Agricultural and Forest Meteorology 149 (2009) 1829–1842

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Review

A review of applications of model–data fusion to studies of terrestrial carbon fluxes at different scales

Ying-Ping Wang [a,*], Cathy M. Trudinger [a], Ian G. Enting [b]

[a] CSIRO Marine and Atmospheric Research, Centre for Australian Weather and Climate Research, Private Bag #1, Aspendale, Victoria 3195, Australia
[b] MASCOS, 139 Barry St, The University of Melbourne, Victoria 3010, Australia

ARTICLE INFO

Article history:
Received 20 January 2009
Received in revised form 16 April 2009
Accepted 14 July 2009

Keywords:
Model–data fusion
Statistical estimation
Carbon flux
Model error
Measurement error
Parameter estimation
Data assimilation
Kalman filter

ABSTRACT

Model–data fusion is defined as matching model prediction and observations by varying model parameters or states using statistical estimation. In this paper, we review the history of applications of various model–data fusion techniques in studies of terrestrial carbon fluxes in two approaches: top-down approaches that use measurements of global CO_2 concentration and sometimes other atmospheric constituents to infer carbon fluxes from the land surface, and bottom-up approaches that estimate carbon fluxes using process-based models. We consider applications of model–data fusion in flux estimation, parameter estimation, model error analysis, experimental design and forecasting. Significant progress has been made by systematically studying the discrepancies between the predictions by different models and observations. As a result, some major controversies in global carbon cycle studies have been resolved, robust estimates of continental and global carbon fluxes over the last two decades have been obtained, and major deficiencies in the atmospheric models for tracer transport have been identified. In the bottom-up approaches, various optimization techniques have been used for a range of process-based models. Model–data fusion techniques have been successfully used to improve model predictions, and quantify the information content of carbon flux measurements and identify what other measurements are needed to further constrain model predictions. However, we found that very few studies in both top-down and bottom-up approaches have quantified the errors in the observations, model parameters and model structure systematically and consistently. We therefore suggest that future research will focus on developing an integrated Bayesian framework to study both model and measurement errors systematically.

Contents

* Corresponding author. Tel.: +61 3 9239 4577; fax: +61 3 9239 4444.
E-mail address: yingping.wang@csiro.au (Y.-P. Wang).

0168-1923/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.agrformet.2009.07.009

1. Introduction

Feedback between the terrestrial carbon cycle and climate change has been shown to be positive at the global scale, and has significant impact on the predictions of climate change in the future, but the magnitude of this positive feedback and its regional variation remains quite uncertain (Friedlingstein et al., 2006). To improve the understanding of this positive feedback, we need to constrain our model simulations and predictions using biophysical and atmospheric observations available at different time and spatial scales. Two approaches have traditionally been used to estimate carbon fluxes. The top-down approach uses the observed surface atmospheric CO_2 concentration, other atmospheric constituents such as $^{13}CO_2$ or O_2/N_2, and an atmospheric transport model to infer surface CO_2 sinks and sources in time and space (Ciais et al., 1995; Rayner et al., 1999; Bousquet et al., 2000; Gurney et al., 2002). The bottom-up approach uses a process-based model developed using observations at smaller scales to estimate the exchange of carbon between the land biosphere and the atmosphere at larger scale (Cramer et al., 1999). The process-based models can be quite complex, including detailed descriptions of physical, chemical and biological processes at leaf to ecosystem scales. Improving the representation of these processes is essential for using process-based models in a predictive mode.

