Data assimilation with carbon cycle models

Ewan Pinnington

December 13, 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). 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]. We therefore discuss data assimilation for joint state and parameter estimation. 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 $\mathbf{z}^t \in \mathbb{R}^n$ and true parameters $\mathbf{p}^t \in \mathbb{R}^q$. We then define the true augmented state as

$$\mathbf{x}^t = \begin{pmatrix} \mathbf{p}^t \\ \mathbf{z}^t \end{pmatrix} \in \mathbb{R}^{q+n}. \tag{1}$$

The initial guess to this model augmented state $\mathbf{x}^b \in \mathbb{R}^{q+n}$ (often referred to as the prior or background) and observations of the system $\mathbf{y} \in \mathbb{R}^m$ will only be approximations to the true system state, such that

$$\mathbf{x}^b = \mathbf{x}^t + \boldsymbol{\epsilon}^b, \tag{2}$$

$$\mathbf{y} = h(\mathbf{x}^t) + \boldsymbol{\epsilon}^o, \tag{3}$$

where $\boldsymbol{\epsilon}^b$ and $\boldsymbol{\epsilon}^o$ are the prior and observation errors respectively, and $h : \mathbb{R}^{q+n} \to \mathbb{R}^m$ is the observation operator (can be linear or non-linear) mapping the augmented 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 $\mathbf{R} = \mathbb{E}[\boldsymbol{\epsilon}^o (\boldsymbol{\epsilon}^o)^T]$.

The best estimate to $\mathbf{x}^t$ satisfying both equation (2) and (3) is often called the analysis or the maximum a posteriori estimate, here denoted $\mathbf{x}^a$. It is possible to derive this analysis by applying Bayesian methods to probability density functions. Bayes' theorem, first discussed in Bayes and Price [1763] but formalised by Laplace [1781], states that the posterior probability of event A given that event B occurs, is proportional to the prior probability of A multiplied by the probability of event B given that event A occurs, this can be expressed mathematically as

$$\mathbb{P}(A|B) \propto \mathbb{P}(A)\mathbb{P}(B|A). \tag{4}$$

For data assimilation event A represents the augmented state of the system $\mathbf{x}$ and event B the observations $\mathbf{y}$. Maximising the probability $\mathbb{P}(A|B)$ is then equivalent to finding the augmented state that best represents the observations.

If we make the assumption of Gaussian probability density functions (pdf) with

$$\mathbb{P}^b(\mathbf{x}) = \frac{1}{\sqrt{|2\pi\mathbf{B}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)\right) \tag{5}$$

and

$$\mathbb{P}^o(\mathbf{y}|\mathbf{x}) = \frac{1}{\sqrt{|2\pi\mathbf{R}|}} \exp\left(-\frac{1}{2}(\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{x}))\right), \tag{6}$$

where $\mathbb{P}^b(\mathbf{x})$ is the pdf for the prior and $\mathbb{P}^o(\mathbf{y}|\mathbf{x})$ the pdf of the observations given the augmented state. Then from Bayes' theorem (equation (4)) the posterior pdf for the augmented state

$$\mathbb{P}^a(\mathbf{x}|\mathbf{y}) \propto \frac{1}{\sqrt{|2\pi\mathbf{B}|}\sqrt{|2\pi\mathbf{R}|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \frac{1}{2}(\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{x}))\right)$$

$$\propto \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \frac{1}{2}(\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{x}))\right), \tag{7}$$

here we can ignore the constant multiplying the exponential function as it is independent of $\mathbf{x}$. We want to maximise the probability of the augmented state $\mathbf{x}$ given the observations $\mathbf{y}$, from equation (7) we can see that to maximise $\mathbb{P}^a(\mathbf{x}|\mathbf{y})$ we must maximise the terms in the exponent, this is equivalent to minimising the quadratic cost function

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y} - h(\mathbf{x}))^T \mathbf{R}^{-1}(\mathbf{y} - h(\mathbf{x})). \tag{8}$$

This is the cost function minimised in three-dimensional variational data assimilation (3D-Var), where the minimum is found using a descent algorithm evaluating equation (8) and its gradient [Courtier et al., 1998]. We can approximate the minimum of (8) by finding its gradient and setting it to zero to obtain the best linear unbiased estimate (BLUE) [Talagrand, 1997] where

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{K}(\mathbf{y} - h(\mathbf{x}^b)), \tag{9}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}, \tag{10}$$

where $\mathbf{K}$ is the Kalman gain matrix specifying the weight of the analysis increment and $\mathbf{H} = \frac{\partial h(\mathbf{x})}{\partial \mathbf{x}}$ is the linearised observation operator. We can also approximate the analysis error covariance matrix as

$$\mathbf{A} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{B}^{-1})^{-1}, \tag{11}$$

if h is linear then (9) and (11) are exact solutions.

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.

Sequential algorithms approximate the BLUE formula to update the model trajectory at each observation time.

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 increasing 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 compliment 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...

References

M. Bayes and M. Price. An essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfrs. *Philosophical Transactions (1683-1775)*, pages 370–418, 1763.

P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ecmwf implementation of three-dimensional variational assimilation (3d-var). i: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

P. d. Laplace. Mémoire sur les probabilités. *Mémoires de l?Académie Royale des sciences de Paris*, 1778:227–332, 1781.

Y. Luo, T. F. Keenan, and M. Smith. Predictability of the terrestrial carbon cycle. *Global change biology*, 21(5):1737–1751, 2015.

O. Talagrand. Assimilation of observations, an introduction. *Journal-Meteorological Society of Japan Series 2*, 75:81–99, 1997.