecast accuracy being increased at longer lead times, with the four day forecast in 2014 having the same level of accuracy as the one day forecast in 1979 (Bauer et al., 2015). This increase in forecast skill is obviously not solely due to data assimilation but also increased quality and resolution of observations along with improvements in model structure, however the introduction and evolution of data assimilation has played a large part (Dee et al., 2011). The current method implemented at many leading operational numerical weather prediction centres is known as Four-Dimensional Variational data assimilation (4D-Var) (Bonavita et al., 2015; Clayton et al., 2013), which has been shown to be a significant improvement over its predecessor three-dimensional variational data assimilation (Lorenc and Rawlins, 2005). Variational assimilation techniques minimise a cost function to find the optimal state of a system given all available knowledge of errors in the model and observations. The minimisation routine typically requires the derivative of the model which can sometimes prove difficult to calculate. Using techniques such as automatic-differentiation (Renaud, 1997) can reduce the time taken to implement the derivative of a model.

In numerical weather prediction data assimilation has been predominately used for state estimation whilst keeping parameters fixed. This is because numerical weather prediction is mainly dependent on the initial state with model physics being well understood. Ecosystem carbon cycle models are more dependent on finding the correct set of parameters to describe the ecosystem of interest (Luo et al., 2015). This is possibly why Monte Carlo Markov chain (MCMC) data assimilation methods have been used more with ecosystem carbon cycle models. Smaller ecosystem models are much less computationally expensive to run than large numerical weather prediction models, meaning that MCMC methods (requiring many more model runs than variational assimilation methods) are more easily implemented. For larger scale and more complex ecosystem models variational methods represent a much more computationally efficient option for data assimilation. Variational data assimilation can be used for joint parameter and state estimation by augmenting the state vector with the parameters (Navon,

1998). By including the parameters in the state vector we must also specify error statistics and error correlations for them. Smith et al. (2009) show that the prescription of these error statistics and their correlations can have a significant impact on parameter-state estimates obtained from the assimilation.

Many different observations relevant to the carbon balance of forests have now been combined with functional ecology models, using data assimilation, in order to improve our knowledge of ecological systems (Zobitz et al., 2011; Fox et al., 2009; Richardson et al., 2010; Quaife et al., 2008; Zobitz et al., 2014; Niu et al., 2014). Two such models that have been used extensively with data assimilation are the Data Assimilation Linked Ecosystem Carbon (DALEC) model (Williams et al., 2005) and the Simplified Photosynthesis and Evapo-Transpiration (SIPNET) model (Braswell et al., 2005). Nearly all data assimilation routines built with these models have used sequential and Monte Carlo Markov chain (MCMC) data assimilation methods with the exception of a variational routine being implemented for DALEC by Delahaies et al. (2013). There have been examples of global land surface models being implemented with variational methods such as the ORganizing Carbon and Hydrology In Dynamic EcosystEms model (ORCHIDEE) (Krinner et al., 2005) and the Biosphere Energy Transfer HYdrology scheme (BETHY) in a Carbon Cycle Data Assimilation System (CCDAS) (Kaminski et al., 2013). These examples have mainly been used to assimilate data from satellite and atmospheric CO_2 observations with only a few cases where site level data has also been assimilated (Verbeeck et al., 2011; Bacour et al., 2015).

Forest carbon balance model parameters are often determined in advance of using the model for forecasting by calibration of the model against observations (Richardson et al., 2010; Bloom and Williams, 2015). Here we take the alternative approach of concurrent state-parameter estimation. A key difference between the joint state-parameter estimation approach and a priori calibration is the way that the observational data is used. Pre-calibration approaches train the model against historical data and so become infeasible when there is a lack of sufficient observational information prior to the model forecast period. Joint state-parameter estimation methods have the advantage that observations could be used as they arrive in real time, by sequential assimilation cycling. This approach also gives the possibility of adapting to changes in the forest (e.g., tree thinning, fires etc.) that may change the parameter values over time.

Background errors (describing our knowledge of error in prior model estimates before data assimilation) and observation errors have largely been treated as uncorrelated and independent in ecosystem model data assimilation schemes. In 3D and 4DVar schemes background and observation errors are represented by the error covariance matrices **B** and **R** respectively. The off-diagonal elements of these matrices indicate the correlations between errors in the parameter and state variables for **B** and the correlations between observation errors for **R**. In the assimilation, the off-diagonal terms in the **B** matrix act to spread information between the state and augmented parameter variables (Kalnay, 2003). This means that assimilating observations of one state variable can act to update different state and parameter variables in the assimilation when correlations are included in **B**. In 4D-Var the **B** matrix is propagated implicitly by the forecast model, so that even a propagated diagonal **B** matrix can develop correlations throughout an assimilation window. These correlations will only be in the propagated **B** matrix, with the **B** matrix valid at the initial time remaining unchanged. Including correlations in **B** has been shown to significantly improve

data assimilation results in numerical weather prediction (Bannister, 2008).

Including correlations between observation errors has only started to be explored recently in numerical weather prediction, with **R** still often treated as diagonal (Stewart et al., 2013). Including some correlation structure in **R** has been shown to improve forecast accuracy (Weston et al., 2014). Currently the correlations included in **R** have been mainly between observations made at the the same time rather than correlations between observations throughout time. When assimilating observations, data streams with many more observations can have a greater impact on the assimilation than those with fewer observations. In Richardson et al. (2010) this problem is discussed when assimilating large numbers of NEE observations along with smaller numbers of leaf area index and soil respiration observations. To address this problem Richardson et al. uses a cost function that calculates the product of the departures from the observations rather than a cost function which sums these departures, giving a relative rather than absolute measure of the goodness-of-fit to the observations. This problem is also encountered in Bacour et al. (2015) when assimilating daily eddy covariance data with weekly observations of the FrAction of Photosynthetically Active Radiation (FAPAR). In Bacour et al. (2015) the error in observations of FAPAR is divided by two in order to give these less frequent observations more weight in the assimilation algorithm. Specifying serial time correlations between observations represents another way of addressing this problem, whilst also adding valuable information to the data assimilation routine. Including serial correlations between observations of the same quantity decreases the impact of these observations (Järvinen et al., 1999) therefore increasing the impact of less frequent observations.

In this paper we implement the new version of DALEC (DALEC2 (Bloom and Williams, 2015)) in a 4D-Var data assimilation scheme for joint state and parameter estimation, assimilating daily NEE observations from the Alice Holt flux site in Hampshire, UK (Wilkinson et al., 2012). This assimilation scheme is then subjected to rigorous testing to ensure correctness. A new method is outlined for including parameter and state correlations in the background "prior" error covariance matrix. Currently parameter and state error statistics are largely treated as independent and uncorrelated when data assimilation has been used with models of carbon balance. We also introduce a novel method for including serial time correlations in the observation error covariance matrix. The idea of including time correlations between observation error statistics is new and has not been previously explored in carbon balance model data assimilation. These correlated matrices are then used in a series of experiments in order to examine the effect that including correlations in the assimilation scheme has on the results.

2. Model and Data Assimilation Methods

2.1. Alice Holt research forest

Alice Holt Forest is a research forest area managed by the UK Forestry Commission located in Hampshire, SE England. Forest Research has been operating a CO_2 flux measurement tower in a portion of the forest, the Straits Inclosure, since 1998 so it is one of the longer forest site CO_2 flux records, globally. The Straits Inclosure is a 90ha area of managed deciduous broadleaved plantation woodland, presently approximately 80 years old, on a surface water gley soil. The majority

of the canopy trees are oak (*Quercus robur* L.), with an understory of hazel (*Corylus avellana* L.) and hawthorn (*Crataegus monogyna* Jacq.); but there is a small area of conifers (*Pinus nigra* J. F. Arnold) within the tower measurement footprint area in some weather conditions. Further details of the Straits Inclosure site and the measurement procedures are given in Wilkinson et al. (2012), together with analysis of stand-scale 30 minute average net CO_2 fluxes (NEE) measured by standard eddy covariance methods from 1998-2011. The data used here span from January 1999 to December 2013, and consist of the NEE fluxes and meteorological driving data of temperatures, irradiance and atmospheric CO_2 concentration. The original NEE data were subjected to normal quality control procedures, including u^* filtering to remove unreliable data when there were low turbulence night time conditions, as described in Wilkinson et al. (2012), but were not gap-filled. To compute daily NEE observations we take the sum over the 48 measurements made each day. We only select days where there is no missing data and over 90% of CO_2 flux observations have a quality control flag associated with the best observations and no observations associated with the worst from the EddyPro flux processing software (LI-COR, Inc., 2015).

2.2. *The DALEC2 model*

The DALEC2 model is a simple process-based model describing the carbon balance of a forest ecosystem (Bloom and Williams, 2015) and is the new version of the original DALEC (Williams et al., 2005). The model is constructed of six carbon pools (labile (C_{lab}), foliage (C_f), fine roots (C_r), woody stems and coarse roots (C_w), fresh leaf and fine root litter (C_l) and soil organic matter and coarse woody debris (C_s)) linked via fluxes. The aggregated canopy model (ACM) (Williams et al., 1997) is used to calculate daily gross primary production (*GPP*) of the forest, taking meteorological driving data and the modelled leaf area index (a function of C_f) as arguments. Figure 1 shows a schematic of how the carbon pools are linked in DALEC2.

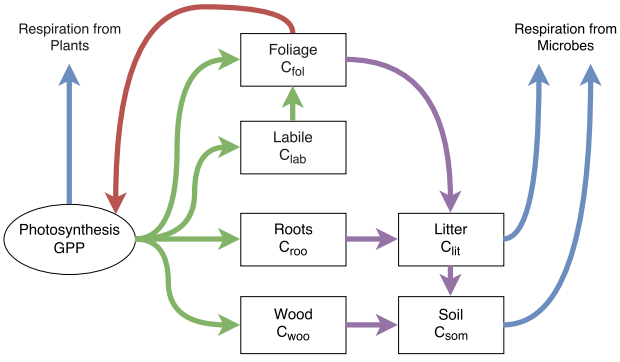


Figure 1: Representation of the fluxes in the DALEC2 carbon balance model. Green arrows represent C allocation, purple arrows represent litter fall and decomposition fluxes, blue arrows represent respiration fluxes and the red arrow represents the influence of leaf area index in the *GPP* function.

The model equations for the carbon pools at day i are as follows:

$$GPP^i = ACM(C_{fol}^{i-1}, c_{lma}, c_{eff}, \Psi) \tag{1}$$

$$C_{lab}^i = C_{lab}^{i-1} + (1-f_{auto})(1-f_{fol})f_{lab}GPP^i - \Phi_{on}C_{lab}^{i-1}, \tag{2}$$

$$C_{fol}^i = C_{fol}^{i-1} + \Phi_{on}C_{lab}^{i-1} + (1-f_{auto})f_{fol}GPP^i - \Phi_{off}C_{fol}^{i-1}, \tag{3}$$

$$C_{roo}^i = C_{roo}^{i-1} + (1-f_{auto})(1-f_{fol})(1-f_{lab})f_{roo}GPP^i - \theta_{roo}C_{roo}^{i-1}, \tag{4}$$

$$C_{woo}^i = C_{woo}^{i-1} + (1-f_{auto})(1-f_{fol})(1-f_{lab})(1-f_{roo})GPP^i - \theta_{woo}C_{woo}^{i-1}, \tag{5}$$

$$C_{lit}^i = C_{lit}^{i-1} + \theta_{roo}C_{roo}^{i-1} + \Phi_{off}C_{fol}^{i-1} - (\theta_{lit}+\theta_{min})e^{\Theta T^{i-1}}C_{lit}^{i-1}, \tag{6}$$

$$C_{som}^i = C_{som}^{i-1} + \theta_{woo}C_{woo}^{i-1} + \theta_{min}e^{\Theta T^{i-1}}C_{lit}^{i-1} - \theta_{som}e^{\Theta T^{i-1}}C_{som}^{i-1}, \tag{7}$$

where T^{i-1} is the daily mean temperature, Ψ represents the meteorological driving data used in the *GPP* function and Φ_{on}/Φ_{off} are functions controlling leaf-on and leaf-off. Descriptions for each model parameter used in equations (1) to (7) are included in the appendix in table 3. DALEC2 differs from the original DALEC in that it can be parameterised for both deciduous and evergreen sites with Φ_{on} and Φ_{off} being able to reproduce the phenology of either type of site. The full details of this version of DALEC can be found in Bloom and Williams (2015).

