

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

Ewan Pinnington

March 10, 2015

## 1 Introduction

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

### 2.1 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's Theorem [Lewis et al., 2006]. The cost function is given as,

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

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 ( $m_{0 \rightarrow i}(\mathbf{x}_0) = \mathbf{x}_i$ ) to the observations at this time  $\mathbf{y}_i$  and  $\mathbf{R}_i$  is the observation error covariance matrix 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,

$$\nabla J(\mathbf{x}_0) = \mathbf{B}^{-1}(\mathbf{x}_0 - \mathbf{x}_b) - \sum_{i=0}^N \mathbf{M}_{i,0}^T \mathbf{H}_i^T \mathbf{R}_i^{-1}(\mathbf{y}_i - h_i(\mathbf{x}_i)), \quad (2)$$

where  $\mathbf{H}_i = \frac{\delta h_i(\mathbf{x}_i)}{\delta \mathbf{x}_i}$  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  $\mathbf{M}_i = \frac{\delta m_i(\mathbf{x}_i)}{\delta \mathbf{x}_i}$ .

## 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.

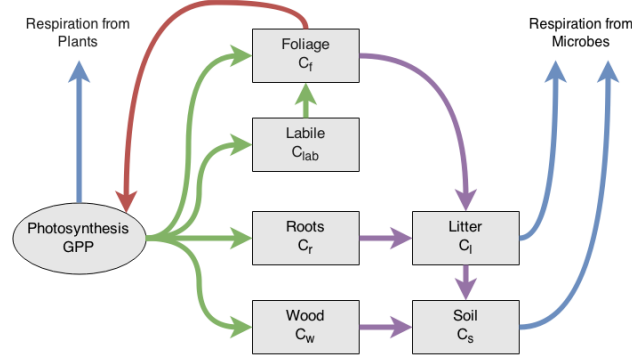


Figure 1: Representation of the fluxes in the DALEC2 carbon balance model. Green arrows represent C allocation, purple arrows represent litterfall and decomposition fluxes, blue arrows represent respiration fluxes and the red arrow represents the feedback of foliar carbon to the  $GPP$  function.

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

$$GPP^t = ACM(C_f^t, c_{lma}, c_{eff}, \Psi) \quad (3)$$

$$C_{lab}^{t+1} = (1 - \Phi_{on})C_{lab}^t + (1 - f_{auto})(1 - f_{fol})f_{lab}GPP^t, \quad (4)$$

$$C_f^{t+1} = (1 - \Phi_{off})C_f^t + \Phi_{on}C_{lab}^t + (1 - f_{auto})f_{fol}GPP^t, \quad (5)$$

$$C_r^{t+1} = (1 - \theta_{roo})C_r^t + (1 - f_{auto})(1 - f_{fol})(1 - f_{lab})f_{roo}GPP^t, \quad (6)$$

$$C_w^{t+1} = (1 - \theta_{woo})C_w^t + (1 - f_{auto})(1 - f_{fol})(1 - f_{lab})(1 - f_{roo})GPP^t, \quad (7)$$

$$C_l^{t+1} = (1 - (\theta_{lit} + \theta_{min})e^{\theta T^t})C_l^t + \theta_{roo}C_r^t + \Phi_{off}C_f^t, \quad (8)$$

$$C_s^{t+1} = (1 - \theta_{som}e^{\theta T^t})C_s^t + \theta_{woo}C_w^t + \theta_{min}e^{\theta T^t}C_l^t, \quad (9)$$

where  $T^t$  is the daily mean temperature,  $p_1, \dots, p_{10}$  are rate parameters and  $\phi$  represents the meteorological driving data used in the  $GPP$  function. The full details of this version of DALEC can be found in Bloom and Williams [2014]. It is parameterized for data from a young pine stand in Ponderosa, Oregon. The model parameters and the equations used to calculate  $GPP$  are included in the appendix. We now see how DALEC can be implemented in a four-dimensional variational data assimilation (4D-Var) framework.

Forecast skill score,  $SS = 1 - \frac{MSE_{forecast}}{MSE_{ref}}$ ,

## 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.