Improvement of forest carbon balance model DALEC2 for flux site Alice Holt using four-dimensional variational data assimilation.

Ewan Pinnington

June 11, 2015

1 Introdution

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. It is vital that we improve understanding in order to better constrain predictions of future carbon budgets

Four-dimensional variational data assimilation (4D-Var) has been used extensively in numerical weather prediction to improve forecasts. Currently efforts to use variational assimilation with carbon balance models have been limited, with sequential and Markov Chain Monte Carlo methods being more prevalent. In this report 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.

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:

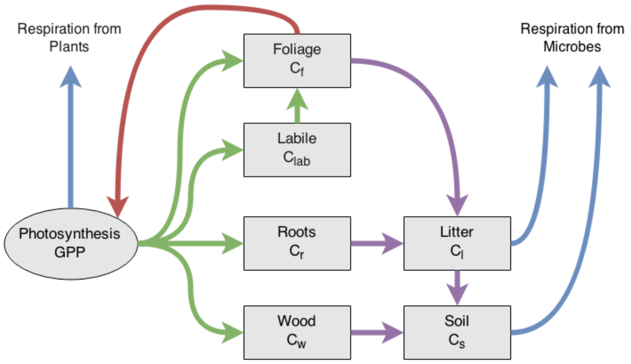

Figure 1: Representation of the fluxes in the DALEC2 carbon balance model. Green arrows represent C allocation, purple arrows represent litter fall and decomposition fluxes, blue arrows represent respiration fluxes and the red arrow represents the feedback of foliar carbon to the GPP function.

$$GPP^t = ACM(C_f^t, c_{lma}, c_{eff}, \Psi) \tag{1}$$

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

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

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

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

$$C_l^{t+1} = (1-(\theta_{lit}+\theta_{min})e^{\Theta T^t})C_l^t + \theta_{roo}C_r^t + \Phi_{off}C_f^t, \tag{6}$$

$$C_s^{t+1} = (1-\theta_{som}e^{\Theta T^t})C_s^t + \theta_{woo}C_w^t + \theta_{min}e^{\Theta T^t}C_l^t, \tag{7}$$

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,

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

3 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 3.1 to 3.3 we will show some tests of our scheme.

3.1 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,

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

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.

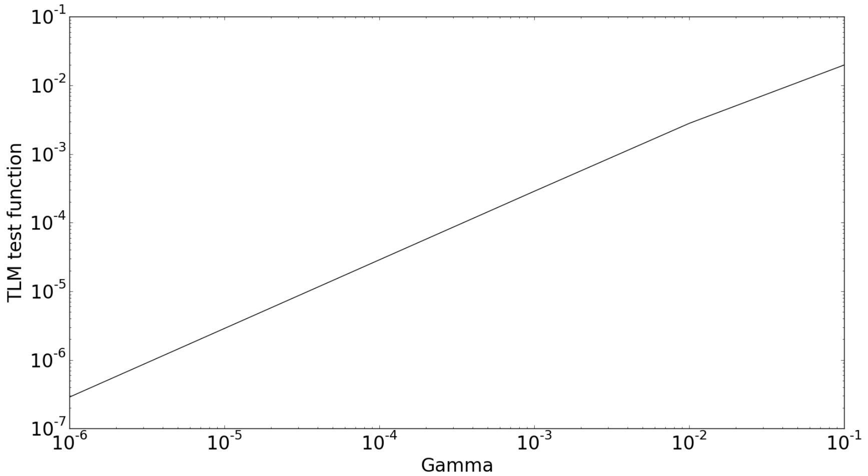

Figure 2: Plot of the tangent linear model test function for DALEC2, for a TLM evolving our state 731 days forward in time and a 5% perturbation, $\delta\mathbf{x}_0$.

