Agricultural and Forest Meteorology 228–229 (2016) 299–314

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

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

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

[a] Department of Meteorology, University of Reading, Reading, UK
[b] Department of Mathematics and Statistics, University of Reading, Reading, UK
[c] Centre for Sustainable Forestry and Climate Change, Forest Research, Alice Holt, Farnham, UK
[d] National Centre for Earth Observation, University of Reading, Reading, UK

ARTICLE INFO

Article history:
Received 7 January 2016
Received in revised form 2 June 2016
Accepted 6 July 2016
Available online 4 August 2016

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

ABSTRACT

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

1. Introduction

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

* Corresponding author.
E-mail address: e.m.pinnington@pgr.reading.ac.uk (E.M. Pinnington).

http://dx.doi.org/10.1016/j.agrformet.2016.07.006

a valuable resource for carbon balance model validation and data assimilation.

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

In numerical weather prediction data assimilation has been predominately used for state estimation whilst keeping parameters fixed. This is because numerical weather prediction is mainly dependent on the initial state with model physics being well understood. Ecosystem carbon cycle models are more dependent on finding the correct set of parameters to describe the ecosystem of interest (Luo et al., 2015). This is possibly why Monte Carlo Markov chain (MCMC) data assimilation methods have been used more with ecosystem carbon cycle models. Smaller ecosystem models are much less computationally expensive to run than large numerical weather prediction models, meaning that MCMC methods (requiring many more model runs than variational assimilation methods) are more easily implemented. For larger scale and more complex ecosystem models variational methods represent a much more computationally efficient option for data assimilation. Variational data assimilation can be used for joint parameter and state estimation by augmenting the state vector with the parameters (Navon, 1998). By including the parameters in the state vector we must also specify error statistics and error correlations for them. Smith et al. (2009) show that the prescription of these error statistics and their correlations can have a significant impact on parameter-state estimates obtained from the assimilation.

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

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

Background errors (describing our knowledge of error in prior model estimates before data assimilation) and observation errors have largely been treated as uncorrelated and independent in ecosystem model data assimilation schemes. In 3D- and 4D-Var schemes background and observation errors are represented by the error covariance matrices **B** and **R** respectively. The off-diagonal elements of these matrices indicate the correlations between errors in the parameter and state variables for **B** and the correlations between observation errors for **R**. In the assimilation, the off-diagonal terms in the **B** matrix act to spread information between the state and augmented parameter variables (Kalnay, 2003). This means that assimilating observations of one state variable can act to update different state and parameter variables in the assimilation when correlations are included in **B**. In 4D-Var the **B** matrix is propagated implicitly by the forecast model, so that even a propagated diagonal **B** matrix can develop correlations throughout an assimilation window. These correlations will only be in the propagated **B** matrix, with the **B** matrix valid at the initial time remaining unchanged. Including correlations in **B** has been shown to significantly improve data assimilation results in numerical weather prediction (Bannister, 2008).

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