Investigating the role of background and observation error correlations in improving a model forecast of forest carbon balance using 4D-Var.

Ewan Pinnington ADD AFFILIATIONS

October 26, 2015

Abstract

Efforts to implement variational data assimilation routines with functional ecology and land surface models have been limited with sequential and Markov chain Monte Carlo data assimilation methods being more prevalent. When data assimilation has been used with models of carbon balance, background errors (describing our knowledge of error in our prior model estimates before data assimilation) and observation errors have largely been treated as independent and uncorrelated. In numerical weather prediction it has been shown that including correlations in these errors can considerably improve data assimilation results and forecasts. In this paper we implement a Four-Dimensional Variational (4D-Var) data assimilation scheme with a simple model of forest carbon balance, for joint parameter and state estimation, assimilating observations of Net Ecosystem Exchange (NEE) taken at the Alice Holt flux site in Hampshire, UK managed by Forest Research. 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 error statistics. In data assimilation background and observation error statistics are often described by the background error covariance matrix and the observation error covariance matrix (commonly treated as diagonal). MORE? We move away from a diagonal representation of the background error covariance matrix, $\mathbf{B}$, and the observation error covariance matrix, $\mathbf{R}$. We outline novel methods for creating a correlated $\mathbf{B}$ and $\mathbf{R}$ and show that using these new correlated matrices can almost halve the root mean square error in our models forecast of NEE in comparison to the results when using an uncorrelated diagonal $\mathbf{B}$ and $\mathbf{R}$, going from 4.22gCm^{-2}day^{-1} to 2.38gCm^{-2}day^{-1}.

1 Introduction

1.1 Tree blurb

Terrestrial ecosystems 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 [Ciais et al., 2014]. Terrestrial ecosystem carbon uptake is the least understood process in the global carbon cycle. 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 [Baldocchi, 2008] providing a valuable resource for model validation and data assimilation.

1.2 DA paragraph

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 the impact of data assimilation has been that the four day forecast in 2014 has 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) [Rabier et al., 2000, Rawlins et al., 2007], which has been shown to be a significant improvement over its predecessor three-dimensional variational 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, this can sometimes prove difficult to calculate. Using techniques such as automatic-differentiation 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. Variational data assimilation can be used for joint parameter and state estimation by augmenting the parameters into the state vector [Navon, 1998]. By including the parameters in the state vector we must also specify error statistics and error correlations for them. In Smith et al. [2009] it is shown that the prescription of these error statistics and their correlations can have a significant impact on parameter-state estimates obtained from the assimilation.

1.3 DALEC and ecosystem models with data assimilation (draw out gaps):

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 ecological systems [Fox et al., 2009, Niu et al., 2014, Quaife et al., 2008, Richardson et al., 2010, Zobitz et al., 2014, 2011]. 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 DALEC being implemented in a variational routine by Delahaies et al. [2013]. There have been examples of global land surface models being implemented with variational methods such as the Carbon Cycle Data Assimilation System (CCDAS) [Kaminski et al., 2013] and the ORganizing Carbon and Hydrology In Dynamic EcosystEms model (ORCHIDEE) [Krinner et al., 2005]. These examples have mainly been used to assimilate data from satellite observations with a few examples where site level data has been assimilated [Bacour et al., 2015, Verbeeck et al., 2011].

Background errors (describing our knowledge of error in our 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 $\mathbf{B}$ and $\mathbf{R}$ respectively. The off-diagonal elements of these matrices indicate the correlations between the parameter and state variables for $\mathbf{B}$ and the correlations between observation errors for $\mathbf{R}$. In the assimilation the off-diagonal terms in the $\mathbf{B}$ 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 variables in the assimilation. In 4D-Var the $\mathbf{B}$ matrix is propagated implicitly by the forecast model, this means that even a diagonal $\mathbf{B}$ matrix can develop correlations throughout an assimilation window. Including correlations in $\mathbf{B}$ 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 $\mathbf{R}$ still often treated as diagonal [Stewart et al., 2013]. Including some correlation structure in $\mathbf{R}$ 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 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. Specifying serial 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] increasing the impact of less frequent observations.

1.4 What does this paper do/results:

In this paper we implement the new version of DALEC (DALEC2 [Bloom and Williams, 2015]) in a 4DVAR data assimilation scheme for joint state and parameter estimation, assimilating daily NEE observations from the Alice Holt flux site in Hampshire, UK which is managed by Forest Research. This assimilation scheme is then subjected to rigorous testing to ensure correctness. We then outline a new method for including parameter and state correlations in the background error covariance matrix and a method for including serial correlations in the observation error covariance matrix. These 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. We show that specifying parameter and state correlations in the prior knowledge and serial correlations between observation errors can significantly improve the predicted carbon balance.

2 Model and Data Assimilation Methods

2.1 Alice Holt research forest

Alice Holt is an established research forest located in Hampshire, England with observational data spanning 50 years. The site is managed by Forest Research. The flux tower site is situated in the Straits Inclosure which is a mainly deciduous part of the forest comprising of mostly oak trees with a hazel understory, although there is a bank of conifer approximately 1km north west of the flux tower site [Pitman and Broadmeadow, 2001]. Flux tower records of Net Ecosystem Exchange (NEE) are available from 1999 up to present day [Wilkinson et al., 2012], meaning that Alice Holt has one of the longest records of flux tower NEE data in the world.

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.

The model equations for the carbon pools at day $t+1$ are as follows:

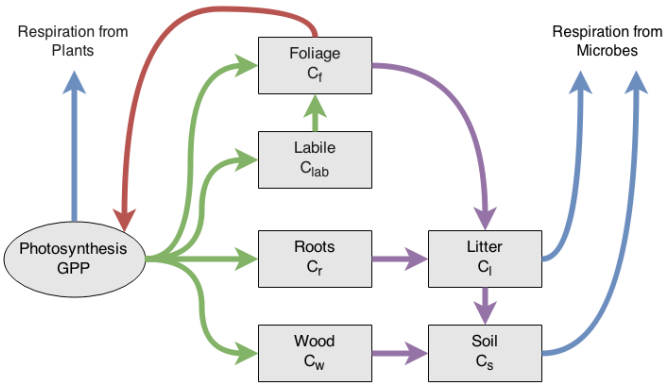


Figure 1: Representation of the fluxes in the DALEC2 carbon balance model. Green arrows represent C allocation, purple arrows represent litter fall and decomposition fluxes, blue arrows represent respiration fluxes and the red arrow represents the influence of leaf area index in the GPP function.

$$GPP^t = ACM(C_f^t, c_{lma}, c_{eff}, \Psi) \tag{1}$$
$$C_{lab}^{t+1} = (1-\Phi_{on})C_{lab}^t + (1-f_{auto})(1-f_{fol})f_{lab}GPP^t, \tag{2}$$
$$C_f^{t+1} = (1-\Phi_{off})C_f^t + \Phi_{on}C_{lab}^t + (1-f_{auto})f_{fol}GPP^t, \tag{3}$$
$$C_r^{t+1} = (1-\theta_{roo})C_r^t + (1-f_{auto})(1-f_{fol})(1-f_{lab})f_{roo}GPP^t, \tag{4}$$
$$C_w^{t+1} = (1-\theta_{woo})C_w^t + (1-f_{auto})(1-f_{fol})(1-f_{lab})(1-f_{roo})GPP^t, \tag{5}$$
$$C_l^{t+1} = (1-(\theta_{lit}+\theta_{min})e^{\Theta T^t})C_l^t + \theta_{roo}C_r^t + \Phi_{off}C_f^t, \tag{6}$$
$$C_s^{t+1} = (1-\theta_{som}e^{\Theta T^t})C_s^t + \theta_{woo}C_w^t + \theta_{min}e^{\Theta T^t}C_l^t, \tag{7}$$

where T^t 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. The model parameters used in equations 1 to 7 are included in the appendix. 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

