

Data assimilation with carbon cycle models

Ewan Pinnington

December 8, 2016

1 Introduction

This chapter reviews recent efforts in using data assimilation with carbon cycle models in order to improve current estimates of ecosystem carbon balance.

2 Data assimilation methods

Data assimilation provides techniques for combining observations and prior knowledge of a system in an optimal way to find a consistent solution referred to as the analysis. The prior knowledge of a system often takes the form of a numerical model and an initial guess of the model state/parameters. Many statistical methods have been developed for data assimilation. These methods can largely be categorised as either sequential or variational. Sequential algorithms solve the system of equations needed to find an optimal solution explicitly at each observation time. Variational methods solve the equations needed for an optimal solution implicitly by minimising a cost function for all available observations over some time window. This thesis is mainly concerned with the variational technique of four-dimensional variational data assimilation (4D-Var). The majority of applications of data assimilation with ecosystem carbon balance models have used other sequential and variational techniques. In the next sections we give a general introduction to data assimilation, then expand this to 4D-Var and finally we briefly discuss other data assimilation methods not directly used in this thesis but applicable to subsequent discussion.

2.1 Introduction to data assimilation

We consider a system that can be described by a numerical model with a true model state $\mathbf{x}^t \in \mathbb{R}^n$. The initial guess to this model state (often referred to as the prior or background) $\mathbf{x}^b \in \mathbb{R}^n$ and observations of the system $\mathbf{y} \in \mathbb{R}^m$ will only be approximations to the true system state, such that

$$\mathbf{x}^b = \mathbf{x}^t + \boldsymbol{\epsilon}^b, \tag{1}$$

$$\mathbf{y} = h(\mathbf{x}^t) + \boldsymbol{\epsilon}^o, \tag{2}$$

where $\boldsymbol{\epsilon}^b$ and $\boldsymbol{\epsilon}^o$ are the prior and observation errors respectively, and $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the observation operator mapping the state to the observations. The errors in the prior and

observations are assumed to be unbiased and mutually independent with known covariance matrices $\mathbf{B} = \mathbb{E}[\boldsymbol{\epsilon}^b(\boldsymbol{\epsilon}^b)^T]$ and $\mathbf{R} = \mathbb{E}[\boldsymbol{\epsilon}^o(\boldsymbol{\epsilon}^o)^T]$.

(BLUE) introduce basic concept of DA for a linear Gaussian time-invariant system... will use this in Info Con chapter

BLUE \rightarrow 3D-Var

2.2 4D-Var

Brief but inclusive of all notation need for results chapter on information content

2.3 Sequential and Monte-Carlo approaches

Brief overview of these methods.

3 Applications to the carbon cycle

DA for NWP is considered a state estimation problem as the physics of the problem are well understood and therefore parameterisations should not change over time. For the C cycle DA is more of a joint parameter and state estimation problem with the vast majority of studies using DA to estimate both parameter and state variables for a given system. Parameters governing land surface C uptake can change over time with developing forest and disturbance events.

3.1 Site-level applications

Many MCMC routines, at the global scale these will become increasingly difficult to implement due to computational expense

3.2 Global inversions

adjoints used for faster optimisation... also allows for finding posterior error distributions and propagating these to future estimates [Scholze]

Current efforts use a stepwise approach to assimilate different distinct data streams... this has been shown to not be optimal [MacBean]

3.3 Issues faced in C cycle DA

- Equifinality: Many different combinations of parameters and states able to recreate assimilated obs.
- Information content in obs: In order to reduce the problem of equifinality it is important to combine as many distinct data streams as possible, it is of great importance that we understand the information content in potential new data streams so that we can focus efforts on campaigns that will add the most information possible to DA schemes. Important to understand what measurements best complement EC data.

- Representation of prior and observational errors: Current DA schemes take a very simple approach to defining errors. Improving the representation of error in DA schemes will also help reduce the problem of equifinality.

4 Conclusion

Many efforts and much progress being made in the field of C cycle DA. Currently there are areas that need addressing... the specification of errors, information content in available and possible new data streams and continued application of DA to new problems involving the C cycle are all important areas for progress...