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,

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

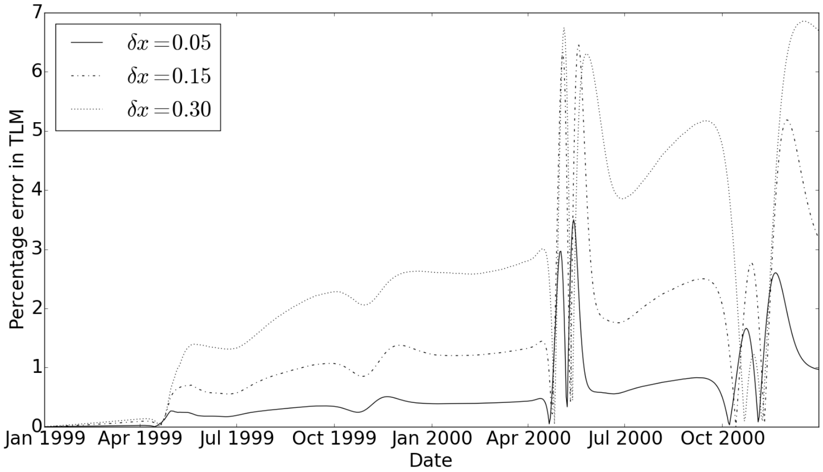

Figure 3: Plot of the percentage error in our tangent linear model for DALEC2 when evolving our state forward over a period of two years with a 5% perturbation, $\delta\mathbf{x}_0$.

In figure 3 we can see that our TLM for DALEC2 performs very well after being run forward a year with less than a 1% error. By the second year we see some peaks in 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%.

3.2 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

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

for any inner product $<,>$ and perturbation $\delta\mathbf{x}_0$. We can use this to show that our adjoint is implemented correctly.

3.3 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,

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

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.

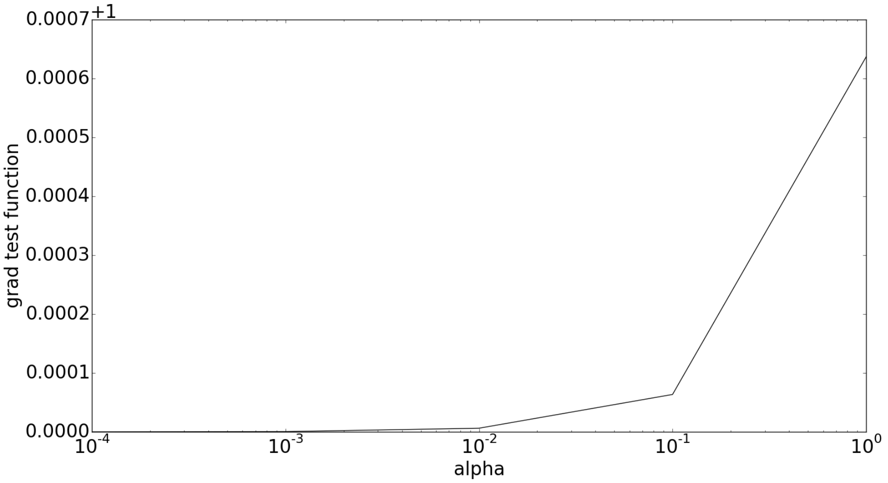

Figure 4: Test of the gradient of the cost function for a 731 day assimilation window with $\mathbf{h} = \mathbf{x}_0||\mathbf{x}_0||^{-1}$.

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.

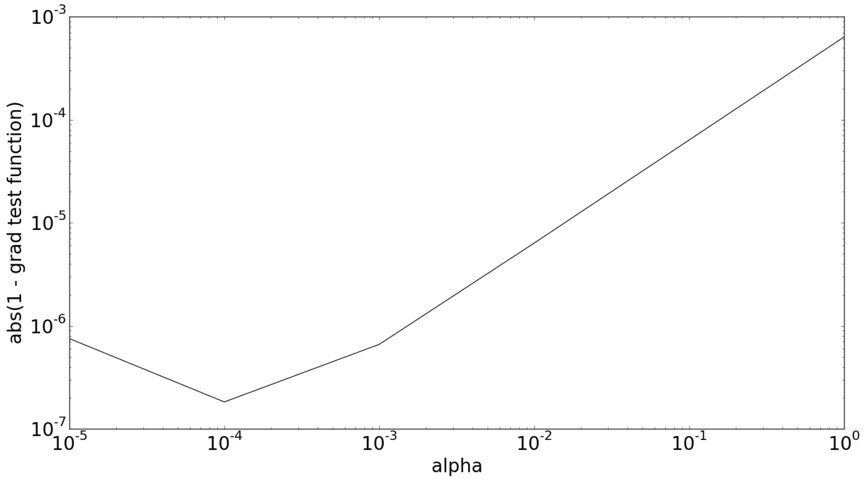

Figure 5: Test of the gradient of the cost function, $|f(\alpha)-1|$. As $\alpha \rightarrow 0$ we have $abs(1-f(\alpha)) \rightarrow 0$ up to $O(\alpha)=10^{-4}$ where we have gone past the precision of the computer.

4 Results

5 Discussion

6 Conclusion

References

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/`.

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

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.

Appendix