2.3. 4D-Var

Following the approach of Smith et al. (2011) for joint state and parameter estimation, we consider the discrete nonlinear dynamical system given by

$$\mathbf{z}_i = \mathbf{f}_{i-1\to i}(\mathbf{z}_{i-1}, \mathbf{p}_{i-1}), \tag{8}$$

where $\mathbf{z}_i \in \mathbb{R}^n$ is the state vector at time t_i, $\mathbf{f}_{i-1\to i}$ is the nonlinear model operator propagating the state at time t_{i-1} to time t_i for $i = 1,2,\ldots,N$ and $\mathbf{p}_{i-1} \in \mathbb{R}^q$ is a vector of q model parameters at time t_{i-1}. For DALEC2 the state vector $\mathbf{z}_i = (C_{lab}^i, C_{for}^i, C_{roo}^i, C_{woo}^i, C_{lit}^i, C_{som}^i)^T$, with the parameters shown in table 3. Given a set of fixed parameters, the value of the forecast at time t_i is uniquely determined by the initial value. The model parameters are not updated by the nonlinear model operator, therefore the evolution of the parameters is given by,

$$\mathbf{p}_i = \mathbf{p}_{i-1}, \tag{9}$$

for $i = 1,2,\ldots,N$. We define the new vector $\mathbf{x}$ by joining the parameter vector $\mathbf{p}$ with the model state vector $\mathbf{z}$, giving us the augmented state vector

$$\mathbf{x} = \begin{pmatrix} \mathbf{p} \\ \mathbf{z} \end{pmatrix} \in \mathbb{R}^{q+n}. \tag{10}$$

We define the augmented system model by

$$\mathbf{x}_i = \mathbf{m}_{i-1\to i}(\mathbf{x}_{i-1}), \tag{11}$$

where

$$\mathbf{m}_{i-1\to i}(\mathbf{x}_{i-1}) = \begin{pmatrix} \mathbf{p}_{i-1} \\ \mathbf{f}_{i-1\to i}(\mathbf{z}_{i-1}, \mathbf{p}_{i-1}) \end{pmatrix} = \begin{pmatrix} \mathbf{p}_i \\ \mathbf{z}_i \end{pmatrix} \in \mathbb{R}^{q+n}. \tag{12}$$

The available observations at time t_i are represented by the vector $\mathbf{y}_i \in \mathbb{R}^{r_i}$ which are related to the augmented state vector through the equation

$$\mathbf{y}_i = \mathbf{h}_i(\mathbf{x}_i) + \boldsymbol{\varepsilon}_i, \tag{13}$$

where $\mathbf{h}_i : \mathbb{R}^{q+n} \to \mathbb{R}^{r_i}$ is the observation operator mapping the augmented state vector to observation space and $\boldsymbol{\varepsilon}_i \in \mathbb{R}^{r_i}$ represents the observation errors. These errors are usually assumed to be unbiased, Gaussian and serially uncorrelated with known covariance matrices $\mathbf{R}_i$.

In the 4D-Var data assimilation detailed here we aim to find the parameter and initial state values such that the model trajectory best fits the data over some time window, given some prior information about the system. The output from 4D-Var is an updated set of parameters, and an updated model state, valid at the beginning of the time window. The updated model state may be used as initial conditions for a forecast using the full nonlinear DALEC2 model. We assume that at time t_0 we have an initial estimate to the augmented state, usually referred to as the background vector denoted $\mathbf{x}^b$. This background is assumed to have unbiased, Gaussian errors with known covariance matrix $\mathbf{B}$. Adding the background term ensures that our problem is well posed and that we can find a locally unique solution (Tremolet, 2006). In 4D-Var we aim to find the initial state that minimises the weighted least squares distance to the background while minimising the weighted least squares distance of the model trajectory to the observations over the time window $t_0, \ldots, t_N$ (Lawless, 2013). We do this by finding the state $\mathbf{x}_0^a$ at time t_0 that minimises the cost function

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i))^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)), \tag{14}$$

subject to the augmented states $\mathbf{x}_i$ satisfying the nonlinear dynamical model (11). The state that minimises the cost function, $\mathbf{x}_0^a$, is commonly called the analysis. This state is found using a minimisation routine that takes as its input arguments the cost function, the background vector ($\mathbf{x}^b$) and also the gradient of the cost function given as,

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) - \sum_{i=0}^{N} \mathbf{M}_{i,0}^T \mathbf{H}_i^T \mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{h}_i(\mathbf{x}_i)) \tag{15}$$

where $\mathbf{H}_i = \frac{\partial \mathbf{h}_i(\mathbf{x}_i)}{\partial \mathbf{x}_i}$ is the linearized observation operator and $\mathbf{M}_{i,0} = \mathbf{M}_{i-1}\mathbf{M}_{i-2}\cdots\mathbf{M}_0$ is the tangent linear model with $\mathbf{M}_i = \frac{\partial \mathbf{m}_{i-1\to i}(\mathbf{x}_i)}{\partial \mathbf{x}_i}$. In practice $\nabla J(\mathbf{x}_0)$ is calculated using the method of Lagrange multipliers as shown in Lawless (2013). We can

rewrite the cost function and its gradient to avoid the sum notation as,

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2}(\hat{\mathbf{y}} - \hat{\mathbf{h}}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{h}}(\mathbf{x}_0)) \tag{16}$$

and

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}^b) - \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{\mathbf{h}}(\mathbf{x}_0)), \tag{17}$$

where,

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \ \hat{\mathbf{h}}(\mathbf{x}_0) = \begin{pmatrix} \mathbf{h}_0(\mathbf{x}_0) \\ \mathbf{h}_1(\mathbf{m}_{0\to1}(\mathbf{x}_0)) \\ \vdots \\ \mathbf{h}_N(\mathbf{m}_{0\to N}(\mathbf{x}_0)) \end{pmatrix}, \ \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \dots & \mathbf{R}_{0,N} \\ \mathbf{R}_{1,0} & \mathbf{R}_{1,1} & \dots & \mathbf{R}_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{N,0} & \mathbf{R}_{N,1} & \dots & \mathbf{R}_{N,N} \end{pmatrix} \text{ and } \hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{N,0} \end{pmatrix}. \tag{18}$$

Solving the cost function in this form also allows us to build serial time correlations into the observation error covariance matrix $\hat{\mathbf{R}}$. The off-diagonal blocks of $\hat{\mathbf{R}}$ represent correlations in time between assimilated observations and are usually taken to be zero. In section 2.6 we show how these off-diagonal blocks can be specified. We can also calculate the posterior or analysis error covariance matrix after assimilation as,

$$\mathbf{A} = (\mathbf{B}^{-1} + \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1} \hat{\mathbf{H}})^{-1}. \tag{19}$$

We can use this matrix to estimate the uncertainty in our parameter and initial state variables after assimilation.

2.4. Implementation and testing of 4D-Var system

In our DALEC2 4D-Var scheme we are performing joint parameter and state estimation. Typically MCMC techniques have been used for joint parameter and state estimation with functional ecology models, such as DALEC2. However 4D-Var has been used for joint parameter and state estimation with global carbon cycle models (Kaminski et al., 2013). The variational approach is computationally efficient and robust, making it particularly suited to large problems with complex models. The augmented state vector, $\mathbf{x}_0$, corresponds to the vector of the 17 model parameters and 6 initial carbon pool values, which can be found in the appendix in table 3. Here the nonlinear model (DALEC2) only updates the initial carbon pool values when evolving the augmented state vector forward in time with the parameters being held constant. To find the background estimate, $\mathbf{x}^b$, to the augmented state vector we can either use a previous DALEC2 model forecast estimate of the state of the system for the site (when available) or use expert elicitation to define likely state and parameter values and ranges for the site. The background vector $(\mathbf{x}^b)$ and its corresponding standard deviations (see table 3) used in this paper were provided from existing runs of the the CARbon DAta-MOdel fraMework (CARDAMOM) (Exbrayat et al., 2015). The CARDAMOM

output is a dataset derived from satellite observations of leaf area index which provides a reasonable first guess to DALEC2 state and parameter values for the Alice Holt research site. In this paper we assimilate observations of daily NEE. From Richardson et al. (2008) the measurement error in observations of daily NEE is between 0.2 to 0.8 $\mathrm{gCm^{-2}day^{-1}}$. Richardson et al. (2008) also shows that flux errors are heteroscedastic. We assume a constant standard deviation of 0.5 $\mathrm{gCm^{-2}day^{-1}}$ in the assimilated observations of daily NEE as we found this standard deviation gave the best weighting to the observations in the assimilation algorithm, producing the best results for the forecast of NEE after assimilation. Assuming this constant standard deviation also allows for correlations in time between observation errors to be included more easily. Ignoring the heteroscedastic nature of NEE errors may influence results by giving observations of larger magnitude a higher weight than would be realistic. Future work should try to incorporate the heteroscedastic nature of NEE errors.

In order to find the tangent linear model (TLM) for DALEC2 it is necessary to find the derivative of the model at each time step with respect to the 17 model parameters and the 6 carbon pools. We use the AlgoPy automatic differentiation package (Walter and Lehmann, 2013) in Python to calculate the TLM at each time step. This package uses forward mode automatic differentiation to calculate the derivative of the model. In the following tests we use a diagonal approximation to the background and observation error covariance matrices so that, $\mathbf{B}_{diag} = \mathrm{diag}(\boldsymbol{\sigma}_b)^2$ and $\hat{\mathbf{R}}_{diag} = \mathrm{diag}(\boldsymbol{\sigma}_o)^2$, where $\boldsymbol{\sigma}_b$ is the vector of background standard deviations found in table 3 and $\boldsymbol{\sigma}_o$ is the vector of observational standard deviations, for a single observation of NEE $\sigma_o = 0.5$ $\mathrm{gCm^{-2}day^{-1}}$. To minimise the cost function we use the truncated Newton iteration method (Nocedal and Wright, 1999) from the Python package Scipy.optimize (Jones et al., 2001). This method uses a number of stopping criteria to ensure convergence to a minimum of our cost function. In sections 2.4.1 to 2.4.3 we show tests of our scheme.

2.4.1. Test of tangent linear model

The TLM is used in the calculation of the gradient of our cost function in 4D-Var. We can have confidence that our implementation of the TLM for DALEC2 is correct as it passes the following relevant tests (Li et al., 1994). In 4D-Var we assume the tangent linear hypothesis,

$$\mathbf{m}_{0\rightarrow i}(\mathbf{x}_0 + \gamma\delta\mathbf{x}_0) \approx \mathbf{m}_{0\rightarrow i}(\mathbf{x}_0) + \gamma\mathbf{M}_{i,0}\delta\mathbf{x}_0, \tag{20}$$

where $\delta\mathbf{x}_0$ is a perturbation of the initial augmented state $\mathbf{x}_0$ and γ is a parameter controlling the size of this perturbation. The validity of this assumption depends on how nonlinear the model is, the length of the assimilation window and the size of the augmented state perturbation $\delta\mathbf{x}_0$. We can test this by rearranging equation (20) to find,

$$\frac{||\mathbf{m}_{0\rightarrow i}(\mathbf{x}_0 + \gamma\delta\mathbf{x}_0) - \mathbf{m}_{0\rightarrow i}(\mathbf{x}_0) - \gamma\mathbf{M}_{i,0}\delta\mathbf{x}_0||}{||\gamma\mathbf{M}_{i,0}\delta\mathbf{x}_0||} \rightarrow 0, \tag{21}$$

as $\gamma \rightarrow 0$ (here we are using the Euclidean norm). Equation (21) should hold if our implementation of the TLM is correct, even for a weakly non-linear model. Figure 2 shows equation (21) plotted for DALEC2 with i fixed at 731 days, a fixed

5% perturbation $\delta\mathbf{x}_0$ and values of γ approaching zero . Figure 2 shows that the TLM behaves as expected for values of γ approaching 0. This was also tested for different choices of $\mathbf{x}_0$ and sizes of perturbation with similar results.

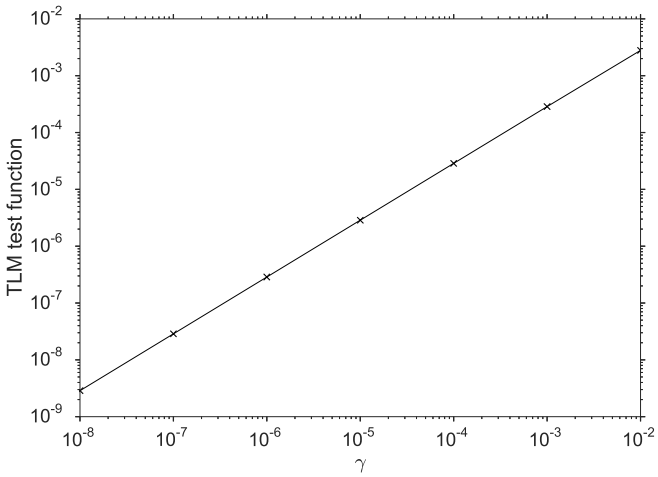


Figure 2: Plot of the tangent linear model test function (equation (21)) for DALEC2, for a fixed TLM evolving the perturbed augmented state 731 days forward in time and a fixed 5% perturbation, $\delta\mathbf{x}_0$.