In 4D-Var we aim to maximise the probability of $P(\mathbf{x}_0|\mathbf{y})$, the initial state $\mathbf{x}_0$ given a set of observations $\mathbf{y}$ over some time window, $0,\ldots,N$. The probability $P(\mathbf{x}_0|\mathbf{y})$ is maximised by minimising a cost function $J(\mathbf{x})$ derived from Bayes Theorem [Lawless et al., 2013]. The cost function is given as,

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0-\mathbf{x}_b)^T\mathbf{B}^{-1}(\mathbf{x}_0-\mathbf{x}_b) + \frac{1}{2}\sum_{i=0}^{N}(\mathbf{y}_i - h_i(\mathbf{x}_i))^T\mathbf{R}_i^{-1}(\mathbf{y}_i-h_i(\mathbf{x}_i)), \tag{8}$$

where $\mathbf{x}_b$ is the so-called background and acts as the initial guess to the state $\mathbf{x}_0$, $\mathbf{B}$ is the background error covariance matrix and quantifies our knowledge of the error in the background, h_i is the observation operator at time t_i and maps the state vector evolved by the nonlinear model ($m_{0\rightarrow i}(\mathbf{x}_0) = \mathbf{x}_i$) to the observations at this time ($\mathbf{y}_i$) and $\mathbf{R}_i$ is the observation error covariance matrix at time t_i and represents our knowledge of the uncertainty in the observations. The state that minimises the cost function is called the analysis and is denoted as $\mathbf{x}_a$. This state is found using a minimisation routine that takes as arguments the cost function, the initial guess ($\mathbf{x}_b$) and also the gradient of the cost function given as,

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) - \sum_{i=0}^{N} \mathbf{M}_{i,0}^T \mathbf{H}_i^T \mathbf{R}_i^{-1}(\mathbf{y}_i - h_i(\mathbf{x}_i)), \tag{9}$$

where $\mathbf{H}_i = \frac{\partial h_i(\mathbf{x}_i)}{\partial \mathbf{x}_i}$ is our 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 $\mathbf{M}_i = \frac{\partial m_i(\mathbf{x}_i)}{\partial \mathbf{x}_i}$. In practice $\nabla J(\mathbf{x}_0)$ is calculated using Lagrange multipliers as shown in Lawless et al. [2013]. We can rewrite the cost function and its gradient to avoid the sum notation as,

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}_b)^T \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) + \frac{1}{2}(\hat{\mathbf{y}} - \hat{h}(\mathbf{x}_0))^T \hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{h}(\mathbf{x}_0)) \tag{10}$$

and

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) - \hat{\mathbf{H}}^T \hat{\mathbf{R}}^{-1}(\hat{\mathbf{y}} - \hat{h}(\mathbf{x}_0)), \tag{11}$$

where,

$$\hat{\mathbf{y}} = \begin{pmatrix} \mathbf{y}_0 \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, \ \hat{h}(\mathbf{x}_0) = \begin{pmatrix} h_0(\mathbf{x}_0) \\ h_1(m_{0\to1}(\mathbf{x}_0)) \\ \vdots \\ h_N(m_{0\to N}(\mathbf{x}_0)) \end{pmatrix}, \ \hat{\mathbf{R}} = \begin{pmatrix} \mathbf{R}_{0,0} & \mathbf{R}_{0,1} & \dots & \mathbf{R}_{0,N} \\ \mathbf{R}_{1,0} & \mathbf{R}_{1,1} & \dots & \mathbf{R}_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{N,0} & \mathbf{R}_{N,1} & \dots & \mathbf{R}_{N,N} \end{pmatrix} \text{ and } \hat{\mathbf{H}} = \begin{pmatrix} \mathbf{H}_0 \\ \mathbf{H}_1\mathbf{M}_0 \\ \vdots \\ \mathbf{H}_N\mathbf{M}_{N,0} \end{pmatrix}. \tag{12}$$

Solving the cost function in this form also allows us to build serial correlations into the observation error covariance matrix $\hat{\mathbf{R}}$. WHAT ARE OFF-DIAGONAL R'S HERE!

2.4 Implementation and testing of 4D-Var system

In our DALEC2 4D-Var scheme we are performing joint parameter and state estimation. Therefore the 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. Here the nonlinear model (DALEC2) only updates the initial carbon pool values when evolving the state vector forward in time with the parameters being held constant. To find the background guess, $\mathbf{x}_b$, to the state vector we can either use a previous DALEC2 model forecast's estimate to 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 variances used in this paper were provided from existing runs of the the CARbon DAta-MOdel fraMework (CARDAMOM) [Exbrayat et al., 2015]. This is a worse resolution dataset which provides a reasonable first guess at DALEC2 state and parameter values for the Alice Holt research site.

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 our model. In the following tests we use a diagonal approximation to our background and observational error covariance matrices so that, $\mathbf{B}_{diag} = \text{diag}(\boldsymbol{\sigma}_b)^2$ and $\hat{\mathbf{R}}_{diag} = \text{diag}(\boldsymbol{\sigma}_o)^2$, where $\boldsymbol{\sigma}_b$ and $\boldsymbol{\sigma}_o$ are the vectors of the background and observational standard deviations respectively.

In this paper we assimilate observations of daily NEE. The flux tower actually produces an estimate of NEE every half-hour, 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 observations have a quality control flag associated with the best observations from the EddyPro flux processing software [LI-COR, Inc., 2015]. We take a variance of $0.5\text{gCm}^{-2}\text{day}^{-1}$ in our assimilated observations of daily NEE [Williams et al., 2005]. The minimisation routine used in our data assimilation experiments is the truncated Newton method [Nocedal and Wright, 1999] from the Python package Scipy.optimize. In sections 2.4.1 to 2.4.3 we show tests of our scheme.

2.4.1 Test of tangent linear model

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,

$$m_{0\to i}(\mathbf{x}_0 + \gamma\delta\mathbf{x}_0) \approx m_{0\to i}(\mathbf{x}_0) + \mathbf{M}_{i,0}\gamma\delta\mathbf{x}_0, \tag{13}$$

where $\delta\mathbf{x}_0$ is a perturbation of the initial state 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 perturbation $\delta\mathbf{x}_0$. We can test this by rearranging equation 13 to find the relative error,

$$E_R = \frac{||m_{0\to i}(\mathbf{x}_0 + \gamma\delta\mathbf{x}_0) - m_{0\to i}(\mathbf{x}_0)||}{||\mathbf{M}_{i,0}\gamma\delta\mathbf{x}_0||}, \tag{14}$$

where we expect $E_R \to 0$ as $\gamma \to 0$ (here we are using the Euclidean norm). Figure 2 shows equation 14 plotted for DALEC2 with a TLM evolving our state 731 days forward in time for different values γ, with a 5% perturbation $\delta\mathbf{x}_0$. Figure 2 shows that the TLM behaves as expected for values of γ approaching 0.

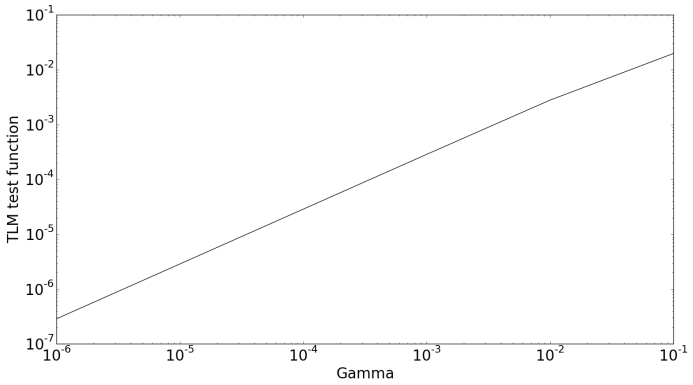


Figure 2: Plot of the tangent linear model test function (equation 14) for DALEC2, for a TLM evolving our state 731 days forward in time and a 5% perturbation, $\delta\mathbf{x}_0$.