Over time, there has been increasing sophistication of both the models and the techniques for integrating model and data in these studies (Keeling et al., 1989; Tans et al., 1990; Enting et al., 1995; Enting, 2002; Rödenbeck et al., 2003). Model and data integration, also called model–data fusion or model–data synthesis, is defined as combining models and observations by varying some properties of the model, to give the optimal combination of both (Raupach et al., 2005). Model–data fusion therefore encompasses both parameter estimation and data assimilation. Mathieu and O'Neill (2008) describe the combination of model and data in data assimilation as "a carefully constructed procedure that brings to bear all our knowledge of the system and measurement process as well as the known inaccuracies in (i) measurements (e.g. instrumental error), (ii) governing equations of the system (e.g. parameterisations error of sub-grid scale processes), and (iii) numerical representation of these equations (i.e. discretization and computational errors)". Early examples combining models and data are not as all-encompassing as in this description, but are precursors to such an approach. The description given by Mathieu and O'Neill (2008) can be seen as the goal for model–data fusion. Here we review the application of model–data fusion techniques, and their precursors, in top-down and bottom-up approaches to estimate terrestrial carbon fluxes. As we will see, the top-down and bottom-up approaches address the same problem by different routes, but many features of the techniques used to solve them are common, so a combined review is appropriate.

In the top-down approach, early studies of CO_2 flux inversions used ad hoc scaling of flux patterns to fit available CO_2 concentration data (Keeling et al., 1989; Tans et al., 1990). Enting et al. (1995) introduced a systematic method, known as 'Bayesian synthesis inversion', for estimating fluxes and, importantly, the uncertainties in these fluxes, which was further developed by Rayner et al. (1999) for a time dependent Bayesian synthesis inversion. Since then there have been numerous flux inversions, some considering only CO_2 observations (e.g. Bousquet et al., 2000; Rödenbeck et al., 2003; Peylin et al., 2005), some considering $\delta^{13}CO_2$ in addition to CO_2 to help distinguish fluxes to and from the land and oceans (Rayner et al., 2008). Early studies were reviewed by Enting (2002) and much subsequent work has been undertaken within the TransCom intercomparison community (e.g. Gurney et al., 2003; Baker et al., 2006a). Apart from the matrix inversion approach used in early Bayesian synthesis inversion studies, a range of other computational techniques have been applied to flux inversions, including sequential estimation with the Kalman filter (Baker, 2001; Bruhwiler et al., 2005; Peters et al., 2005) and variational data assimilation (Baker et al., 2006b).

In the bottom-up approach, precursors to model–data fusion in studies of carbon exchange between the atmosphere and the land biosphere by plant scientists before the 1970s were largely limited to fitting the response functions of photosynthesis or respiration to environmental variables, such as light, temperature, atmospheric vapor pressure or soil moisture using data collected from controlled environments (Harley et al., 1992). Questions were raised how useful those complex models really were if many of their parameters and underlying assumptions cannot be verified (Finnigan and Raupach, 1988). Because of the complex response of processes to multiple environmental variables, traditional methods of curve fitting became quite limited for extracting information from the data and therefore for improving the accuracies of model predictions.

Technology advances in the late 1980s made the continuous measurement of carbon fluxes in the field possible. As a result, the predictions of process-based models can be compared directly with field measurements over multiple years at different sites. Some early examples of parameter estimation in terrestrial carbon models using these measurements were Wang et al. (2001) and White and Luo (2002) using gradient-based parameter estimation methods. Other methods were later used for estimating parameters, such as Kalman filter techniques (Williams et al., 2005; Gove and Hollinger, 2006); Markov Chain Monte Carlo sampling method (Braswell et al., 2005; Richardson and Hollinger, 2005), Generalized Likelihood Uncertainty Estimation (Mo and Beven, 2004) and so on.

The top-down approach for estimating CO_2 fluxes has a number of limitations: the inversion is poorly constrained and is diagnostic, therefore does not readily allow for prediction. Furthermore, including other kinds of observations is difficult in synthesis inversions except as prior constraints on fluxes (formally this is not a limitation, but in practice calculations have been restricted to linear relations with Gaussian statistics). Much more information can be readily used in the bottom-up approaches using process models, such as biomass inventory, eddy fluxes and remote sensing. However, the bottom-up approach is generally used to calibrate models at individual sites, and it is difficult to know how representative these sites are, and therefore how accurate fluxes are when integrated over larger regions. A series of studies combining the top-down and bottom-up approaches, known as the Carbon Cycle Data Assimilation System (CCDAS), addresses these