It is also useful to show how the TLM behaves over a time window to see how the error in the TLM grows as we evolve the augmented state further forward in time. We again rearrange equation (20) with an additional error term to find,

$$\text{percentage error in TLM} = \frac{||m_{0\to i}(\mathbf{x}_0+\gamma\delta\mathbf{x}_0)-m_{0\to i}(\mathbf{x}_0)-\gamma\mathbf{M}_{i,0}\delta\mathbf{x}_0||}{||\gamma\mathbf{M}_{i,0}\delta\mathbf{x}_0||}\times 100. \tag{22}$$

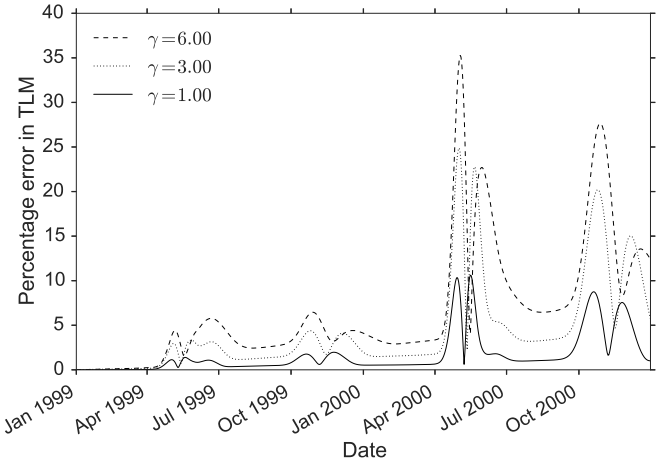


Figure 3: Plot of the percentage error in the tangent linear model (equation (22)) for DALEC2 when evolving the model state forward over a period of two years with three different values of γ and a fixed 5% perturbation $\delta\mathbf{x}_0$.

In figure 3 we plot the percentage error in the TLM tested throughout a two-year period as DALEC2 is run forward. From figure 3 we can see that the TLM for DALEC2 performs well after being run forward a year with less than a 7% error for all values of γ. By the second year we see some peaks in the error in spring and autumn. This is due to leaf on and leaf off functions in the TLM going out of phase with the nonlinear DALEC2. At these peaks the error reaches a maximum at 35% then coming back to around 10% before growing again in the autumn. Although this level of error is still acceptable we

present results using a one year assimilation window in this paper as in practice we could cycle assimilation windows to make use of multiple years of data (Moodey et al., 2013).

2.4.2. Test of adjoint model

The adjoint model we have implemented for DALEC2 passes correctness tests. For the TLM $\mathbf{M}_{i,0}$ and its adjoint $\mathbf{M}_{i,0}^T$ we have the identity

$$< \mathbf{M}_{i,0}\delta\mathbf{x}_0, \mathbf{M}_{i,0}\delta\mathbf{x}_0 > = < \delta\mathbf{x}_0, \mathbf{M}_{i,0}^T\mathbf{M}_{i,0}\delta\mathbf{x}_0 > \tag{23}$$

for any inner product $<,>$ and perturbation $\delta\mathbf{x}_0$. This is derived from the adjoint identity (Lawless, 2013). Using the Euclidean inner product, equation (23) is equivalent to

$$(\mathbf{M}_{i,0}\delta\mathbf{x}_0)^T(\mathbf{M}_{i,0}\delta\mathbf{x}_0) = \delta\mathbf{x}_0^T(\mathbf{M}_{i,0}^T(\mathbf{M}_{i,0}\delta\mathbf{x}_0)). \tag{24}$$

We evaluated the left hand side and right hand side of this identity for differing values of $\mathbf{x}_0$ and size of perturbation $\delta\mathbf{x}_0$ and showed that they were equal to machine precision.

2.4.3. Gradient test

The 4D-Var system we have developed passes tests for the gradient of the cost function (Navon et al., 1992). In the implementation of the cost function and its gradient we regularise the problem using a variable transform (Freitag et al., 2010). For the cost function J and its gradient ∇J we can show that we have implemented ∇J correctly using the identity,

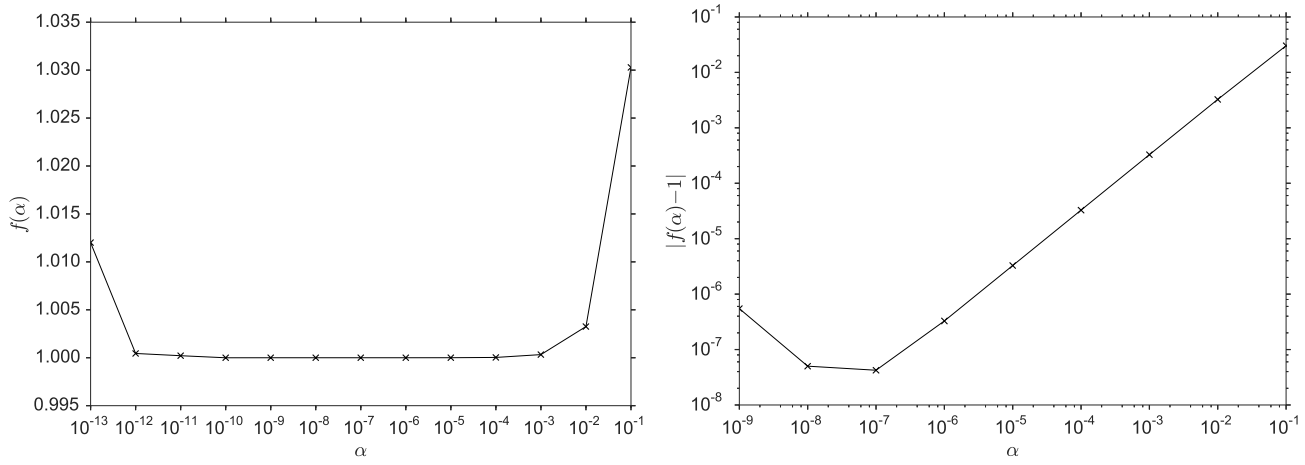
$$f(\alpha) = \frac{|J(\mathbf{x}_0 + \alpha\mathbf{b}) - J(\mathbf{x}_0)|}{\alpha\mathbf{b}^T\nabla J(\mathbf{x}_0)} = 1 + O(\alpha), \tag{25}$$

where $\mathbf{b}$ is a vector of unit length and α is a parameter controlling the size of the perturbation. For small values of α not too close to machine precision we should have $f(\alpha)$ close to 1. Figure 4a shows $f(\alpha)$ for a 365 day assimilation window with $\mathbf{b} = \mathbf{x}_0||\mathbf{x}_0||^{-1}$, we can see that $f(\alpha) \to 1$ as $\alpha \to 0$, as expected until $f(\alpha)$ gets too close to machine zero at order $\alpha = 10^{-11}$. This was also tested with $\mathbf{b}$ in different directions and similar results obtained.

We can also plot $|f(\alpha) - 1|$, where we expect $|f(\alpha) - 1| \to 0$ as $\alpha \to 0$. In figure 4b we have plotted $|f(\alpha) - 1|$ for the same conditions as in figure 4a, we can see that $|f(\alpha) - 1| \to 0$ as $\alpha \to 0$, as expected. This gives us confidence that the gradient of the cost function is implemented correctly.

2.5. Including correlations in the background error covariance matrix

As discussed in section 1, including correlations in $\mathbf{B}$ impacts how information from assimilated observations is spread between different types of analysis variables (Bannister, 2008). We explored a number of different methods in order to



(a) $f(\alpha)$ test

(b) $|f(\alpha)-1|$ test

Figure 4: Tests of the gradient of the cost function for a 365 day assimilation window with $\mathbf{b} = \mathbf{x}_0||\mathbf{x}_0||^{-1}$.

include parameter-state correlations in $\mathbf{B}$. In this paper we present a method using a set of ecological dynamical constraints, based on expert judgement, on model parameters and state variables from Bloom and Williams (2015). Bloom and Williams (2015) show that implementing these constraints in a Metropolis Hastings MCMC data assimilation routine improves results significantly. The constraints impose conditions on carbon pool turnover and allocation ratios, steady state proximity and growth and the decay of model carbon pools.

In order to create a correlated background error covariance matrix, $\mathbf{B}_{corr}$, using these constraints we create an ensemble of state vectors which we then take the covariance of to give us $\mathbf{B}_{corr}$. To create this ensemble we use the following procedure:

1. Draw a random augmented state vector, $\mathbf{x}_i$, from the multivariate truncated normal distribution described by

$$\mathbf{x}_i \sim \mathcal{N}(\mathbf{x}^b, \mathbf{B}_{diag}), \tag{26}$$

   where $\mathbf{B}_{diag}$ is the diagonal matrix described in section 2.4 and $\mathbf{x}_i$ is bound by the parameter and state ranges given in table 3 in the appendix.

2. Test this $\mathbf{x}_i$ with the ecological dynamical constraints (requiring us to run the DALEC2 model using this state).

3. If $\mathbf{x}_i$ passes it is added to our ensemble, else it is discarded.

Once we have a full ensemble we then take the covariance of the ensemble to find $\mathbf{B}_{corr}$. We chose an ensemble size of 1500 as a qualitative assessment using a larger ensemble showed little difference in correlations. In figure 5 we have plotted the correlation matrix or normalised error covariance matrix associated with $\mathbf{B}_{corr}$. This matrix includes both positive and negative correlations between parameter and state variables, with correlations of 1 down the diagonal between variables of the same quantity as expected. The largest positive off-diagonal correlation is 0.42 between f_{lab} and C_{lab}. This makes physical sense as f_{lab} is the parameter controlling the amount of GPP allocated to the labile carbon pool, C_{lab}.

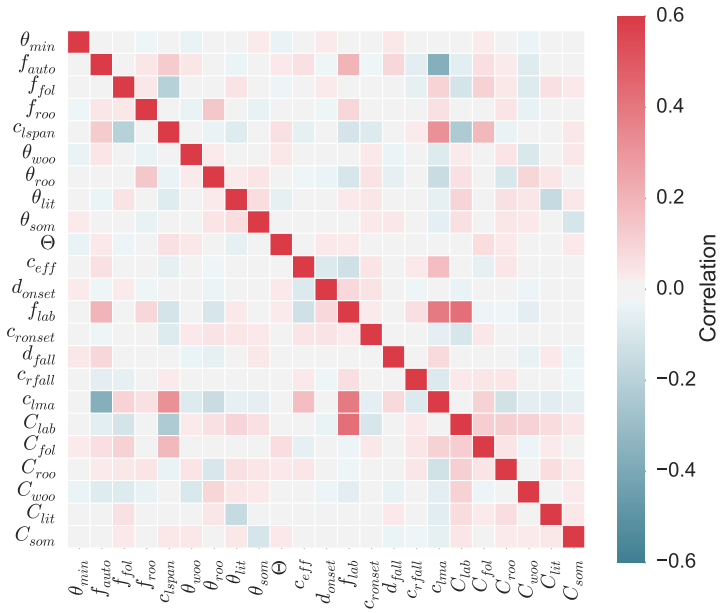


Figure 5: Background error correlation matrix created using method in section 2.5. Here the correlation scale for off-diagonal values ranges from -0.5 to 0.5 with the correlation along the diagonal being 1. For explanation of parameter and state variable symbols see table 3.

2.6. Specifying serial correlations in the observation error covariance matrix

The observation error covariance matrix does not only represent the instrumentation error for an observation but also the error in the observation operator (mapping the model state to the observation) and representativity error (error arising from the model being unable resolve the spatial and temporal scales of the observations). These other sources of error represented in $\hat{\mathbf{R}}$ can also lead to correlations between observation errors (Waller et al., 2014). Errors in NEE observations come from different sources such as instrument errors, sampled ecosystem structure from the variable footprint of the flux tower and turbulent conditions (when there is low turbulence and limited air mixing the magnitude of NEE is underestimated). These errors due to turbulence can still have effect even after u^* filtering (Papale et al., 2006). Due to this dependence on atmospheric conditions we expect the errors in observations of NEE to be serially correlated, as the atmospheric signal itself is serially correlated (Daley, 1992). If we were assimilating half hourly observations of NEE we would expect stronger correlations between observation errors, as atmospheric conditions are more constant at this time scale, with correlations between observation errors getting weaker with lower frequency observations. Although some studies suggest that the correlation between NEE measurement errors on the scale of a day is negligible (Lasslop et al., 2008), it is also likely that error in the observation operator and representativity error will lead to observation error correlations for NEE (Waller et al., 2014).

In section 2.3 we have re-written the 4D-Var cost function in equation (16) in order to allow the specification of serial observation error correlations in our assimilation scheme. These serial correlations are represented by the off-diagonal blocks of $\hat{\mathbf{R}}$. In work carried out with spatial correlations it has been shown that the structure of the correlation is not critical and that it is better to include some estimate of error correlation structure in the observation error covariance matrix than wrongly assume that errors are independent (Stewart et al., 2013; Healy and White, 2005). As a first attempt we try including temporal

correlations on the scale of the observation frequency. We adapt the simple Gaussian model found in Järvinen et al. (1999) (a second order autoregressive correlation function was also tested but is not presented here). The correlation r between 2 observations at times t_1 and t_2 is given as,

$$r = \begin{cases} a\exp\left[\frac{-(t_1-t_2)^2}{\tau^2}\right] + (1-a)\delta_{t_1-t_2} & |t_1-t_2| \leq \eta \\ 0 & \eta < |t_1-t_2| \end{cases}, \tag{27}$$

where τ is the e-folding time in days, a controls the strength of correlation, δ is the Kronecker delta and η is the cut off time after which the correlation between two observation errors is zero. We have incorporated a cut off for correlations between observation errors as the assumed correlation length scale for the assimilated observations is short. This cut off along with the form of correlation function using the Kronecker delta helps ensure $\hat{\mathbf{R}}$ is positive definite and therefore invertible, as required in the assimilation process. The standard deviation assumed in the observations of NEE is 0.5 $\mathrm{gCm^{-2}day^{-1}}$ as described in section 2.4.