It is also useful to show how our TLM behaves over a time window to see how the error in the TLM grows as we evolve our state further forward in time. We again rearrange equation 13 with an additional error term to find,

$$\text{percentage error in TLM} = \left|\frac{||m_{0\to i}(\mathbf{x}_0+\delta\mathbf{x}_0)-m_{0\to i}(\mathbf{x}_0)||}{||\mathbf{M}_{i,0}\delta\mathbf{x}_0||} - 1\right| \times 100. \tag{15}$$

In figure 3 we can see that our TLM for DALEC2 performs well after being run forward a year with less than a 3% error for all values of $\delta\mathbf{x}_0$. By the second year we see some peaks in our 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. Even at these peaks our error is still reasonable reaching a maximum at 7% and then coming back to around 1%. For this reason we present results using a one year assimilation window in this paper.

2.4.2 Test of adjoint model

The adjoint model we have implemented for DALEC2 passes correctness tests [Lawless et al., 2013]. For our TLM $\mathbf{M}_{i,0}$ and its adjoint $\mathbf{M}_{i,0}^T$ we have the identity

$$< \mathbf{M}_{i,0}\delta\mathbf{x}_0, \mathbf{M}_{i,0}\delta\mathbf{x}_0 > = < \delta\mathbf{x}_0, \mathbf{M}_{i,0}^T\mathbf{M}_{i,0}\delta\mathbf{x}_0 > \tag{16}$$

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

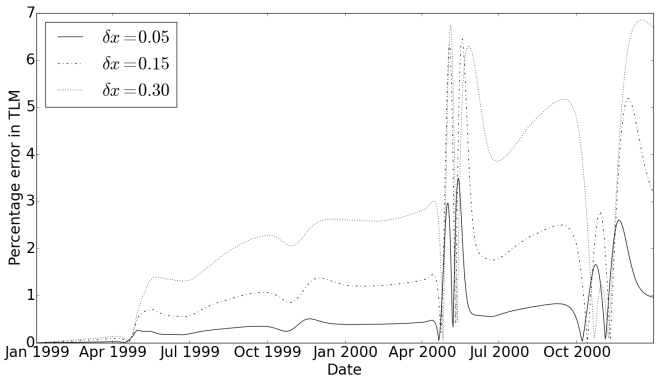


Figure 3: Plot of the percentage error in the tangent linear model (equation 15) for DALEC2 when evolving the model state forward over a period of two years with three differing values of perturbation, $\delta \mathbf{x}_0$.

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]. For our cost function J and its gradient ∇J we can show that we have implemented ∇J correctly using the identity,

$$f(\alpha) = \frac{|J(\mathbf{x}_0 + \alpha \mathbf{b}) - J(\mathbf{x}_0)|}{\alpha \mathbf{b}^T \nabla J(\mathbf{x}_0)} = 1 + O(\alpha), \tag{17}$$

where $\mathbf{b}$ is a vector of unit length and α is a parameter controlling the size of $\mathbf{b}$. For small values of α not too close to machine zero we should have $f(\alpha)$ close to 1. Figure 4 shows $f(\alpha)$ for a 731 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.

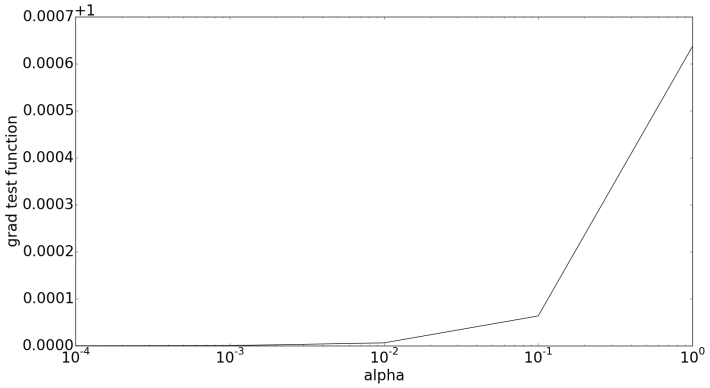


Figure 4: Test of the gradient of the cost function for a 731 day assimilation window with $\mathbf{h} = \mathbf{x}_0 ||\mathbf{x}_0||^{-1}$. REDO

We can also plot $|f(\alpha) - 1|$, where we expect $|f(\alpha) - 1| \to 0$ as $\alpha \to 0$. In figure 5 we have plotted $|f(\alpha) - 1|$ for the same conditions as in figure 4, we can see that $|f(\alpha) - 1| \to 0$ as $\alpha \to 0$, as expected (before $|f(\alpha) - 1|$ gets too close to machine zero at $O(\alpha) = 10^{-5}$). This gives us confidence that the gradient of our cost function is implemented correctly.

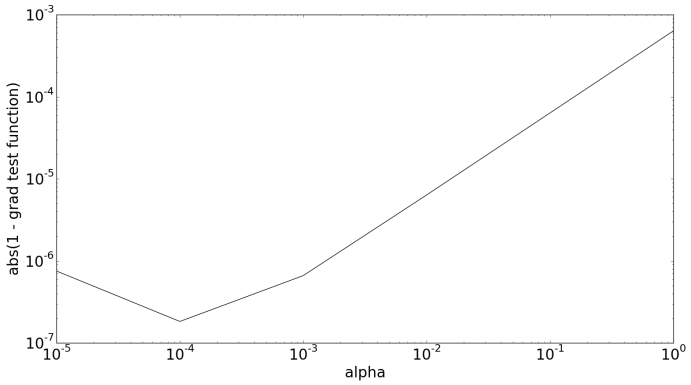


Figure 5: Test of the gradient of the cost function, $|f(\alpha) - 1|$. As $\alpha \to 0$ we have $|1 - f(\alpha)| \to 0$ up to $O(\alpha) = 10^{-4}$ where we have gone past the precision of the computer.REDO

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 include parameter-state correlations in $\mathbf{B}$. In this paper we present a method using a set of ecological dynamical constraints on model parameters and state variables from Bloom and Williams [2015]. In Bloom and Williams [2015] implementing these constraints in a Metropolis Hastings MCMC data assimilation routine is shown to improve results significantly. The constraints impose conditions on carbon pool turnover and allocation ratios, steady state proximity and growth and 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 state vector, $\mathbf{x}_i$, from the multivariate truncated normal distribution described by our $\mathbf{x}_b$, associated variances and parameter-state ranges given in table REF (in 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 to our ensemble, else it is discarded.

Once we have a full ensemble (we chose an ensemble size of 1500 as past this point values of correlations did not appear to change significantly) we then take the covariance of the ensemble to find our $\mathbf{B}_{corr}$. In figure 7 we have plotted the correlation matrix or normalised error covariance matrix of $\mathbf{B}_{corr}$. 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 being 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}.

