Investigating the role of **R** and **B** in improving a model of forest carbon balance using 4D-Var.

Ewan Pinnington

September 9, 2015

Abstract

Efforts to implement functional ecology and land surface models in variational data assimilation routines 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 simple model of forest carbon balance in a Four-Dimensional Variational (4D-Var) data assimilation scheme, for parameter and state estimation, assimilating observations of Net Ecosystem Exchange (NEE) taken at the Alice Holt flux site in Hampshire managed by Forest Research. We then investigate the effect of specifying correlations in background and observation errors by moving away from a diagonal representation of the background error covariance matrix, **B**, and the observation error covariance matrix, **R**. We outline novel methods for creating a correlated **B** and **R** and show that using these new correlated matrices can almost half the root mean square error in our models forecast of NEE in comparison to the results when using an uncorrelated diagonal **B** and **R**.

1 Introdution

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 greatly reduce 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 forest ecosystems and the atmosphere are now routinely made at flux tower sites world-wide [Baldocchi, 2008] providing a great 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 been used in many applications with great results for improving models and forecast. One such application has been In

numerical weather prediction where the impact of data assimilation has been vast with the four day forecast in 2014 having the same level of accuracy as the one day forecast when data assimilation was first introduced operationally in 1979 [Kalnay, 2003, Rabier, 2005]. 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. The current method implemented at several leading operational numerical weather prediction centres is Four-Dimensional Variational data assimilation (4D-Var), which has been shown to be a significant improvement over its predecessor three-dimensional variational assimilation [Lorenc and Rawlins, 2005].

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

[Zobitz et al., 2011] (say something more about this ref) Currently efforts to use variational assimilation with carbon balance models have been limited, with sequential and Markov Chain Monte Carlo methods being more prevalent.

CCDAS example of global ecosystem model implemented in 4D-Var for assimilation of observations of atmospheric CO2 levels.

DALEC implemented in a variaitional assimilation scheme by Sylvain (find ref for this)

1.4   What does this paper do:

In this paper we will paramterize a model of forest carbon balance (DALEC2) using a 4D-Var scheme in order to produce better forecasts of forest carbon balance. We will use data from the research site at Alice Holt forest run by Forest Research.

1.5   Results:

2   Model and Data Assimilation Methods

2.1   Alice Holt research forest

Alice Holt is a well 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. Flux tower records of Net Ecosystem Exchange (NEE) are available from 1999 up to present day, meaning that Alice Holt has one of the longest records of flux tower data in the UK.

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, 2014] 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 site's leaf area index (a function of C_f) as arguments.

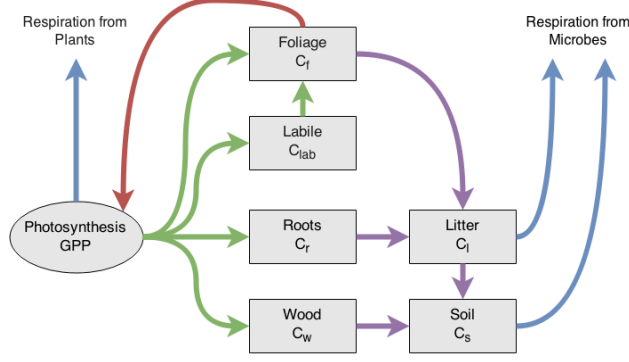The model equations for the carbon pools at day $t + 1$ are as follows:

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 feedback of foliar carbon to 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 and the equations used to calculate GPP, Φ_{on} and Φ_{off} are included in the appendix. The full details of this version of DALEC can be found in Bloom and Williams [2014].

2.3 4D-Var

In 4D-Var we aim to maximise the probability of our initial state $\mathbf{x}_0$ given a set of observations $\mathbf{y}$, $P(\mathbf{x}_0|\mathbf{y})$, over some time window, N. $P(\mathbf{x}_0|\mathbf{y})$ is maximised by minimising a cost function $J(\mathbf{x})$ derived from Bayes Theorem [Lewis et al., 2006]. 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 our background and acts as our initial guess to our state $\mathbf{x}_0$, $\mathbf{B}$ is the background error covariance matrix and quantifies our knowledge of the error in our background, h_i is our observation operator at time t_i and maps our state vector evolved by our nonlinear model ($m_{0 \to i}(\mathbf{x}_0) = \mathbf{x}_i$) 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 the cost function, our initial guess ($\mathbf{x}_b$) and also the gradient of the cost function defined 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 our tangent linear model with $\mathbf{M}_i = \frac{\partial m_i(\mathbf{x}_i)}{\partial \mathbf{x}_i}$. 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,

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

2.4 Testing of 4D-Var system

In our DALECV2 4D-Var scheme the state vector, $\mathbf{x}_0$, corresponds to the vector of the 17 model parameters and 6 initial carbon pool values. We use a diagonal approximation to our background and observational error covariance matrices so that, $\mathbf{B} = \text{diag}(\underline{\sigma}_b^2)$ and $\hat{\mathbf{R}} = \text{diag}(\underline{\sigma}_o^2)$, where $\underline{\sigma}_b$ and $\underline{\sigma}_o$ are the vectors of the background and observational standard deviations respectively.