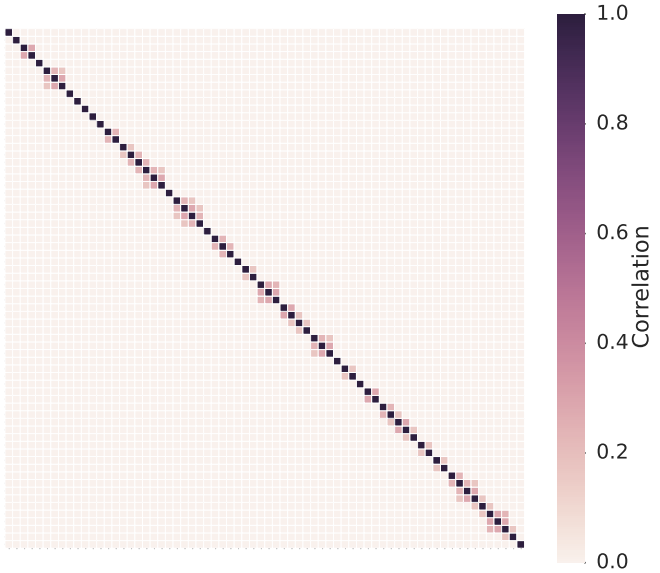


Figure 6: Observation error correlation matrix for the 67 observations used in assimilation created using method in section 2.6 with $\tau = 4$, $a = 0.3$ and $\eta = 4$.

Figure 6 shows the correlation matrix for $\hat{\mathbf{R}}$ created using equation (27). Here observations made on adjacent days will have an error correlation of 0.3; this will then decay exponentially for observations farther apart in time. There are 67 NEE observations in this one year assimilation window. These observations are not all on adjacent days and this is evident in the structure of $\hat{\mathbf{R}}$. The effect of the short e-folding time chosen here ($\tau = 4$) provides the desired structure.

3. Results

3.1. Experiments

In the following sections we present the results of four experiments where we vary the representations of $\mathbf{B}$ and $\hat{\mathbf{R}}$ while assimilating the same NEE observations in the window from the beginning of January 1999 to the end of December 1999. As shown in figure 3 the performance of the tangent linear model deteriorates after the first year. We then forecast the NEE over the next 14 years (Jan 2000 - Dec 2013) and compare with the observed data. Using this shorter analysis window with a long forecast allows us to see the effect of including correlations in the error statistics more clearly, as we have a longer time-series of data with which to judge our forecast after data assimilation. These experiments are outlined in table 1 where $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ are the diagonal matrices of the parameter and state variances and the observations variances respectively and $\mathbf{B}_{corr}$ and $\hat{\mathbf{R}}_{corr}$ are the matrices as specified in section 2.5 and section 2.6 respectively.

Experiment	$\mathbf{B}_{diag}$	$\hat{\mathbf{R}}_{diag}$	$\mathbf{B}_{corr}$	$\hat{\mathbf{R}}_{corr}$
A	$\times$	$\times$		
B		$\times$	$\times$	
C	$\times$			$\times$
D			$\times$	$\times$

Table 1: The combination of error covariance matrices used in each data assimilation experiment.

3.2. Experiment A

In this experiment $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ were used in the assimilation as described in section 3.1. Because these contain no correlations this experiment forms the baseline by which the subsequent results from assimilation experiments are judged.

Figure 7a shows assimilation and forecast results for NEE. We can see that assimilating the observations of NEE has improved the background with the analysis trajectory (green line) fitting well with the observations during the assimilation window (Jan 1999- Dec 1999). The analysis trajectory then diverges in the forecast (Jan 2000 - Dec 2013). This can be seen more clearly in figue 8a, where there is an over prediction of respiration in the winter and the seasonal cycle does not match that of the observations. This is also shown in figure 9a where we have plotted the model-data differences for a year's period averaged over the 14 years in the forecast period. Figure 9a shows that the largest errors in our posterior model forecast occur as a result of not capturing the phenology of the site correctly, in particular the start of the season from April to June.

To see how well the forecast performs after assimilation we show a scatter plot of modelled NEE against observed NEE in figure 10b. From table 2 the predictions have a Root-Mean-Square Error (RMSE) of 4.22 $\mathrm{gCm^{-2}day^{-1}}$ and a bias of -0.3 $\mathrm{gCm^{-2}day^{-1}}$ for the forecast of NEE, whereas the analysis (Jan 1999 - Dec 1999) has a RMSE of 1.36 $\mathrm{gCm^{-2}day^{-1}}$ and a bias of -0.03 $\mathrm{gCm^{-2}day^{-1}}$. The background trajectory is the model trajectory for DALEC2 when run using the prior

estimate of the parameter and initial state values described in section 2.4. The background or prior model trajectory has a RMSE of 3.86 $\mathrm{gCm^{-2}day^{-1}}$ and a bias of -1.60 $\mathrm{gCm^{-2}day^{-1}}$ in the analysis window (Jan 1999 - Dec 1999) and the same RMSE of 3.86 $\mathrm{gCm^{-2}day^{-1}}$ but a bias of -1.36 $\mathrm{gCm^{-2}day^{-1}}$ during the forecast period (Jan 2000 - Dec 2013). Although using $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ in the assimilation has considerably reduced the RMSE in the analysis period, it has also increased the RMSE in the forecast of NEE. However it has reduced the bias in the model forecast considerably from -1.36 $\mathrm{gCm^{-2}day^{-1}}$ to -0.3 $\mathrm{gCm^{-2}day^{-1}}$. The bias in the background is due to the background model predicting less negative values of NEE than observed (i.e. above the 1:1 line shown in figure 10a). This leads to considerably worse results for the background trajectory than the analysis and its forecast for total forest carbon uptake. It is important to compare our results here with the background trajectory. The background acts as our initial prior model estimate and is the starting point for our minimisation in 4D-Var. Comparing our assimilation results with our background trajectory give us confidence that our 4D-Var scheme is improving the results of our model after assimilation.

3.3. Experiment B

Here $\mathbf{B}_{corr}$ (as defined in section 2.5) and $\hat{\mathbf{R}}_{diag}$ are used in the assimilation. Figure 7b shows assimilation and forecast results for NEE. In figure 8b we can see that the forecast performs considerably better than in experiment A, with the analysis trajectory no longer over predicting winter respiration and matching the observed seasonal cycle of NEE more closely in the forecast period (Jan 2000 - Dec 2013). This can be seen more clearly in figure 9b where the improvement in the period April-June is considerable as we capture green-up at the site more closely. Even though we have improved the representation of leaf-on in our model significantly here we can see from figure 8b that this is still where we have the largest uncertainty for our model after assimilation. From figure 10c and table 2 we see that the forecast RMSE has almost halved (now 2.56 $\mathrm{gCm^{-2}day^{-1}}$) with a reduction in bias also, now -0.2 $\mathrm{gCm^{-2}day^{-1}}$. In comparison using $\mathbf{B}_{corr}$ in the assimilation very slightly degrades the fit for the analysis (Jan 1999 - Dec 1999), with a RMSE of 1.42 $\mathrm{gCm^{-2}day^{-1}}$ and a bias of -0.04 $\mathrm{gCm^{-2}day^{-1}}$, as shown in table 2.

As discussed in section 1 previous work has shown the importance of specifying parameter-state correlations when using variational data assimilation for joint parameter and state estimation (Smith et al., 2009). In 4D-Var the initial correlation structure is evolved implicitly through time. However, in order to make full use of the observations it is essential to specify an accurate estimate to the initial correlation structure. Therefore by not specifying these correlations in experiment A we allow the parameter and state variables to attain unrealistic values in order to find the best fit to the observations in the analysis window (Jan 1999 - Dec 1999), leading to the divergence seen in the forecast (1999-2014) in experiment A.

We can see the effect that including correlations in $\mathbf{B}$ has on the analysis update in figure 11. For some variables including correlations in $\mathbf{B}$ has had a large impact on the analysis update after assimilation. This is particularly clear for the f_{lab} parameter. The largest positive off-diagonal correlation in $\mathbf{B}_{corr}$ is between C_{lab} and f_{lab}, with f_{lab} also having a large positive correlation with c_{lma} as shown in section 2.5. The effect of these correlations has been to change the analysis increment for

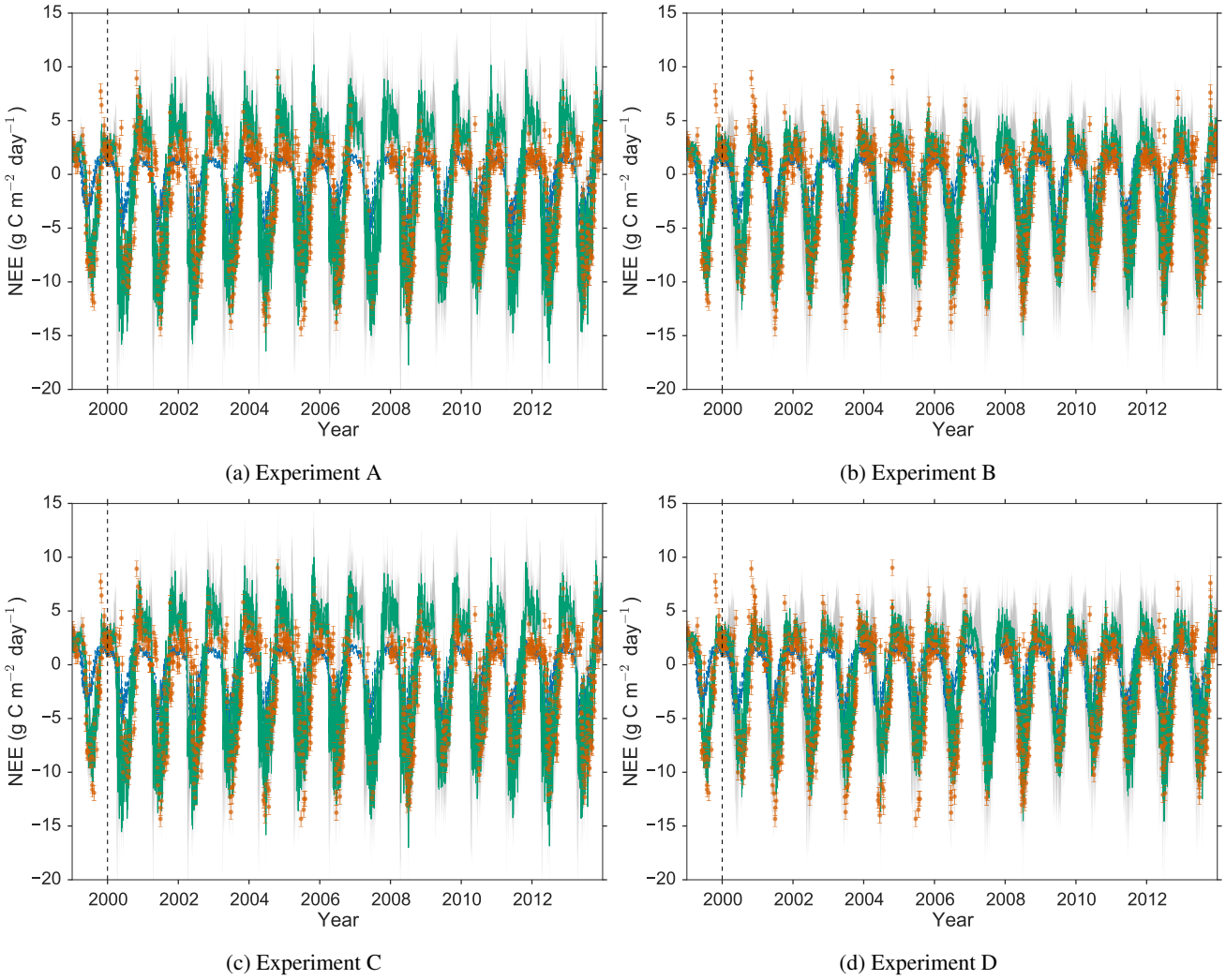
f_{lab} from being slightly positive in experiment A to being strongly negative by following the analysis update of its correlated variables C_{lab} and c_{lma}. From figure 11 we can also see some of the possible reasons for the improved fit to the observations in experiment B. From figure 9a the largest errors in our model forecast of NEE in experiment A stem from a misrepresentation of the phenology of the site in the months April-June. We see that the parameter controlling day of leaf on (d_{onset}) has been updated slightly differently in comparison to experiment A, with day of leaf on now being slightly later in the year (day 124 instead of 119), again this is due to the included correlations in $\mathbf{B}$. Even this small change in d_{onset} appears to reduce the errors at the start of the season for experiment B as seen from figure 9b. The forecast is also no longer over predicting winter respiration to the same extent as in experiment A. From figure 11 we see that the main parameters controlling ecosystem respiration in NEE (f_{auto}, θ_{lit}, θ_{som}, Θ) have been reduced in comparison with experiment A, which we believe have led to an improved fit to observations in experiment B. In experiment A we also had an over prediction of peak carbon uptake in summer which has been improved in this experiment. From figure 11 we see that one of the parameters controlling the magnitude of gross primary productivity (c_{eff}) has been decreased in comparison to experiment A. This appears to lead to less extreme predictions of peak summer carbon uptake than in experiment A. Two parameters with a significant change from experiment A are f_{fol} and C_{lit}; however in Chuter (2013) the DALEC model prediction of NEE is shown to be largely insensitive to variations in these parameters.