2.6 Specifying serial correlations in the observational error covariance matrix

In this paper we use observations of daily NEE. The flux tower actually produces an estimate of NEE every half-hour, 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 observations have a quality control flag associated with the best observations from the EddyPro flux processing software [LI-COR, Inc., 2015]. We take a variance of $0.5\text{gCm}^{-2}\text{day}^{-1}$ in our assimilated observations of daily NEE [Williams et al., 2005]. Errors in NEE observations come

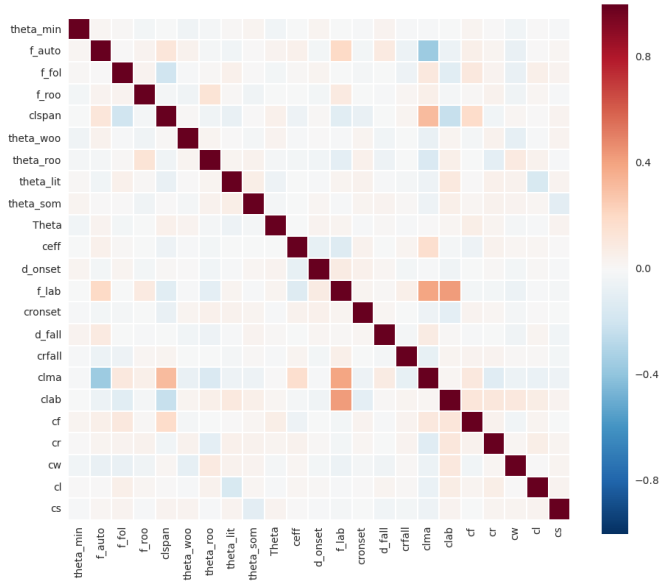


Figure 6: Background error correlation matrix created using method in section 2.5

from different sources such as instrument errors, sampled ecosystem structure and turbulent conditions (when we have low turbulence and limited air mixing NEE is underestimated) [Papale et al., 2006]. Due to this dependance 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.

In section 2.3 we have re-written the 4D-Var cost function in equation 10 in order to allow the specification of serial observation error correlations in our assimilation scheme. These serial correlations are represented by the off-diagonal terms of $\hat{\mathbf{R}}$. As we do not have a practical method of estimating the serial correlation in NEE observation error we adapt the simple Gaussian model found in Järvinen et al. [1999] (a second order autoregressive correlation function was also tested but not presented here). The correlation r between 2 observations at times t_1 and t_2 is given as,

$$r = \begin{cases} a\exp\left[\frac{-(t_1-t_2)^2}{\tau^2}\right] + (1-a)\delta_{t_1-t_2} & |t_1 - t_2| \leq \eta \\ 0 & \eta < |t_1 - t_2| \end{cases}, \tag{18}$$

where τ is the e-folding time, a is the 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 we believe the correlation length scale to be quite short for our assimilated observations but also because this makes $\hat{\mathbf{R}}$ better conditioned and therefore easier to invert in the assimilation process.

Figure 2.6 shows $\hat{\mathbf{R}}$ created using equation 18, 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 as correlations are likely to be on the scale of a day REFS.

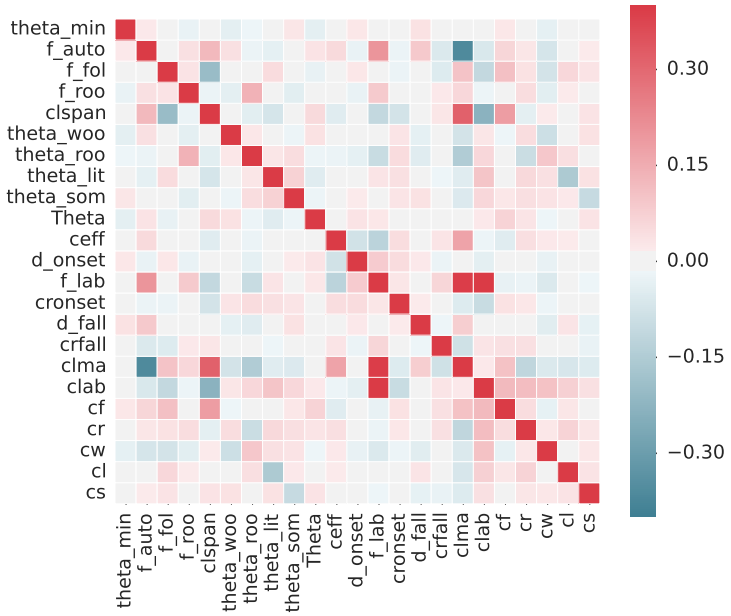


Figure 7: Background error correlation matrix created using method in section 2.5

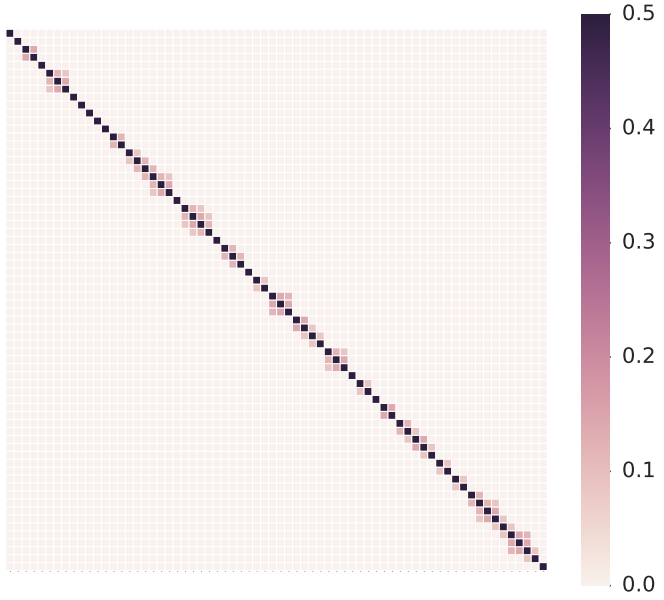


Figure 8: Observation error covariance matrix for the 67 observations used in assimilation created using method in section 2.6 with $\tau = 4$, $a = 0.3$ and $\eta = 4$.

3 Results

3.1 Experiments

In the following sections we present the results of 4 experiments where we vary the representations of $\mathbf{B}$ and $\hat{\mathbf{R}}$ while assimilating the same NEE observations in the window from 1999-2000. We then forecast the NEE over the next 14 years and compare with the observed data. These experiments are outlined in table 4 where $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ are the diagonal matrices of the parameter-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	×	×		
B		×	×	
C	×			×
D			×	×

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

In all the following experiments we present results when assimilating NEE observations in the period 1999-2000 from the Alice Holt flux site, other assimilation windows were also tested but are not shown here.

3.2 Experiment A

In this experiment $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ were used in our 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 9a shows assimilation and forecast results for NEE. We can see that our analysis trajectory (green line) fits well with the observations during the the assimilation window (1999-2000) and then diverges in the forecast (2000-2014). One reason for this is that our analysis is being over constrained by the observations and under constrained by the background. Using a diagonal background error covariance matrix ($\mathbf{B}_{diag}$) we are not adding much background information to our system and therefore the observations dominate and result in the over fitting of our analysis to the assimilated observations of NEE.

To see how well our forecast performs after assimilation we show a scatter plot of modelled NEE against observed NEE in figure 11a. Here we have a Root-Mean-Square Error (RMSE) of 4.22gCm^{-2} and a bias of -0.3gCm^{-2} for our forecast of NEE, whereas our analysis (1999-2000) has a RMSE of 1.36gCm^{-2} and a bias of -0.03gCm^{-2}. The background trajectory meanwhile has a RMSE of 3.86gCm^{-2} and a bias of -1.60gCm^{-2} in the analysis window (1999-2000) and the same RMSE of 3.86gCm^{-2} but a bias of -1.36gCm^{-2} during the forecast period (2000-2014). Although using $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$ in our assimilation has considerably reduced the RMSE in our analysis period, it has also increased the RMSE in our forecast of NEE. However it has reduced the bias in the model forecast considerably from -1.36gCm^{-2} to -0.3gCm^{-2}. The bias in the background comes from a constant under prediction of the more extreme negative values of NEE and this leads to considerably worse results than our analysis and its forecast for total forest carbon uptake.