In order to find the tangent linear model (TLM) for DALECV2 we need 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 in Python to calculate the TLM at each time step. This package uses forward mode automatic differentiation to calculate the derivative of our model. AlgoPy was selected after testing other automatic differentiation packages (PyAutoDiff and ad.py) and finding that AlgoPy could compute the TLM in the fastest time. We now have all the tools to create our 4D-Var scheme. In sections 2.4.1 to 2.4.3 we will show some 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 relevant tests. In 4D-Var we assume the tangent linear hypothesis,

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

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,

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

4

where we should have $E_R \to 0$ as $\gamma \to 0$. In figure 2 we have plotted equation 14 for DALEC2 with a TLM evolving our state 731 days forward in time for a 5% perturbation $\delta\mathbf{x}_0$. Figure 2 shows that our TLM behaves as expected for values of γ approaching 0.
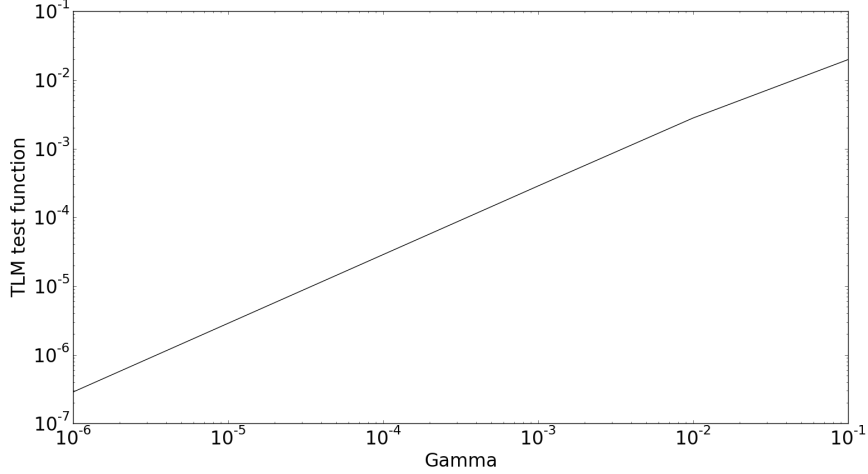

Figure 2: Plot of the tangent linear model test function for DALEC2, for a TLM evolving our state 731 days forward in time and a 5% perturbation, $\delta\mathbf{x}_0$.

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

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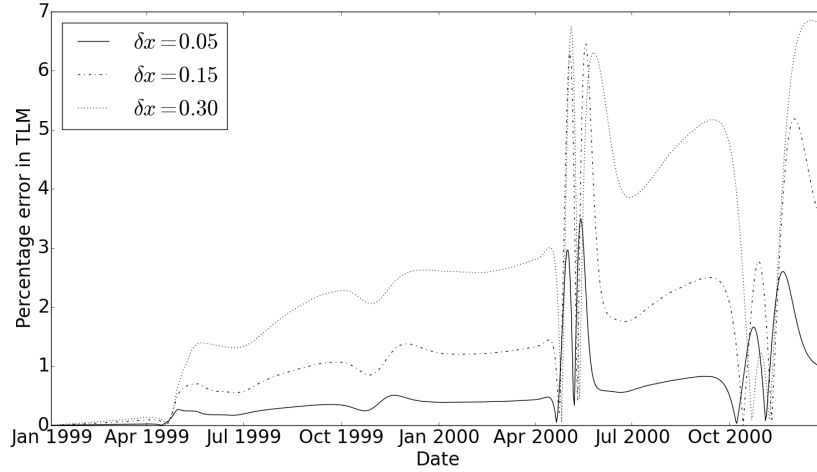

Figure 3: Plot of the percentage error in our tangent linear model for DALEC2 when evolving our state forward over a period of two years with three differing values of perturbation, $\delta\mathbf{x}_0$.

In figure 3 we can see that our TLM for DALEC2 performs very 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

5

our error in spring and autumn, this is where our leaf on and leaf off functions in the TLM have gone 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%.

2.4.2 Test of adjoint model

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

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

for any inner product $<,>$ and perturbation $\delta\mathbf{x}_0$. This identity has been used with differing values of $\delta\mathbf{x}_0$ and i to show that our adjoint model is implemented correctly.