The added constraints provided by the correlations in $\mathbf{B}_{corr}$ acts to regularise the data assimilation problem and avoid overfitting to the assimilated data by limiting the parameter space of the problem (Smith et al., 2009). These correlations have been diagnosed using the EDC's from Bloom and Williams (2015), as shown in section 2.5, and help to limit unrealistic behaviour for a mature forest site. Although this has led to a slightly degraded fit to the observations in the analysis window (Jan 1999 - Dec 1999) it has also significantly improved the fit to observations for the forecast (Jan 2000 - Dec 2013).

3.4. Experiment C

Here we use $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{corr}$ (as defined in section 2.6) in the assimilation. Results shown in figure 7c, 8c and 9c appear similar to those in section 3.2 however there are some differences. From table 2 and figure 10d we see a slight reduction in RMSE for the forecast (now 4.09 $\mathrm{gCm^{-2}day^{-1}}$) in comparison with experiment A. As in experiment B the fit to the observations in the analysis window (Jan 1999 - Dec 1999) is very slightly degraded as the added correlations in $\hat{\mathbf{R}}_{corr}$ act to reduce the weight of the observations in the assimilation (Järvinen et al., 1999). The changes seen when using $\hat{\mathbf{R}}_{corr}$ in the assimilation are less than when using $\mathbf{B}_{corr}$ as the correlations specified in $\hat{\mathbf{R}}_{corr}$ are on a short timescale and much weaker than those in $\mathbf{B}_{corr}$. In figure 11 we can see that the changes between experiment A and C in the analysis increment are much less than when using $\mathbf{B}_{corr}$.

We also expect that specifying time correlations in $\hat{\mathbf{R}}$ will help when assimilating other less frequently sampled data streams along with NEE as the serial correlations reduce the weight given to the mean of the more frequently sampled observations (here NEE) and also reduce the information content of these more frequently sampled observations (Järvinen et al., 1999;



(a) Experiment A

(b) Experiment B

(c) Experiment C

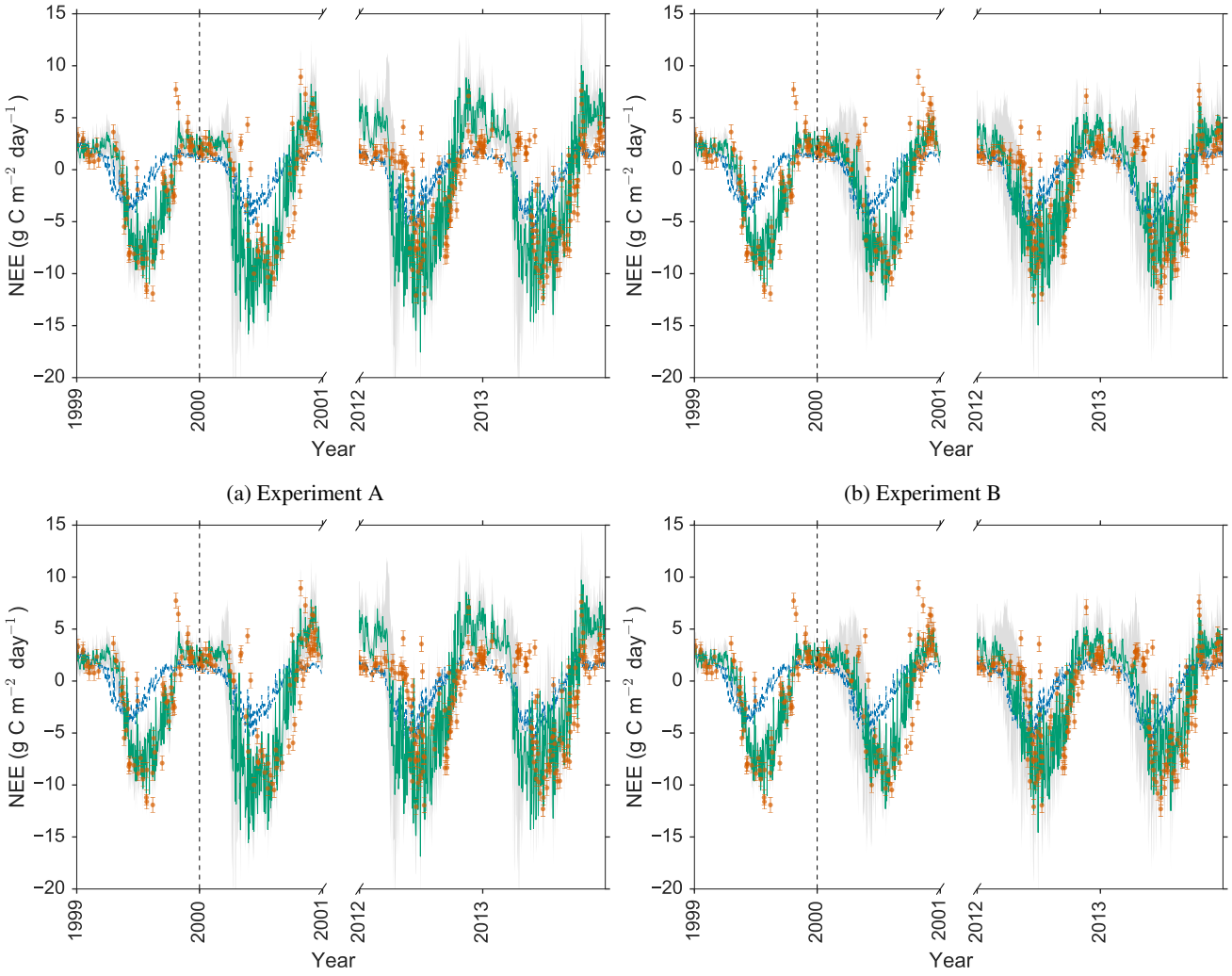
(d) Experiment D

Figure 7: One year assimilation and fourteen year forecast of Alice Holt NEE with DALEC2, blue dotted line: background model trajectory, green line: analysis and forecast after assimilation, grey shading: Error in model after assimilation (+/- 3 standard deviations), red dots: observations from Alice Holt flux site with error bars.

Daley, 1992), meaning that less frequently sampled data streams can have more impact on the assimilation.

3.5. Experiment D

In the final experiment we use $\mathbf{B}_{corr}$ and $\hat{\mathbf{R}}_{corr}$ in the assimilation. Figure 8d, figure 8b and 9a shows that using both correlated matrices gives similar results as experiment B when $\mathbf{B}_{corr}$ is used with $\hat{\mathbf{R}}_{diag}$. However using $\hat{\mathbf{R}}_{corr}$ in addition to $\mathbf{B}_{corr}$ provides similar improvements as in experiment C. From table 2 and figure 10e we see the forecast RMSE is slightly reduced from results in experiment B to 2.38 gCm^{-2}day^{-1}. Using both matrices appears to combine the beneficial effects described in both section 3.3 and section 3.4. In figure 11 we can see that the analysis increment is very similar to experiment B.



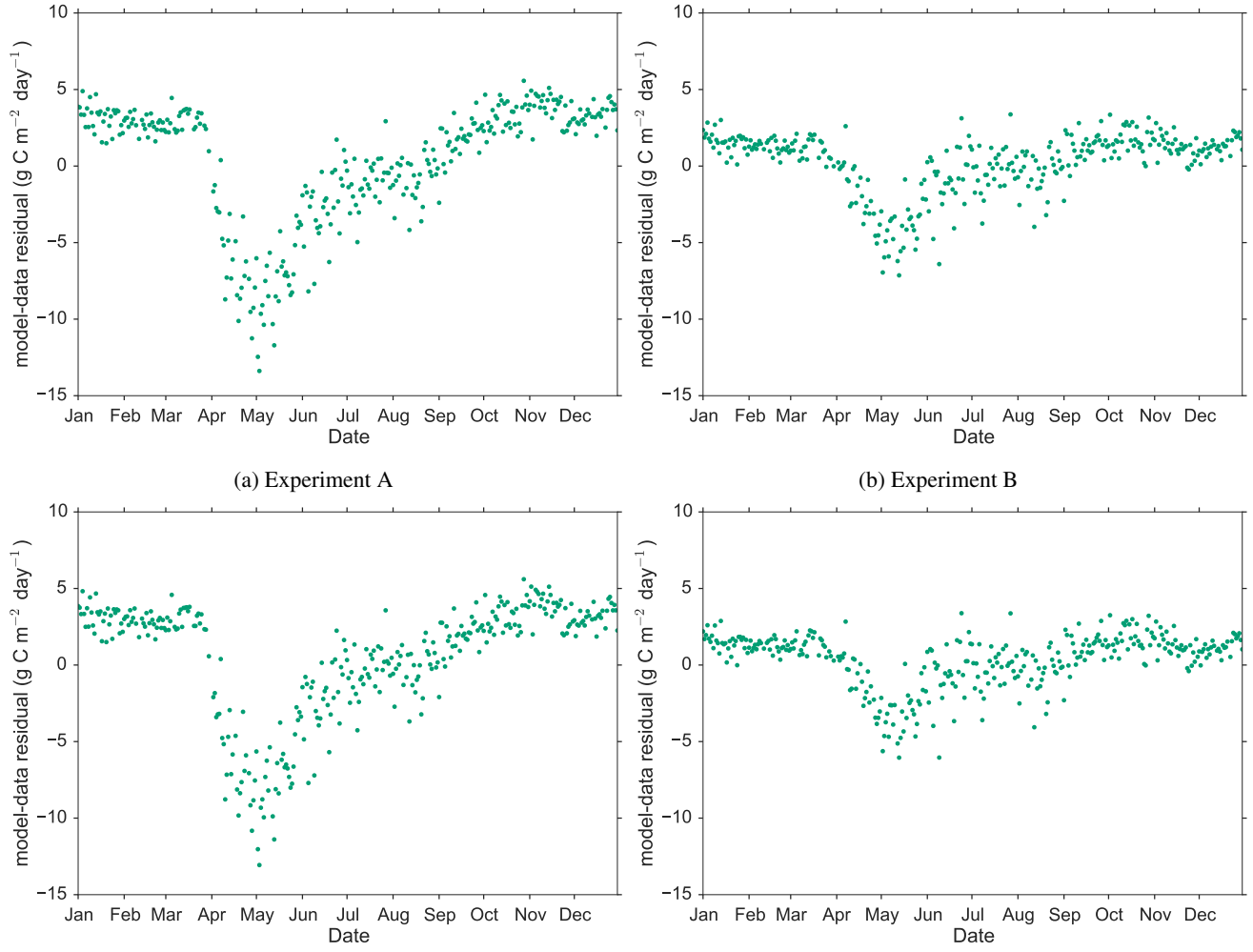
(a) Experiment A

(b) Experiment B

(c) Experiment C

(d) Experiment D

Figure 8: As figure 7 but only showing the first and final two years results from the one year assimilation and fourteen year forecast of Alice Holt NEE with DALEC2, blue dotted line: background model trajectory, green line: analysis and forecast after assimilation, grey shading: Error in model after assimilation (+/- 3 standard deviations), red dots: observations from Alice Holt flux site with error bars.



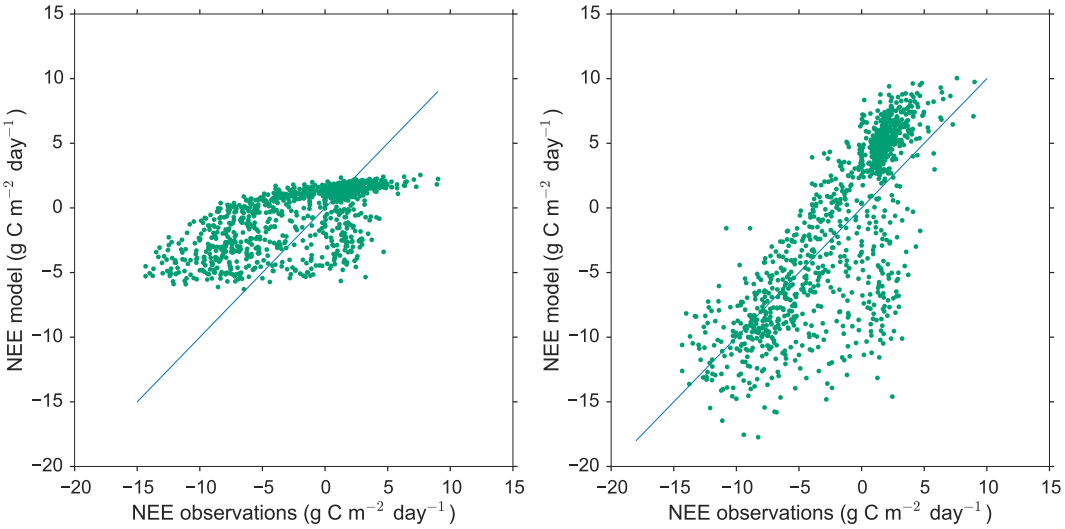
(a) Experiment A

(b) Experiment B

(c) Experiment C

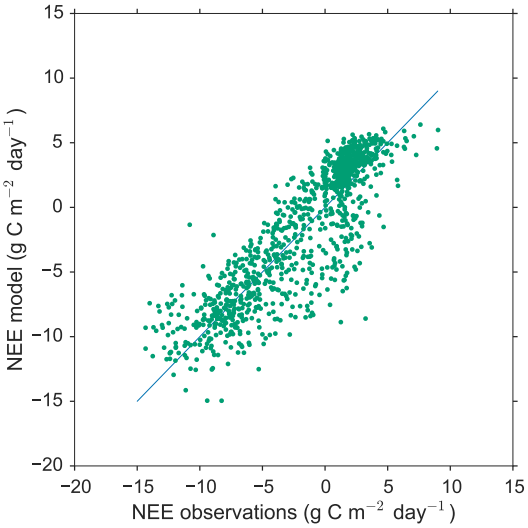
(d) Experiment D

Figure 9: Net ecosystem exchange model-data differences for the four experiments. Here each point corresponds to the mean model-data difference for that day of the year over the 14 year model forecast (Jan 2000 - Dec 2013).