Maybe include something about the estimated reduction in error for parameter and state variables using the diagonal terms of **B and **A** (analysis error covariance matrix). How best to compare and present these?

3.3 Experiment B

Here $\mathbf{B}_{corr}$ (as defined in section 2.5) and $\hat{\mathbf{R}}_{diag}$ are used in our assimilation. In figure 9b we can see that the forecast performs considerably better than in experiment A and indeed from figure 11b we see that our forecasts RMSE has almost halved (now 2.56gCm^{-2}) with a reduction in bias also, now -0.2gCm^{-2}. In comparison using $\mathbf{B}_{corr}$ in our assimilation slightly degrades the fit for our analysis (1999-2000), with a RMSE of 1.42gCm^{-2} and a bias of -0.04gCm^{-2}, this is because our assimilation scheme is now more constrained by the background than in experiment A. Therefore using $\mathbf{B}_{corr}$ in our assimilation reduces the problem of overfitting to our assimilated observations of NEE as seen in experiment A. Another improvement made by using $\mathbf{B}_{corr}$ in our assimilation is that our minimisation routine converges to a solution more quickly, taking 218 fewer function iterations than experiment A to converge.

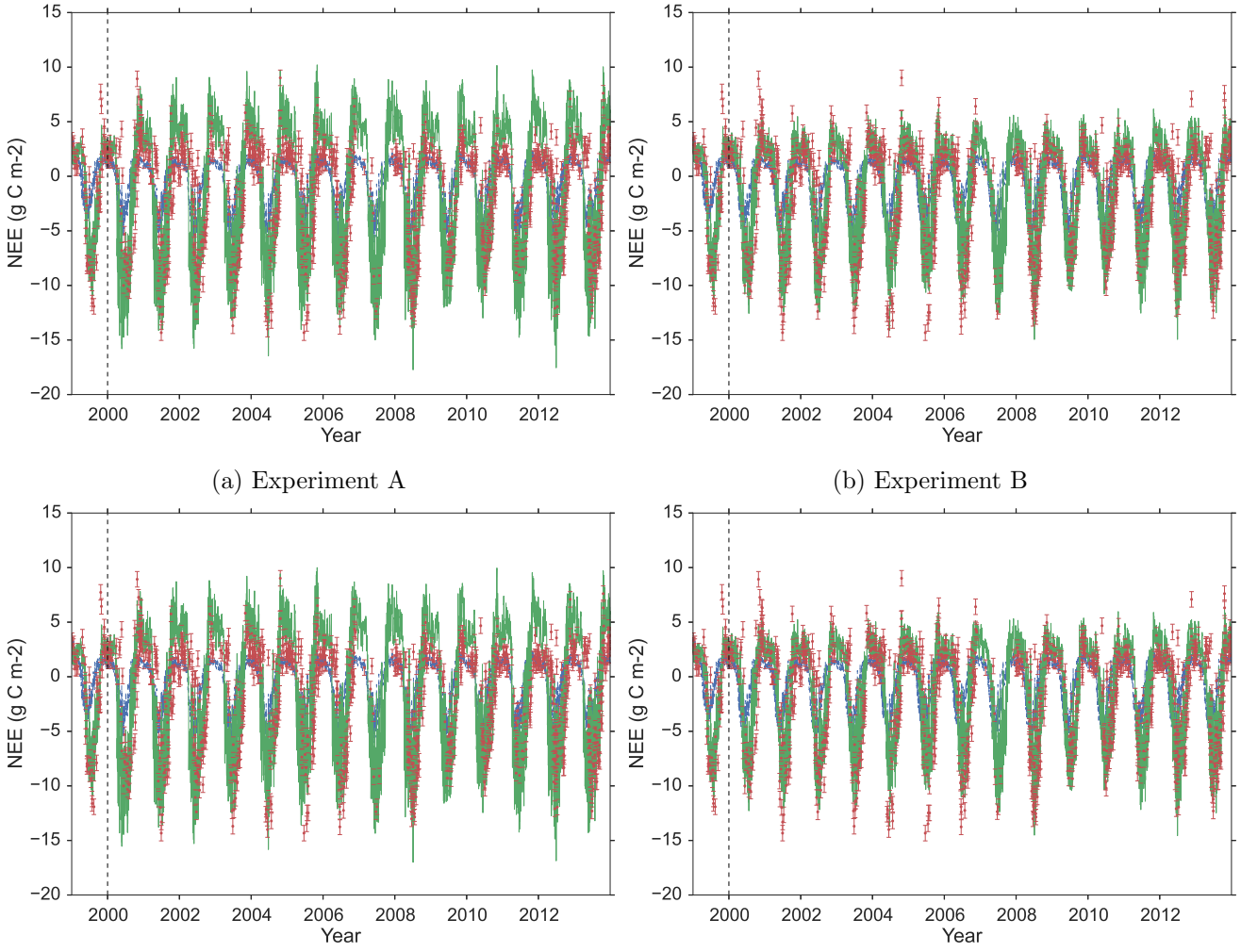
3.4 Experiment C

Here we use $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{corr}$ (as defined in section 2.6) in the assimilation. Results shown in figure **??** appear similar to those in section 3.2 however there are some differences. Using $\hat{\mathbf{R}}_{corr}$ in our assimilation has also reduced the minimisation routines convergence time, taking 127 less function iterations to converge than experiment A. We also have a slight reduction in RMSE for our forecast (now 4.09gCm^{-2}) from experiment A. As in experiment B our analysis gets slightly worse as using $\hat{\mathbf{R}}_{corr}$ acts to reduce the overfitting of the analysis to the assimilated observations. 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 quite small, we have therefore added less information to our system when using $\hat{\mathbf{R}}_{corr}$ in the assimilation. We would expect the effect to be larger when trying to predict half hourly observations of NEE using a model with a half-hourly time step and assimilating half-hourly observations of NEE as we would then have much stronger correlations in $\hat{\mathbf{R}}_{corr}$ (as discussed in section 2.6). We also expect that specifying these 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 observations and also reduce the information content of the data streams with more observations [Daley, 1992, Järvinen et al., 1999].

3.5 Experiment D

In the final experiment we use $\mathbf{B}_{corr}$ and $\hat{\mathbf{R}}_{corr}$ in the assimilation. Figure 9d 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. The number of function evaluations taken for the minimisation routine to converge is 316; the lowest of all the experiments. Also the forecast RMSE is reduced again still from results in experiment B to 2.38gCm^{-2}. Using both matrices appears to combine the beneficial effects described in both section 3.3 and section 3.4.

From the results in experiment D and integrating over the area of the Alice Holt flux site (~ 0.93km^2) the forecasted (2000-2014) total carbon uptake for the Alice Holt research site is 5.04×10^9gC. In comparison the less accurate results from experiment A (using $\mathbf{B}_{diag}$ and $\hat{\mathbf{R}}_{diag}$) predict a total carbon uptake of 5.26×10^9gC, a difference of 2.25×10^8gC. This is quite a substantial difference as we are considering the carbon uptake of a small (~ 0.93km^2) research site only.



(a) Experiment A

(b) Experiment B

(c) Experiment C

(d) Experiment D

Figure 9: 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, red dots: observations from Alice Holt flux site with error bars.