2.4.3 Gradient test

The 4D-Var system we have developed passes tests for the gradient of the cost function. For our cost function J and its gradient ∇J we can show that we have implemented ∇J correctly using the identity,

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

where $\mathbf{h}$ is a vector of unit length. For small values of α not too close to machine zero we should have $f(\alpha)$ close to 1. In figure 4 we have plotted $f(\alpha)$ for a 731 day assimilation window with $\mathbf{h} = \mathbf{x}_0||\mathbf{x}_0||^{-1}$, we can see that $f(\alpha) \to 1$ as $\alpha \to 0$, as expected.
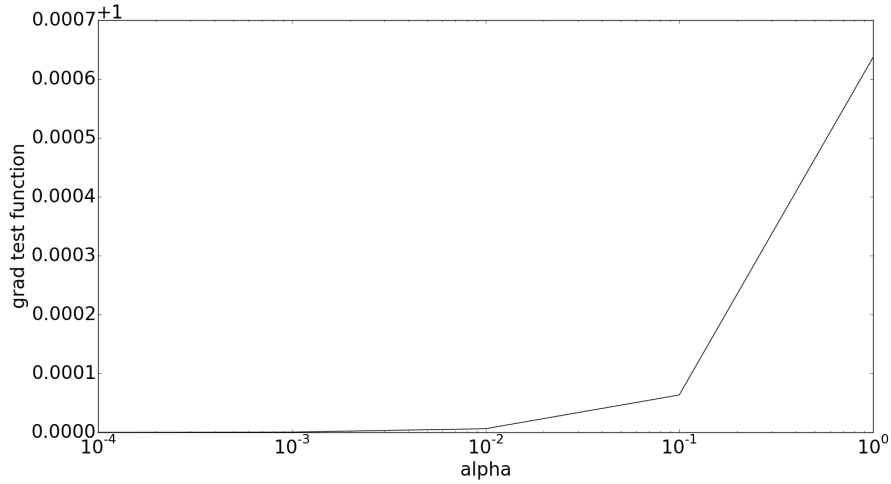

Figure 4: Test of the gradient of the cost function for a 731 day assimilation window with $\mathbf{h} = \mathbf{x}_0||\mathbf{x}_0||^{-1}$.

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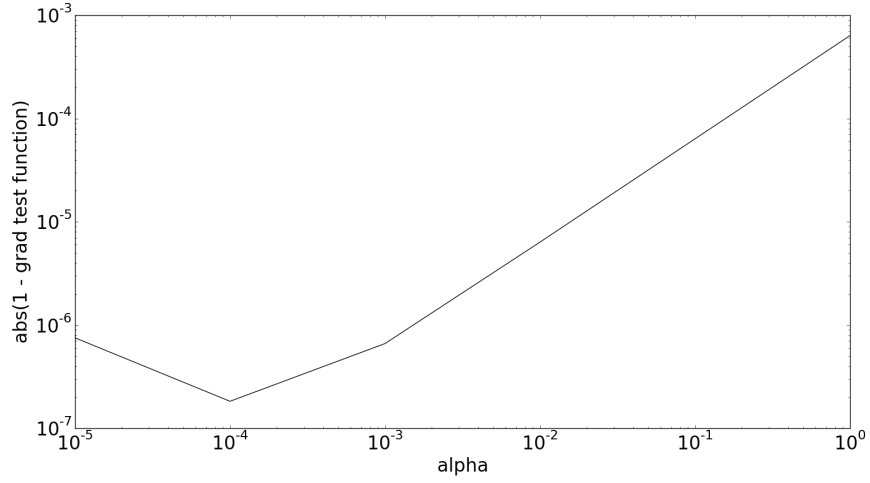
Figure 5: Test of the gradient of the cost function, $|f(\alpha) - 1|$. As $\alpha \to 0$ we have $abs(1 - f(\alpha)) \to 0$ up to $O(\alpha) = 10^{-4}$ where we have gone past the precision of the computer.

Experiment	$\mathbf{B}_{diag}$	$\mathbf{R}_{diag}$	$\mathbf{B}_{corr}$	$\mathbf{R}_{corr}$
1	×	×		
2		×	×	
3	×			×
4			×	×

Table 1: Here are the experiments!!!

2.5 Computing a correlated background error covariance matrix

Using ecological dynamical constraints (EDC's) and drawing from a truncated distribution.

2.6 Specifying serial correlations in the observational error covariance matrix

Use correlation function found in [Järvinen et al., 1999], short e-folding time as correlations thought to be on the scale of a day.

3 Results

Experiments

4 Discussion

- Improving B improves forecast for 1 year window. What is the effect on the analysis, if any?

- Using correlated B improves convergence time (check this with new experiments and maybe also CVT)

- Using correlated R improves forecast both for diagonal and correlated B but change is not significant

7

5   Conclusion

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

- Using B created with ecological dynamical constraints improves our forecast after assimilation (sometimes greatly) in comparison to using diagonal B.

- Using a correlated R improves our forecast, not as much as using the EDCB but this is expected as we are not adding as much information with our correlations in R.

References

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.

Anthony Bloom and Mathew Williams. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological "common sense" in a model-data-fusion framework. *Biogeosciences Discussions*, 11(8):12733–12772, 2014. URL `http://www.biogeosciences-discuss.net/11/12733/2014/`.

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.

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.

John M Lewis, Sivaramakrishnan Lakshmivarahan, and Sudarshan Dhall. *Dynamic data assimilation: a least squares approach*, volume 13. Cambridge University Press, 2006.

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.

Florence Rabier. Overview of global data assimilation developments in numerical weather-prediction centres. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3215–3233, 2005.

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.

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