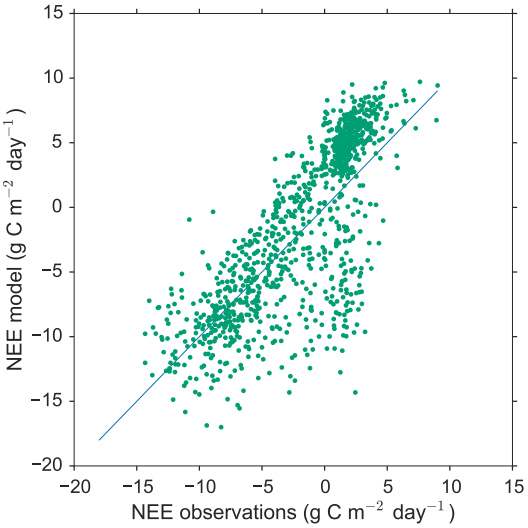


(a) Background

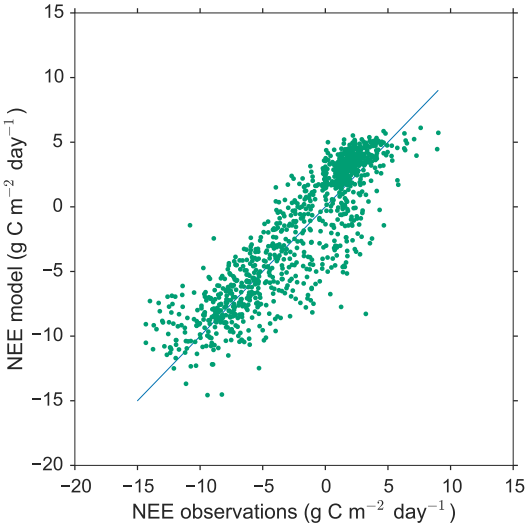
(b) Experiment A



(c) Experiment B



(d) Experiment C



(e) Experiment D

Figure 10: Forecast scatter plot of modelled daily NEE vs. observations for Jan 2000 - Dec 2013 (green dots). Blue line represents the 1-1 line.

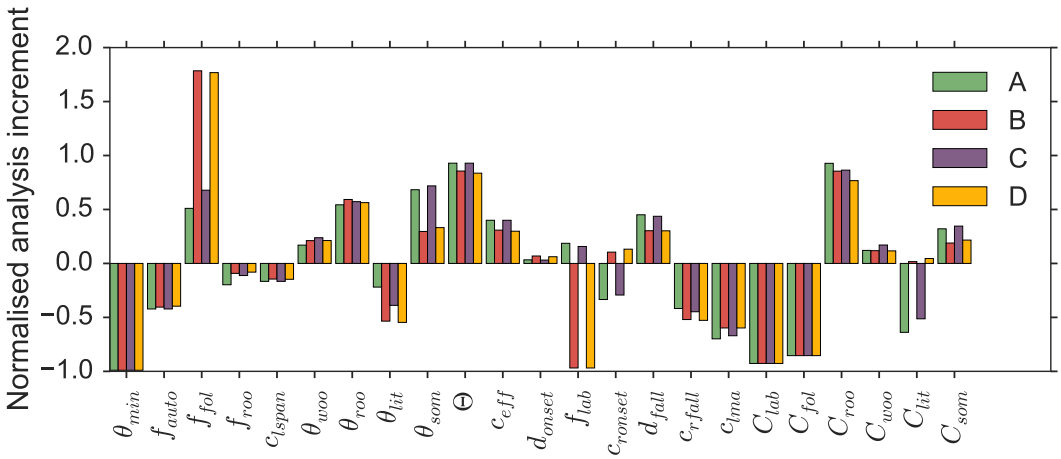
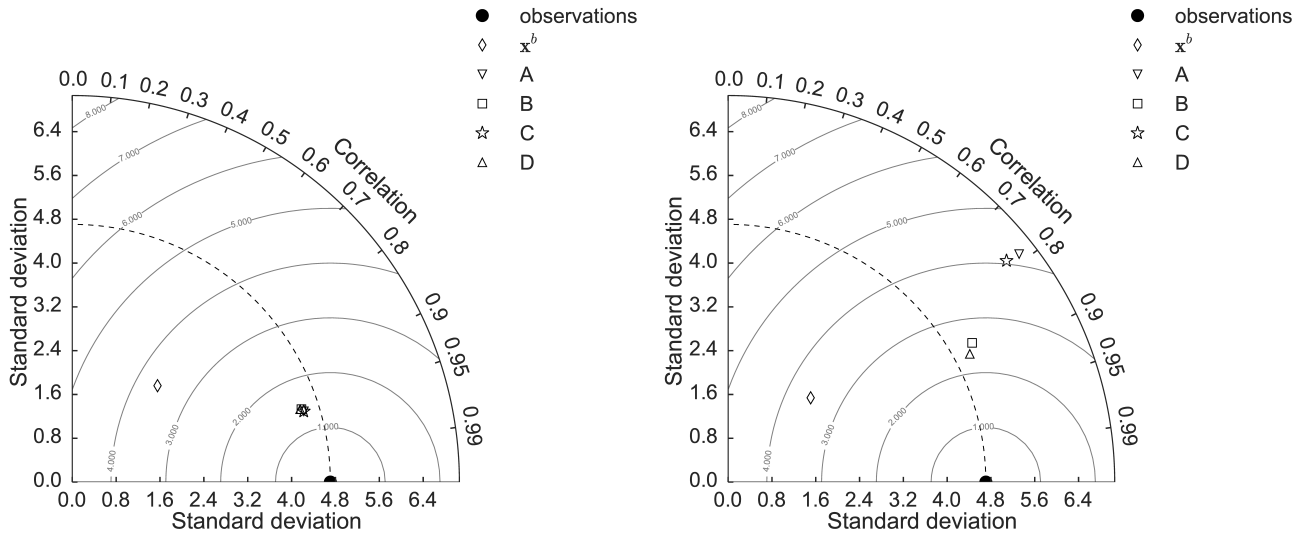


Figure 11: Normalised analysis increment $\left(\frac{(\mathbf{x}^a-\mathbf{x}^b)}{\mathbf{x}^b}\right)$ for the four experiments. Explanation of parameter and state variable symbols in table 3.

3.6. Summary

In our experiments we have shown that both $\mathbf{B}_{corr}$ and $\hat{\mathbf{R}}_{corr}$ have the effect of improving the model forecast of NEE. As it can be difficult to inspect the skill of a certain model by only plotting model trajectories, in figure 12 we show Taylor diagrams displaying a statistical comparison of the four experiment and background analysis (Jan 1999 - Dec 1999) and forecast (Jan 2000 - Dec 2013) results with the observations of NEE. Here the radial distances from the origin to the points are proportional to the standard deviations of the observations and modelled observations and the azimuthal positions give the correlation coefficient between the modelled and observed NEE (Taylor, 2001). If a model predicted the observations perfectly it would have a correlation coefficient of 1 and a radial distance matching that of the observations (represented by the dotted line). Figure 12a shows that all the experiments give very similar results in the analysis window (Jan 1999 - Dec 1999) with all the experiment points closely grouped on top of each other, whereas figure 12b shows the significant difference between the experiment results in the forecast (Jan 2000 - Dec 2013), with experiments B and D being closer to the dotted line. In all our experiments we find that θ_{min}, C_{lab} and C_{fol} reach the bounds after assimilation. In the case of θ_{min} this is most likely due to the fact that we do not have enough information to recover this parameter when only assimilating observations of NEE, as the DALEC model prediction of NEE is insensitive to variations in this parameter (Chuter, 2013). Assimilating more distinct data streams could help avoid this edge-hitting behaviour. For C_{lab} and C_{fol} this could suggest a flaw in the model or the fact that the prescribed bounds need to be relaxed slightly for the studied ecosystem. Our hypothesis is that the mechanism by which C_{lab} is distributed to the leaves is over simplified; we intend to test this in future work. In table 4 we show the standard deviations for our parameter and state variables after assimilation. We can see that we have improved our confidence for most of these variables after assimilation when compared with the standard deviations in table 3.



(a) Analysis (Jan 1999 - Dec 1999)

(b) Forecast (Jan 2000 - Dec 2013)

Figure 12: Taylor diagrams displaying statistical comparison of the four experiment and background analysis (Jan 1999 - Dec 1999) and forecast (Jan 2000 - Dec 2013) results with observations of NEE ($\mathrm{gCm^{-2}day^{-1}}$). The dotted line represents the standard deviation of the observations and the contours represent values of constant root mean square error between model and observations.

4. Discussion

In this paper we have implemented the DALEC2 functional ecology model in a 4D-Var data assimilation scheme, building an adjoint of the DALEC2 model and applying rigorous tests to our scheme. Using 4D-Var can provide much faster assimilation results than MCMC techniques as we have knowledge of the derivative of the model. For our experiments the 4D-Var routine has taken in the order of 10^2 function evaluations to converge to a minimum, whereas MCMC techniques using the same model take in the order of 10^8 function evaluations (Bloom and Williams, 2015). However, we do assume that the statistics of the problem are Gaussian whereas MCMC techniques do not. We have shown that 4D-Var is a valid tool for improving the DALEC2 model estimate of NEE and that even when assimilating only a single year of NEE observations we can improve the forecast significantly. If more than one year was required, this type of data assimilation routine could be run in cycling mode, allowing for the assimilation of multiple years of data (Moodey et al., 2013). This also avoids any possible unstable behaviour associated with much longer single assimilation windows. However, here our aim is to investigate the effect of specifying correlations in background and observation error statistics on the forecast of NEE. We have therefore assimilated just one year of NEE observations and produced a long 14 year forecast in order to see more clearly the effect of including these correlations on the forecast when judging against observations. The observations of daily NEE from the Alice Holt flux site are quite variable year to year, peak summer uptake varies from -14.35 $\mathrm{gCm^{-2}day^{-1}}$ to -9.04 $\mathrm{gCm^{-2}day^{-1}}$, and therefore provide a reasonable test for the ability of the DALEC2 model forecast, especially over a 14 year period.

We then considered the nature of background and observation errors. The effect of specifying parameter-state correlations in the background information and serial correlations between the observation errors was explored.

The technique presented here to specify $\mathbf{B}_{corr}$ has been shown to have significantly improved forecasts of NEE over using

Analysis (Jan 1999 - Dec 1999)			
Experiment	RMSE ($gCm^{-2}day^{-1}$)	Bias ($gCm^{-2}day^{-1}$)	Correlation coefficient
Background	3.86	-1.60	0.70
A	1.36	-0.03	0.96
B	1.42	-0.04	0.95
C	1.37	-0.09	0.96
D	1.43	-0.09	0.95
Forecast (Jan 2000 - Dec 2013)			
Experiment	RMSE ($gCm^{-2}day^{-1}$)	Bias ($gCm^{-2}day^{-1}$)	Correlation coefficient
Background	3.86	-1.36	0.66
A	4.22	-0.30	0.79
B	2.56	-0.20	0.87
C	4.09	-0.51	0.78
D	2.38	-0.33	0.88

Table 2: Analysis (Jan 1999 - Dec 1999) and forecast (Jan 2000 - Dec 2013) results for experiments and background when judged against observed NEE.

a diagonal representation of $\mathbf{B}$. In section 3.3 we discuss how the correlations in $\mathbf{B}_{corr}$ impact the analysis update for the parameter and state variables (see figure 11) causing the seasonal cycle of carbon uptake and magnitude of fluxes to fit more closely with the observations than when using a diagonal $\mathbf{B}$ in the assimilation algorithm. These results agree with those of Smith et al. (2009) where the importance of specifying parameter-state correlations when performing joint parameter and state estimation with variational data assimilation was shown. The added constraint provided by including correlations in the prior error covariance matrix, $\mathbf{B}$, acts to regularise the assimilation problem. Hence, the parameter and initial state values we retrieve from our data assimilation are more likely to be realistic, leading to better insight into the studied system. For example we see from figure 11 that when using $\mathbf{B}_{corr}$ in our assimilation we find a much longer labile release period (c_{ronset}) than when using $\mathbf{B}_{diag}$. This means that the period of green-up in our study site is possibly much longer than we would have estimated had we based our analysis on our assimilation results using a matrix $\mathbf{B}$ with no correlations. The method for specifying $\mathbf{B}_{corr}$ in this paper used a series of ecological dynamical constraints taken from Bloom and Williams (2015). Implementing correlations in the prior error statistics in this way may prove difficult for models where these type of constraints are not available; however there are other methods to build correlations into $\mathbf{B}$. One technique we also tested (not presented here) to create a correlated $\mathbf{B}$ involved evolving an ensemble of state vectors over the length of the chosen assimilation window using the model (DALEC2) and then taking the covariance of the evolved ensemble. This gave us a $\mathbf{B}$ with parameter-state and state-state correlations, but no parameter-parameter correlations as the parameters are not updated by the model. Using the $\mathbf{B}$ created with this method also improved assimilation results significantly over using a diagonal $\mathbf{B}$. A larger number of different tests were run using different background vectors and variances and it was found that specifying some form of

correlation structure in $\mathbf{B}$ always made an improvement to the results of the assimilation. As this work has only considered a single deciduous site, it would be useful to apply the techniques detailed here for an evergreen site. Evergreen ecosystems usually have less extreme seasonal variation, it will therefore be of interest to see if a similar improvement for evergreen ecosystem forecast results is found when using a $\mathbf{B}_{corr}$ created using the same method.