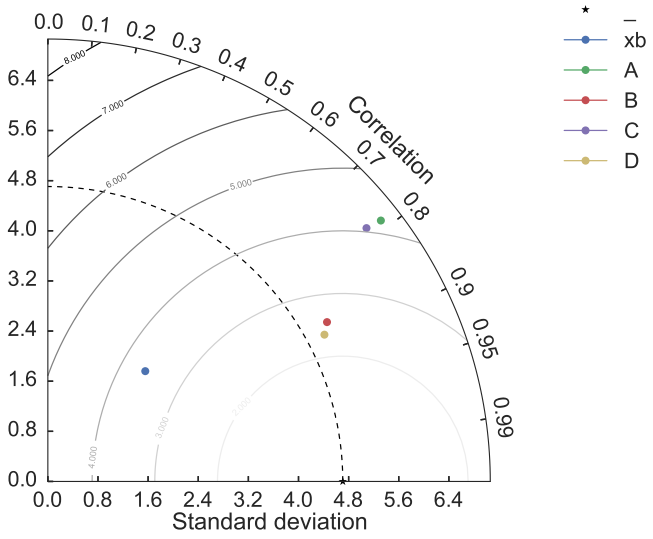


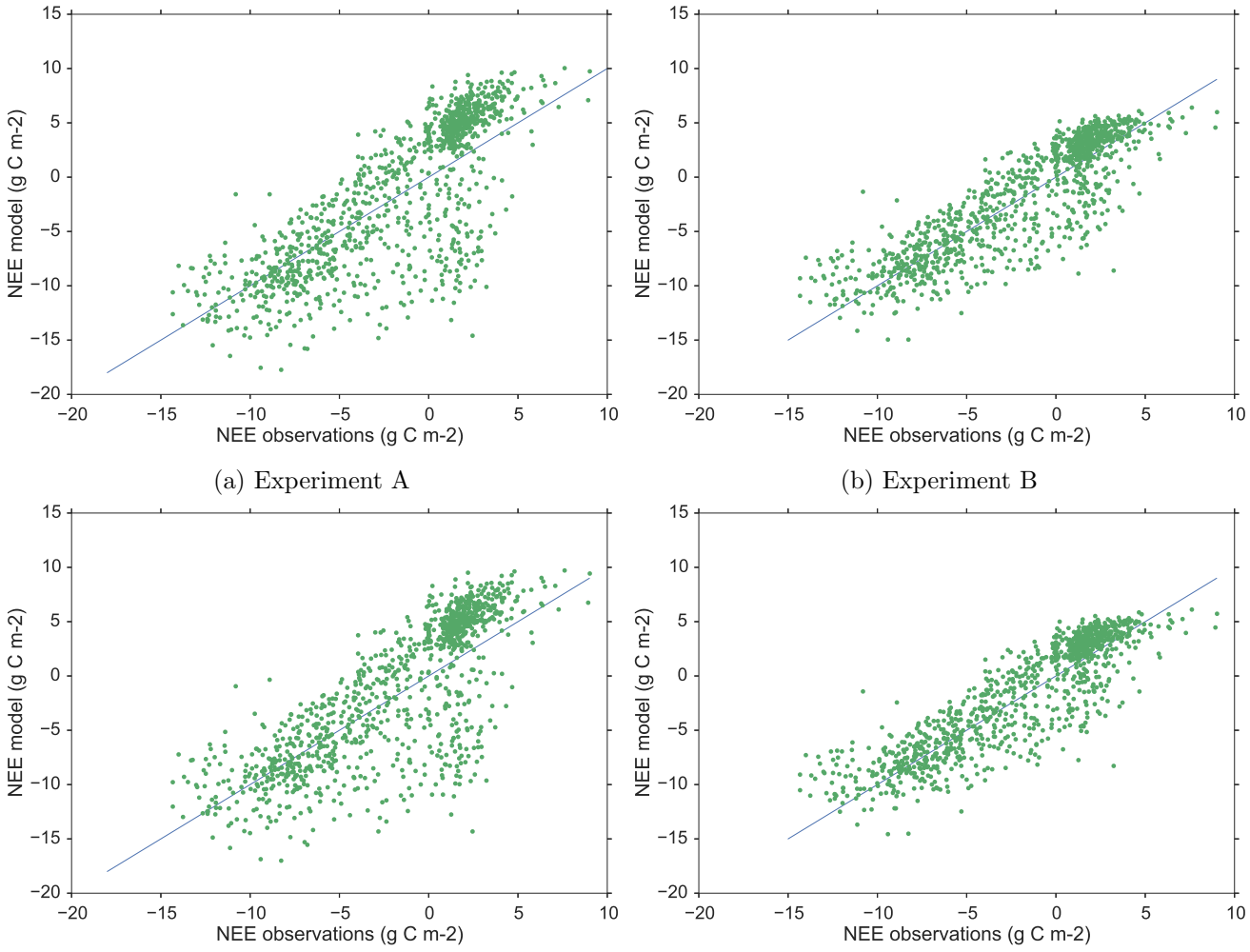
Figure 10: Taylor diagram displaying statistical comparison of the four experiment and background forecast (2000-2014) results with observations of NEE (gCm^{-2}).

Experiment	RMSE (gCm^{-2})	Bias (gCm^{-2})	Correlation coefficient	Minimisation function evaluations
Background	3.86	-1.60	0.70	n/a
A	1.36	-0.03	0.96	571
B	1.42	-0.04	0.95	353
C	1.37	-0.09	0.96	444
D	1.43	-0.09	0.95	316

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

Experiment	RMSE (gCm^{-2})	Bias (gCm^{-2})	Correlation coefficient	Minimisation function evaluations
Background	3.86	-1.36	0.66	n/a
A	4.22	-0.30	0.79	571
B	2.56	-0.20	0.87	353
C	4.09	-0.51	0.78	444
D	2.38	-0.33	0.88	316

Table 3: Forecast (2000-2014) results for experiments and background when judged against observed NEE.



(a) Experiment A

(b) Experiment B

(c) Experiment C

(d) Experiment D

Figure 11: Forecast scatter plot of modelled NEE vs. observations for 2000-2014 (green dots). Blue line represents the 1-1 line.

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 quicker assimilation results than MCMC techniques as we have knowledge of the derivative of the model, however we do also assume that the problem is Gaussian whereas MCMC techniques do not. We have shown that 4D-Var is a valid tool for improving the DALEC2 model estimate of NEE.

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

- The technique presented here to specify $\mathbf{B}_{corr}$ has been shown to significantly improve forecasts of NEE. Other techniques were also tested (not presented here) to create a correlated $\mathbf{B}$, one of which added correlations to $\mathbf{B}$ using the DALEC2 model. Here we evolved an ensemble of state vectors over the length of the chosen assimilation window, 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 considerably in some cases. Many different tests were run using different background vectors and variances and it was found that specifying some form of correlations in $\mathbf{B}$ always made some improvement to the results of our assimilation.

- Here the $\hat{\mathbf{R}}_{corr}$ used in our experiments has improved our forecast of NEE, however many other choices of $\hat{\mathbf{R}}_{corr}$ tested for this paper degraded the forecast. This is probably due to specifying unrealistic correlations and suggests a more diagnostic approach is needed for the calculation of serial correlations in $\hat{\mathbf{R}}$. One option would be to adapt the Deroziers diagnostic [Desroziers et al., 2005], which has been used successfully in NWP for diagnosing observation error correlations for observations taken at the same time, and extending this technique to diagnose serial correlations.

- Specifying serial correlations also allows us to address the issues discussed by Richardson et al. [2010], that when assimilating multiple data streams more frequently sampled observation types (such as NEE) have much more impact on the assimilation than data streams sampled less frequently. Specifying serial correlations between observations of the same type has the effect of reducing the weight given to the mean of the observations [Järvinen et al., 1999], thus allowing less frequent data streams to have more impact on the assimilation.

  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 model with a day-night time step this will allow us to build in the type of correlations investigated by Baldocchi et al. [2015] between ecosystem respiration and canopy photosynthesis.

- We have also seen an improvement in convergence times for our minimisation routine when specifying correlations in our $\mathbf{B}$ and $\hat{\mathbf{R}}$ matrices, this has been the case in most other tests not presented here. This is due to the correlated matrices constraining the assimilation more and making the problem less ill-posed.

5 Conclusion

- 4DVar is a valid tool for improving the DALEC2 model estimate of NEE.

- Including correlations in the background error covariance matrix improves our forecast after assimilation (sometimes greatly) in comparison to using a diagonal representation of $\mathbf{B}$.

- Specifying serial correlations between observation errors in $\hat{\mathbf{R}}$ can improve our forecast, however more work is needed to find a method for diagnosing these correlations.

6 Acknowledgements

Luke, FR, NERC, NCEO,

References

C. Bacour, P. Peylin, N. MacBean, P. J. Rayner, F. Delage, F. Chevallier, M. Weiss, J. Demarty, D. Santaren, F. Baret, D. Berveiller, E. Dufrêne, and P. Prunet. Joint assimilation of eddy-covariance flux measurements and FAPAR products over temperate forests within a process-oriented biosphere model. *Journal of Geophysical Research: Biogeosciences*, pages n/a–n/a, 2015. ISSN 21698953. doi: 10.1002/2015JG002966. URL `http://doi.wiley.com/10.1002/2015JG002966`.

Dennis Baldocchi. 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):1–26, 2008.

Dennis Baldocchi, Cove Sturtevant, and Fluxnet Contributors. Does day and night sampling reduce spurious correlation between canopy photosynthesis and ecosystem respiration? *Agricultural and Forest Meteorology*, 207:117–126, 2015. ISSN 01681923. doi: 10.1016/j.agrformet.2015.03.010. URL `http://linkinghub.elsevier.com/retrieve/pii/S016819231500088X`.

Ross N Bannister. 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(637):1951–1970, 2008.

Peter Bauer, Alan Thorpe, and Gilbert Brunet. The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55, 2015.

A. A. Bloom and M. Williams. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a modeldata fusion framework. *Biogeosciences*, 12(5):1299–1315, 2015. ISSN 1726-4189. doi: 10.5194/bg-12-1299-2015. URL `http://www.biogeosciences.net/12/1299/2015/`.

Bobby H Braswell, William J Sacks, Ernst Linder, and David S Schimel. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biology*, 11(2):335–355, 2005.

Philippe Ciais, Christopher Sabine, Govindasamy Bala, Laurent Bopp, Victor Brovkin, Josep Canadell, Abha Chhabra, Ruth DeFries, James Galloway, Martin Heimann, et al. 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*, pages 465–570. Cambridge University Press, 2014.

Roger Daley. The Effect of Serially Correlated Observation and Model Error on Atmospheric Data Assimilation, 1992. ISSN 0027-0644.

DP Dee, SM Uppala, AJ Simmons, Paul Berrisford, P Poli, S Kobayashi, U Andrae, MA Balmaseda, G Balsamo, P Bauer, et al. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011.

Sylvain Delahaies, Ian Roulstone, and Nancy Nichols. A regularization of the carbon cycle data-fusion problem. In *EGU General Assembly Conference Abstracts*, volume 15, page 4087, 2013.

Gérald Desroziers, Loic Berre, Bernard Chapnik, and Paul Poli. Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3385–3396, 2005.

Jean-françois Exbrayat, T Luke Smallman, A Anthony Bloom, and Mathew Williams. 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, 2015.

Andrew Fox, Mathew Williams, Andrew D Richardson, David Cameron, Jeffrey H Gove, Tristan Quaife, Daniel Ricciuto, Markus Reichstein, Enrico Tomelleri, Cathy M Trudinger, et al. 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(10):1597–1615, 2009.

Heikki Järvinen, Erik Andersson, and François Bouttier. Variational assimilation of time sequences of surface observations with serially correlated errors. *Tellus A*, 51(4):469–488, 1999.

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

T. Kaminski, W. Knorr, G. Schürmann, M. Scholze, P. J. Rayner, S. Zaehle, S. Blessing, W. Dorigo, V. Gayler, R. Giering, N. Gobron, J. P. Grant, M. Heimann, a. Hooker-Stroud, S. Houweling, T. Kato, J. Kattge, D. Kelley, S. Kemp, E. N. Koffi, C. Köstler, P. P. Mathieu, B. Pinty, C. H. Reick, C. Rödenbeck, R. Schnur, K. Scipal, C. Sebald, T. Stacke, a. Terwisscha Van Scheltinga, M. Vossbeck, H. Widmann, and T. Ziehn. The BETHY/JSBACH Carbon Cycle Data Assimilation System: Experiences and challenges. *Journal of Geophysical Research: Biogeosciences*, 118(4):1414–1426, 2013. ISSN 21698961. doi: 10.1002/jgrg.20118.

G. Krinner, Nicolas Viovy, Nathalie de Noblet-Ducoudré, Jérôme Ogée, Jan Polcher, Pierre Friedlingstein, Philippe Ciais, Stephen Sitch, and I. Colin Prentice. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, 19(1):1–33, 2005. ISSN 08866236. doi: 10.1029/2003GB002199.

A. S. Lawless, M.J.P. Cullen, M. A. Freitag, S. Kindermann, and R. Scheichl. *Variational data assimilation for very large environmental problems. In Large Scale Inverse Problems: Computational Methods and Applications in the Earth Sciences*. De Gruyter, 2013.

Yong Li, I. Michael Navon, Weiyu Yang, Xiaolei Zou, J. R. Bates, S. Moorthi, and R. W. Higgins. Four-Dimensional Variational Data Assimilation Experiments with a Multilevel Semi-Lagrangian Semi-Implicit General Circulation Model. *Monthly Weather Review*, 122(5):966–983, may 1994. ISSN 0027-0644. doi: 10.1175/1520-0493(1994)122¡0966:FDVDAE¿2.0.CO;2. URL `http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1994)122%3C0966:FDVDAE%3E2.0.CO%3B2`.

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

Andrew C Lorenc and F Rawlins. Why does 4d-var beat 3d-var? *Quarterly Journal of the Royal Meteorological Society*, 131(613):3247–3257, 2005.

I. M. Navon, X. Zou, J. Derber, and J. Sela. Variational Data Assimilation with an Adiabatic Version of the NMC Spectral Model. *Monthly Weather Review*, 120(7):1433–1446, jul 1992. ISSN 0027-0644. doi: 10.1175/1520-0493(1992)120¡1433:VDAWAA¿2.0.CO;2. URL `http://journals.ametsoc.org/doi/abs/10.1175/1520-0493(1992)120%3C1433%3AVDAWAA%3E2.0.CO%3B2`.

IM Navon. Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans*, 27(1):55–79, 1998.

Shuli Niu, Yiqi Luo, Michael C. Dietze, Trevor F. Keenan, Zheng Shi, Jianwei Li, and F. Stuart Chapin Iii. The role of data assimilation in predictive ecology. *Ecosphere*, 5(5):art65, 2014. ISSN 2150-8925. doi: 10.1890/ES13-00273.1. URL `http://www.esajournals.org/doi/abs/10.1890/ES13-00273.1`.

Jorge Nocedal and Stephen J Wright. *Numerical Optimization*. Springer Science & Business Media, 1999. ISBN 0387987932. URL `http://books.google.co.uk/books/about/Numerical_Optimization.html?id=epc5fX0lqRIC&pgis=1`.