The purpose of this exercise was to see how well we could forecast NEE while also investigating the effect of including correlations between error statistics. It was not an attempt to recover all the parameters and state variables with a high level of accuracy. However, it is still instructive to look at these values and compare with data where available. In Meir et al. (2002) an observed range is given for leaf mass per area (c_{lma}) for the Alice Holt flux site of between 40 to 80 gCm^{-2}. The background value for c_{lma} in our experiments is 128.5 gCm^{-2}. When using diagonal error covariance matrices in experiment A we find a value of 38.7 gCm^{-2} for c_{lma} after assimilation which is almost within the range given by Meir et al. (2002). In experiment D when using error covariance matrices including correlations c_{lma} has a value of 51.6 gCm^{-2} after assimilation, this is well within the observed range given by Meir et al. (2002). From observations made by Forest Research we also have estimates of the above and below ground woody carbon pool (C_{woo}) at the start of 1999, with an observed value of 14258 gCm^{-2}. It is not clear how uncertain this estimate is. The background value for C_{woo} in our experiments is 6506 gCm^{-2}. When using diagonal error covariance matrices in experiment A we find a value of 7291 gCm^{-2} for C_{woo}, an increase but still far away from the observed estimate. In experiment D when using error covariance matrices including correlations C_{woo} has a value of 7262 gCm^{-2} a similar result as experiment A. Here the assimilation has not been able to recover a value of C_{woo} similar to that of the observed estimate. This is not necessarily of concern as we are not able to quantify the error in this observation. Also we are assimilating observations of daily NEE only; NEE is the difference between Gross Primary Productivity (GPP) and Total ecosystem respiration (RT), (NEE = RT - GPP), with neither GPP nor RT being direct functions of C_{woo}. Therefore it is unlikely that we will recover an accurate value of C_{woo}, as the assimilated observations are not significantly impacted by large changes in this state variable; this result is also discussed in Fox et al. (2009). This may also explains why we are able to recover a reasonable value of c_{lma} from the assimilation, as from equation (1) we can see that c_{lma} is one of the input arguments taken by GPP. The function calculating NEE will therefore be sensitive to variations in the c_{lma} parameter and so assimilating observations of NEE could help to constrain this parameter.

In numerical weather prediction it has been shown that including correlations in $\mathbf{R}$ can help improve data assimilation results (Weston et al., 2014; Stewart et al., 2013). However the specified correlations have most commonly been satellite interchannel correlations with observations errors still being considered independent in time. In this paper we have shown that including correlations between observation errors in time can also improve data assimilation results, here providing a slight improvement for the DALEC2 model forecast of NEE. Here we only see a small impact on our results when using $\hat{\mathbf{R}}_{corr}$ in the assimilation as the correlations we have included are weak (especially in comparison to those included in $\mathbf{B}_{corr}$) and on a short time-scale. We expect including correlations in $\hat{\mathbf{R}}$ will have more of an impact on data assimilation results when assimilating data with stronger error correlations (i.e. finer temporal-resolution observations). We also expect including these serial correlations to

have an even greater impact when assimilating more than one data stream as discussed in section 1. Using the form of $\hat{\mathbf{R}}$ given in this paper for specifying serial correlations will also allow us to specify serial correlations between different observation types. When running the DALEC2 model with a day-night time step, instead of the daily time step used for this paper, this will allow us to build in the type of correlations investigated by Baldocchi et al. (2015) between ecosystem respiration and canopy photosynthesis. More work is needed to investigate the effect of including correlations between observations error statistics when assimilating multiple data streams.

The $\hat{\mathbf{R}}_{corr}$ presented in this paper has a weak correlation ($a = 0.3$ as shown in section2.6) in time between observations of NEE, this representation of $\hat{\mathbf{R}}_{corr}$ has slightly improved the model forecast of NEE. However other choices of $\hat{\mathbf{R}}_{corr}$ (with much stronger correlations between observations) tested for this paper degraded the forecast. This is probably due to the specified correlations being unrealistic and highlights the fact that a reasonable estimate of the true correlation structure for $\hat{\mathbf{R}}_{corr}$ is needed to have a positive impact on results. The development of a more diagnostic approach for the calculation of serial correlations in $\hat{\mathbf{R}}$ would be useful. One option would be to adapt the Desroziers et al. (2005) diagnostic, which has been used successfully in numerical weather prediction for diagnosing observation error correlations for observations taken at the same time (Weston et al., 2014), and extending this technique to diagnose serial correlations.

5. Conclusion

Functional ecology and land surface model data assimilation routines largely treat prior estimates of parameter and state uncertainties and observation error statistics as independent and uncorrelated. In this paper we have shown the importance of including estimates of such correlations, especially between background parameter and state error statistics when performing joint parameter and state estimation.

When performing joint parameter and state estimation including correlations in the background error covariance matrix significantly improves the forecast after assimilation, in comparison to using a diagonal representation of $\mathbf{B}$. Specifying serial time correlations between observation errors in $\hat{\mathbf{R}}$ also improves the forecast and we expect these correlations to have a greater impact when assimilating more than one data stream. More work is needed to investigate the effect of including these correlations when assimilating multiple data streams. The development of a more diagnostic tool for the calculation of the error correlation structure in $\hat{\mathbf{R}}$ is also important.

When including both parameter-state correlations in $\mathbf{B}$ and time correlations between observation errors in $\hat{\mathbf{R}}$ and assimilating only a single year of NEE observations we can forecast 14 years of NEE observations with a root-mean square error of 2.38 $\mathrm{gCm^{-2}day^{-1}}$ and a correlation coefficient of 0.88. This is a significant 44% reduction in error from the results when using a $\mathbf{B}$ and $\hat{\mathbf{R}}$ with no specified correlations of 4.22 $\mathrm{gCm^{-2}day^{-1}}$ and a correlation coefficient of 0.79.

6. Acknowledgements

This work was funded by the UK Natural Environment Research Council (NE/K00705X/1) with a CASE award from the UK Forestry Commission. This work was also partly funded by the National Centre for Earth Observation. We are grateful to Luke Smallman for providing the data from the CARDAMOM system.

References

Bacour, C., Peylin, P., MacBean, N., Rayner, P.J., Delage, F., Chevallier, F., Weiss, M., Demarty, J., Santaren, D., Baret, F., Berveiller, D., Dufrêne, E., Prunet, P., 2015. Joint assimilation of eddy-covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model. Journal of Geophysical Research: Biogeosciences , n/a–n/adoi:10.1002/2015JG002966.

Baldocchi, D., 2008. Turner review no. 15.'breathing'of the terrestrial biosphere: lessons learned from a global network of carbon dioxide flux measurement systems. Australian Journal of Botany 56, 1–26.

Baldocchi, D., Sturtevant, C., Contributors, F., 2015. Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration? Agricultural and Forest Meteorology 207, 117–126. doi:10.1016/j.agrformet.2015.03.010.

Bannister, R.N., 2008. A review of forecast error covariance statistics in atmospheric variational data assimilation. i: Characteristics and measurements of forecast error covariances. Quarterly Journal of the Royal Meteorological Society 134, 1951–1970.

Bauer, P., Thorpe, A., Brunet, G., 2015. The quiet revolution of numerical weather prediction. Nature 525, 47–55.

Bloom, A.A., Williams, M., 2015. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a modeldata fusion framework. Biogeosciences 12, 1299–1315. doi:10.5194/bg-12-1299-2015.

Bonavita, M., Hlm, E., Isaksen, L., Fisher, M., 2015. The evolution of the ecmwf hybrid data assimilation system. Quarterly Journal of the Royal Meteorological Society , n/a–n/adoi:10.1002/qj.2652.

Braswell, B.H., Sacks, W.J., Linder, E., Schimel, D.S., 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. Global Change Biology 11, 335–355.

Chuter, A.M., 2013. A Qualitative Analysis of the Data Assimilation Linked Ecosystem Carbon Model , DALEC. Ph.D. thesis. University of Surry, Guildford.

Ciais, P., Sabine, C., Bala, G., Bopp, L., Brovkin, V., Canadell, J., Chhabra, A., DeFries, R., Galloway, J., Heimann, M., et al., 2014. Carbon and other biogeochemical cycles, in: Climate change 2013: the physical science basis. Contribution of

Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, pp. 465–570.

Clayton, A.M., Lorenc, A.C., Barker, D.M., 2013. Operational implementation of a hybrid ensemble/4d-var global data assimilation system at the met office. Quarterly Journal of the Royal Meteorological Society 139, 1445–1461. doi:10.1002/qj.2054.

Daley, R., 1992. The Effect of Serially Correlated Observation and Model Error on Atmospheric Data Assimilation. doi:10.1175/1520-0493(1992)120<0164:TEOSCO>2.0.CO;2.

Dee, D., Uppala, S., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, P., et al., 2011. The era-interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137, 553–597.

Delahaies, S., Roulstone, I., Nichols, N., 2013. A regularization of the carbon cycle data-fusion problem, in: EGU General Assembly Conference Abstracts, p. 4087.

Desroziers, G., Berre, L., Chapnik, B., Poli, P., 2005. Diagnosis of observation, background and analysis-error statistics in observation space. Quarterly Journal of the Royal Meteorological Society 131, 3385–3396.

Exbrayat, J.f., Smallman, T.L., Bloom, A.A., Williams, M., 2015. Using a data-assimilation system to assess the influence of fire on simulated carbon fluxes and plant traits for the Australian continent. EGU General Assembly 17, 6421.

Fox, A., Williams, M., Richardson, A.D., Cameron, D., Gove, J.H., Quaife, T., Ricciuto, D., Reichstein, M., Tomelleri, E., Trudinger, C.M., et al., 2009. The reflex project: comparing different algorithms and implementations for the inversion of a terrestrial ecosystem model against eddy covariance data. Agricultural and Forest Meteorology 149, 1597–1615.

Freitag, M.A., Nichols, N.K., Budd, C.J., 2010. L1-regularisation for ill-posed problems in variational data assimilation. Pamm 10, 665–668. doi:10.1002/pamm.201010324.

Healy, S., White, A., 2005. Use of discrete Fourier transforms in the 1D-Var retrieval problem. Quarterly Journal of the Royal Meteorological Society 131, 63–72. doi:10.1256/qj.03.193.

Järvinen, H., Andersson, E., Bouttier, F., 1999. Variational assimilation of time sequences of surface observations with serially correlated errors. Tellus A 51, 469–488.

Jones, E., Oliphant, T., Peterson, P., et al., 2001. SciPy: Open source scientific tools for Python. URL: http://www.scipy.org/. [Online; accessed 2015-12-04].

Kalnay, E., 2003. Atmospheric modeling, data assimilation, and predictability. Cambridge university press.

Kaminski, T., Knorr, W., Schürmann, G., Scholze, M., Rayner, P.J., Zaehle, S., Blessing, S., Dorigo, W., Gayler, V., Giering, R., Gobron, N., Grant, J.P., Heimann, M., Hooker-Stroud, a., Houweling, S., Kato, T., Kattge, J., Kelley, D., Kemp, S., Koffi, E.N., Köstler, C., Mathieu, P.P., Pinty, B., Reick, C.H., Rödenbeck, C., Schnur, R., Scipal, K., Sebald, C., Stacke, T., Van Scheltinga, a.T., Vossbeck, M., Widmann, H., Ziehn, T., 2013. The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges. Journal of Geophysical Research: Biogeosciences 118, 1414–1426. doi:10.1002/jgrg.20118.

Krinner, G., Viovy, N., de Noblet-Ducoudré, N., Ogée, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., Prentice, I.C., 2005. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. Global Biogeochemical Cycles 19, 1–33. doi:10.1029/2003GB002199.

Lasslop, G., Reichstein, M., Kattge, J., Papale, D., 2008. Influences of observation errors in eddy flux data on inverse model parameter estimation. Biogeosciences Discussions 5, 751–785.

Lawless, A.S., 2013. Variational data assimilation for very large environmental problems., in: Cullen, M.J.P., Freitag, M.A., Kindermann, S., Scheichl, R. (Eds.), Large scale Inverse Problems: Computational Methods and Applications in the Earth Sciences, Radon series on Computational and Applied Mathematics, De Gruyter. pp. 55–90.