D. Papale, M. Reichstein, M. Aubinet, E. Canfora, C. Bernhofer, W. Kutsch, B. Longdoz, S. Rambal, R. Valentini, T. Vesala, and D. Yakir. Towards a standardized processing of Net Ecosystem Exchange measured with eddy covariance technique: algorithms and uncertainty estimation. *Biogeosciences*, 3 (4):571–583, 2006. ISSN 1726-4189. doi: 10.5194/bg-3-571-2006.

Rona Pitman and Mark Broadmeadow. Leaf area, biomass and physiological parameterisation of ground vegetation of lowland oak woodland. *Forestry Commission, Edinburgh*, 2001.

Tristan Quaife, Philip Lewis, Martin De Kauwe, Mathew Williams, Beverly E. Law, Mathias Disney, and Paul Bowyer. Assimilating canopy reflectance data into an ecosystem model with an Ensemble Kalman Filter. *Remote Sensing of Environment*, 112(4):1347–1364, 2008. ISSN 00344257. doi: 10.1016/j.rse.2007.05.020.

Florence Rabier, H Järvinen, E Klinker, J-F Mahfouf, and A Simmons. The ecmwf operational implementation of four-dimensional variational assimilation. i: Experimental results with simplified physics. *Quarterly Journal of the Royal Meteorological Society*, 126(564):1143–1170, 2000.

F Rawlins, SP Ballard, KJ Bovis, AM Clayton, D Li, GW Inverarity, AC Lorenc, and TJ Payne. The met office global four-dimensional variational data assimilation scheme. *Quarterly Journal of the Royal Meteorological Society*, 133(623):347–362, 2007.

Andrew D Richardson, Mathew Williams, David Y Hollinger, David JP Moore, D Bryan Dail, Eric A Davidson, Neal A Scott, Robert S Evans, Holly Hughes, John T Lee, et al. Estimating parameters of a forest ecosystem c model with measurements of stocks and fluxes as joint constraints. *Oecologia*, 164 (1):25–40, 2010.

Polly J Smith, Sarah L Dance, Michael J Baines, Nancy K Nichols, and Tania R Scott. Variational data assimilation for parameter estimation: application to a simple morphodynamic model. *Ocean Dynamics*, 59(5):697–708, 2009.

Laura M. Stewart, Sarah L. Dance, and Nancy K. Nichols. Data assimilation with correlated observation errors: Experiments with a 1-D shallow water model. *Tellus, Series A: Dynamic Meteorology and Oceanography*, 65(1):1–14, 2013. ISSN 02806495. doi: 10.3402/tellusa.v65i0.19546.

Hans Verbeeck, Philippe Peylin, Cédric Bacour, Damien Bonal, Kathy Steppe, and Philippe Ciais. fluxes in Amazon forests: Fusion of eddy covariance data and the ORCHIDEE model. *Journal of Geophysical Research*, 116(G2):1–19, 2011. ISSN 0148-0227. doi: 10.1029/2010JG001544.

Sebastian F. Walter and Lutz Lehmann. Algorithmic differentiation in Python with AlgoPy. *Journal of Computational Science*, 4(5):334–344, sep 2013. ISSN 18777503. doi: 10.1016/j.jocs.2011.10.007. URL `http://www.sciencedirect.com/science/article/pii/S1877750311001013`.

PP Weston, W Bell, and JR Eyre. Accounting for correlated error in the assimilation of high-resolution sounder data. *Quarterly Journal of the Royal Meteorological Society*, 140(685):2420–2429, 2014.

M Wilkinson, EL Eaton, MSJ Broadmeadow, and JIL Morison. Inter-annual variation of carbon uptake by a plantation oak woodland in south-eastern england. *Biogeosciences*, 9(12):5373–5389, 2012.

Mathew Williams, Edward B Rastetter, David N Fernandes, Michael L Goulden, Gaius R Shaver, and Loretta C Johnson. Predicting gross primary productivity in terrestrial ecosystems. *Ecological Applications*, 7(3):882–894, 1997.

Mathew Williams, Paul A Schwarz, Beverly E Law, James Irvine, and Meredith R Kurpius. An improved analysis of forest carbon dynamics using data assimilation. *Global Change Biology*, 11(1):89–105, 2005.

J. M. Zobitz, David J P Moore, Tristan Quaife, Bobby H. Braswell, Andrew Bergeson, Jeremy a. Anthony, and Russell K. Monson. 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, 2014. ISSN 01681923. doi: 10.1016/j.agrformet.2014.04.011. URL `http://dx.doi.org/10.1016/j.agrformet.2014.04.011`.

JM Zobitz, AR Desai, DJP Moore, and MA Chadwick. A primer for data assimilation with ecological models using markov chain monte carlo (mcmc). *Oecologia*, 167(3):599–611, 2011.

Appendix

Parameter	Background	A	B	C	D
θ_{min}	$9.810e-04$	$1.000e-05$	$1.000e-05$	$1.000e-05$	$1.000e-05$
f_{auto}	$5.190e-01$	$3.000e-01$	$3.089e-01$	$3.000e-01$	$3.134e-01$
f_{fol}	$1.086e-01$	$1.640e-01$	$3.025e-01$	$1.822e-01$	$3.006e-01$
f_{roo}	$4.844e-01$	$3.886e-01$	$4.398e-01$	$4.298e-01$	$4.452e-01$
c_{lspan}	$1.200e+00$	$1.000e+00$	$1.026e+00$	$1.000e+00$	$1.023e+00$
θ_{woo}	$1.013e-04$	$1.185e-04$	$1.228e-04$	$1.254e-04$	$1.228e-04$
θ_{roo}	$3.225e-03$	$4.977e-03$	$5.136e-03$	$5.070e-03$	$5.041e-03$
θ_{lit}	$3.442e-03$	$2.688e-03$	$1.601e-03$	$2.107e-03$	$1.563e-03$
θ_{som}	$1.113e-04$	$1.873e-04$	$1.443e-04$	$1.914e-04$	$1.482e-04$
Θ	$4.147e-02$	$8.000e-02$	$7.697e-02$	$8.000e-02$	$7.616e-02$
c_{eff}	$7.144e+01$	$1.000e+02$	$9.347e+01$	$1.000e+02$	$9.276e+01$
d_{onset}	$1.158e+02$	$1.196e+02$	$1.237e+02$	$1.194e+02$	$1.230e+02$
f_{lab}	$3.204e-01$	$3.801e-01$	$1.000e-02$	$3.707e-01$	$1.000e-02$
c_{ronset}	$4.134e+01$	$2.752e+01$	$4.567e+01$	$2.924e+01$	$4.680e+01$
d_{fall}	$2.205e+02$	$3.199e+02$	$2.874e+02$	$3.169e+02$	$2.871e+02$
c_{rfall}	$1.168e+02$	$6.801e+01$	$5.605e+01$	$6.450e+01$	$5.517e+01$
c_{lma}	$1.285e+02$	$3.869e+01$	$5.165e+01$	$4.237e+01$	$5.163e+01$
C_{lab}	$1.365e+02$	$1.000e+01$	$1.000e+01$	$1.000e+01$	$1.000e+01$
C_f	$6.864e+01$	$1.000e+01$	$1.000e+01$	$1.000e+01$	$1.000e+01$
C_r	$2.838e+02$	$5.470e+02$	$5.265e+02$	$5.290e+02$	$5.015e+02$
C_w	$6.506e+03$	$7.292e+03$	$7.275e+03$	$7.614e+03$	$7.262e+03$
C_l	$5.988e+02$	$2.165e+02$	$6.088e+02$	$2.911e+02$	$6.258e+02$
C_s	$1.936e+03$	$2.557e+03$	$2.302e+03$	$2.606e+03$	$2.355e+03$

Table 4: Parameter values for background and difference experiments analysis vectors.

Write appendix!