Li, Y., Navon, I.M., Yang, W., Zou, X., Bates, J.R., Moorthi, S., Higgins, R.W., 1994. Four-Dimensional Variational Data Assimilation Experiments with a Multilevel Semi-Lagrangian Semi-Implicit General Circulation Model. Monthly Weather Review 122, 966–983. doi:10.1175/1520-0493(1994)122<0966:FDVDAE>2.0.CO;2.

LI-COR, Inc., 2015. EddyPro 6 Help and User's Guide. LI-COR, Inc. Lincoln, NE.

Lorenc, A.C., Rawlins, F., 2005. Why does 4d-var beat 3d-var? Quarterly Journal of the Royal Meteorological Society 131, 3247–3257.

Luo, Y., Keenan, T.F., Smith, M., 2015. Predictability of the terrestrial carbon cycle. Global change biology 21, 1737–1751.

Meir, P., Kruijt, B., Broadmeadow, M., Barbosa, E., Kull, O., Carswell, F., Nobre, A., Jarvis, P., 2002. Acclimation of photosynthetic capacity to irradiance in tree canopies in relation to leaf nitrogen concentration and leaf mass per unit area. Plant, Cell & Environment 25, 343–357.

Moodey, A.J.F., Lawless, A.S., Potthast, R.W.E., van Leeuwen, P.J., 2013. Nonlinear error dynamics for cycled data assimilation methods. Inverse Problems 29, 025002.

Navon, I., 1998. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. Dynamics of Atmospheres and Oceans 27, 55–79.

Navon, I.M., Zou, X., Derber, J., Sela, J., 1992. Variational Data Assimilation with an Adiabatic Version of the NMC Spectral Model. Monthly Weather Review 120, 1433–1446. doi:10.1175/1520-0493(1992)120<1433:VDAWAA>2.0.CO;2.

Niu, S., Luo, Y., Dietze, M.C., Keenan, T.F., Shi, Z., Li, J., Iii, F.S.C., 2014. The role of data assimilation in predictive ecology. Ecosphere 5, art65. doi:10.1890/ES13-00273.1.

Nocedal, J., Wright, S.J., 1999. Numerical Optimization. Springer Science & Business Media.

Papale, D., Reichstein, M., Aubinet, M., Canfora, E., Bernhofer, C., Kutsch, W., Longdoz, B., Rambal, S., Valentini, R., Vesala, T., Yakir, D., 2006. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. Biogeosciences 3, 571–583. doi:10.5194/bg-3-571-2006.

Quaife, T., Lewis, P., De Kauwe, M., Williams, M., Law, B.E., Disney, M., Bowyer, P., 2008. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. Remote Sensing of Environment 112, 1347–1364. doi:10.1016/j.rse.2007.05.020.

Renaud, J., 1997. Automatic differentiation in robust optimization. AIAA journal 35, 1072–1079.

Richardson, A.D., Mahecha, M.D., Falge, E., Kattge, J., Moffat, A.M., Papale, D., Reichstein, M., Stauch, V.J., Braswell, B.H., Churkina, G., Kruijt, B., Hollinger, D.Y., 2008. Statistical properties of random {CO2} flux measurement uncertainty inferred from model residuals. Agricultural and Forest Meteorology 148, 38 – 50. doi:http://dx.doi.org/10.1016/j.agrformet.2007.09.001.

Richardson, A.D., Williams, M., Hollinger, D.Y., Moore, D.J., Dail, D.B., Davidson, E.A., Scott, N.A., Evans, R.S., Hughes, H., Lee, J.T., et al., 2010. Estimating parameters of a forest ecosystem c model with measurements of stocks and fluxes as joint constraints. Oecologia 164, 25–40.

Smith, P.J., Dance, S.L., Baines, M.J., Nichols, N.K., Scott, T.R., 2009. Variational data assimilation for parameter estimation: application to a simple morphodynamic model. Ocean Dynamics 59, 697–708.

Smith, P.J., Dance, S.L., Nichols, N.K., 2011. A hybrid data assimilation scheme for model parameter estimation: Application to morphodynamic modelling. Computers & Fluids 46, 436 – 441. doi:http://dx.doi.org/10.1016/j.compfluid.2011.01.010. 10th {ICFD} Conference Series on Numerical Methods for Fluid Dynamics (ICFD 2010).

Stewart, L.M., Dance, S.L., Nichols, N.K., 2013. Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. Tellus, Series A: Dynamic Meteorology and Oceanography 65, 1–14. doi:10.3402/tellusa.v65i0.19546.

Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research 106, 7183. doi:10.1029/2000JD900719.

Tremolet, Y., 2006. Accounting for an imperfect model in 4D-Var. Quarterly Journal of the Royal Meteorological Society 132, 2483–2504. URL: http://doi.wiley.com/10.1256/qj.05.224, doi:10.1256/qj.05.224.

Verbeeck, H., Peylin, P., Bacour, C., Bonal, D., Steppe, K., Ciais, P., 2011. Seasonal patterns of CO2 fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. Journal of Geophysical Research 116, G02018. doi:10.1029/2010JG001544.

Waller, J.A., Dance, S.L., Lawless, A.S., Nichols, N.K., Eyre, J.R., 2014. Representativity error for temperature and humidity using the Met Office high-resolution model. Quarterly Journal of the Royal Meteorological Society 140, 1189–1197. doi:10.1002/qj.2207.

Walter, S.F., Lehmann, L., 2013. Algorithmic differentiation in Python with AlgoPy. Journal of Computational Science 4, 334–344. doi:10.1016/j.jocs.2011.10.007.

Weston, P., Bell, W., Eyre, J., 2014. Accounting for correlated error in the assimilation of high-resolution sounder data. Quarterly Journal of the Royal Meteorological Society 140, 2420–2429.

Wilkinson, M., Eaton, E., Broadmeadow, M., Morison, J., 2012. Inter-annual variation of carbon uptake by a plantation oak woodland in south-eastern england. Biogeosciences 9, 5373–5389.

Williams, M., Rastetter, E.B., Fernandes, D.N., Goulden, M.L., Shaver, G.R., Johnson, L.C., 1997. Predicting gross primary productivity in terrestrial ecosystems. Ecological Applications 7, 882–894.

Williams, M., Schwarz, P.A., Law, B.E., Irvine, J., Kurpius, M.R., 2005. An improved analysis of forest carbon dynamics using data assimilation. Global Change Biology 11, 89–105.

Zobitz, J., Desai, A., Moore, D., Chadwick, M., 2011. A primer for data assimilation with ecological models using markov chain monte carlo (mcmc). Oecologia 167, 599–611.

Zobitz, J.M., Moore, D.J.P., Quaife, T., Braswell, B.H., Bergeson, A., Anthony, J.a., Monson, R.K., 2014. Joint data assimilation of satellite reflectance and net ecosystem exchange data constrains ecosystem carbon fluxes at a high-elevation subalpine forest. Agricultural and Forest Meteorology 195-196, 73–88. doi:10.1016/j.agrformet.2014.04.011.

Appendix

Parameter	Description	Background vector ($\mathbf{x}^b$)	Standard deviation	Range
θ_{min}	Litter mineralisation rate (day^{-1})	9.810×10^{-4}	2.030×10^{-3}	$10^{-5} - 10^{-2}$
f_{auto}	Autotrophic respiration fraction	5.190×10^{-1}	1.168×10^{-1}	$0.3 - 0.7$
f_{fol}	Fraction of GPP allocated to foliage	1.086×10^{-1}	1.116×10^{-1}	$0.01 - 0.5$
f_{roo}	Fraction of GPP allocated to fine roots	4.844×10^{-1}	2.989×10^{-1}	$0.01 - 0.5$
c_{lspan}	Determines annual leaf loss fraction	1.200×10^{0}	1.161×10^{-1}	$1.0001 - 10$
θ_{woo}	Woody carbon turnover rate (day^{-1})	1.013×10^{-4}	1.365×10^{-4}	$2.5 \times 10^{-5} - 10^{-3}$
θ_{roo}	Fine root carbon turnover rate (day^{-1})	3.225×10^{-3}	2.930×10^{-3}	$10^{-4} - 10^{-2}$
θ_{lit}	Litter carbon turnover rate (day^{-1})	3.442×10^{-3}	3.117×10^{-3}	$10^{-4} - 10^{-2}$
θ_{som}	Soil and organic carbon turnover rate (day^{-1})	1.113×10^{-4}	1.181×10^{-4}	$10^{-7} - 10^{-3}$
Θ	Temperature dependance exponent factor	4.147×10^{-2}	1.623×10^{-2}	$0.018 - 0.08$
c_{eff}	Canopy efficiency parameter	7.144×10^{1}	2.042×10^{1}	$10 - 100$
d_{onset}	Leaf onset day (day)	1.158×10^{2}	6.257×10^{0}	$1 - 365$
f_{lab}	Fraction of GPP allocated to labile carbon pool	3.204×10^{-1}	1.145×10^{-1}	$0.01 - 0.5$
c_{ronset}	Labile carbon release period (days)	4.134×10^{1}	1.405×10^{1}	$10 - 100$
d_{fall}	Leaf fall day (day)	2.205×10^{2}	3.724×10^{1}	$1 - 365$
c_{rfall}	Leaf-fall period (days)	1.168×10^{2}	2.259×10^{1}	$10 - 100$
c_{lma}	Leaf mass per area (gCm^{-2})	1.285×10^{2}	6.410×10^{1}	$10 - 400$
C_{lab}	Labile carbon pool (gCm^{-2})	1.365×10^{2}	6.626×10^{1}	$10 - 1000$
C_{fol}	Foliar carbon pool (gCm^{-2})	6.864×10^{1}	3.590×10^{1}	$10 - 1000$
C_{roo}	Fine root carbon pool (gCm^{-2})	2.838×10^{2}	2.193×10^{2}	$10 - 1000$
C_{woo}	Above and below ground woody carbon pool (gCm^{-2})	6.506×10^{3}	7.143×10^{3}	$100 - 10^{5}$
C_{lit}	Litter carbon pool (gCm^{-2})	5.988×10^{2}	5.450×10^{2}	$10 - 1000$
C_{som}	Soil and organic carbon pool (gCm^{-2})	1.936×10^{3}	1.276×10^{3}	$100 - 2 \times 10^{5}$

Table 3: Parameter values and standard deviations for background vector used in experiments.

Parameter	A	B	C	D
θ_{min}	1.822×10^{-6}	3.742×10^{-7}	1.519×10^{-6}	3.854×10^{-7}
f_{auto}	2.913×10^{-3}	1.428×10^{-3}	2.937×10^{-3}	1.510×10^{-3}
f_{fol}	5.459×10^{-3}	4.581×10^{-3}	6.797×10^{-3}	4.591×10^{-3}
f_{roo}	7.907×10^{-2}	9.141×10^{-3}	8.199×10^{-2}	9.149×10^{-3}
c_{lspan}	4.884×10^{-7}	5.894×10^{-4}	5.304×10^{-7}	5.469×10^{-4}
θ_{woo}	1.849×10^{-8}	8.365×10^{-9}	1.849×10^{-8}	8.365×10^{-9}
θ_{roo}	6.870×10^{-6}	3.494×10^{-6}	7.326×10^{-6}	3.508×10^{-6}
θ_{lit}	3.144×10^{-6}	4.808×10^{-7}	2.242×10^{-6}	4.635×10^{-7}
θ_{som}	1.178×10^{-8}	6.848×10^{-9}	1.210×10^{-8}	6.850×10^{-9}
Θ	7.905×10^{-5}	6.808×10^{-5}	8.010×10^{-5}	6.978×10^{-5}
c_{eff}	3.755×10^{2}	2.625×10^{2}	3.724×10^{2}	2.608×10^{2}
d_{onset}	3.552×10^{1}	3.755×10^{1}	3.649×10^{1}	3.766×10^{1}
f_{lab}	1.220×10^{-2}	3.209×10^{-3}	1.225×10^{-2}	3.203×10^{-3}
c_{ronset}	8.304×10^{1}	1.642×10^{2}	1.100×10^{2}	1.644×10^{2}
d_{fall}	5.992×10^{2}	5.294×10^{1}	5.772×10^{2}	6.145×10^{1}
c_{rfall}	1.540×10^{2}	1.521×10^{2}	1.604×10^{2}	1.599×10^{2}
c_{lma}	2.134×10^{2}	2.209×10^{2}	2.503×10^{2}	2.372×10^{2}
C_{lab}	6.142×10^{2}	5.709×10^{2}	8.586×10^{2}	5.618×10^{2}
C_{fol}	7.971×10^{2}	1.212×10^{2}	8.029×10^{2}	1.285×10^{2}
C_{roo}	3.984×10^{4}	2.539×10^{4}	4.114×10^{4}	2.553×10^{4}
C_{woo}	5.075×10^{7}	2.764×10^{7}	5.075×10^{7}	2.764×10^{7}
C_{lit}	4.157×10^{4}	5.416×10^{4}	7.179×10^{4}	5.532×10^{4}
C_{som}	1.454×10^{6}	1.106×10^{6}	1.482×10^{6}	1.105×10^{6}

Table 4: Standard deviations for each experiment after assimilation, calculated using equation 19.