Agricultural and Forest Meteorology 189–190 (2014) 175–186

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Uncertainty in model parameters and regional carbon fluxes: A model-data fusion approach

Jingfeng Xiao[a,*], Kenneth J. Davis[b], Nathan M. Urban[c,d], Klaus Keller[c,e]

[a] Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH 03824, USA
[b] Department of Meteorology, Pennsylvania State University, University Park, PA 16802, USA
[c] Department of Geosciences, Pennsylvania State University, University Park, PA 16802, USA
[d] Energy Security Center, Los Alamos National Laboratory, Los Alamos, NM, USA
[e] Earth and Environmental Systems Institute, Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 31 July 2013
Received in revised form
19 December 2013
Accepted 27 January 2014

Keywords:
Uncertainty
Model-data fusion
Data assimilation
Ecosystem model
Carbon cycle
Light use efficiency

ABSTRACT

Models have been widely used to estimate carbon fluxes at regional scales, and the uncertainty of modeled fluxes, however, has rarely been quantified and remains a challenge. A quantitative uncertainty assessment of regional flux estimates is essential for better understanding of terrestrial carbon dynamics and informing carbon and climate decision-making. We use a simple ecosystem model, eddy covariance (EC) flux observations, and a model-data fusion approach to assess the uncertainty of regional carbon flux estimates for the Upper Midwest region of northern Wisconsin and Michigan, USA. We combine net ecosystem exchange (NEE) observations and an adaptive Markov chain Monte Carlo (MCMC) approach to quantify the parameter uncertainty of the Diagnostic Carbon Flux Model (DCFM). Our MCMC approach eliminates the need for an initial equilibration or "burn-in" phase of the random walk, and also improves the performance of the algorithm for parameter optimization. For each plant functional type (PFT), we use NEE observations from multiple EC sites to estimate parameters, and the resulting parameter estimates are more representative of the PFT than estimates based on observations from a single site. A probability density function (PDF) is generated for each parameter, and the spread of the PDF provides an estimate of parameter uncertainty. We then apply the model with parameter PDFs to estimate NEE for each grid cell across our study region, and propagate the parameter uncertainty through simulations to produce probabilistic flux estimates. Over the period from 2001 to 2007, the mean annual NEE of the region was estimated to be $-30.0\,\mathrm{Tg\,C\,yr^{-1}}$, and the associated uncertainty as measured by standard deviation was $\pm 7.6\,\mathrm{Tg\,C\,yr^{-1}}$. Uncertainty in parameters can lead to a large uncertainty to estimates of regional carbon fluxes, and our model-data approach can provide uncertainty bounds to regional carbon fluxes. Future research is needed to apply our approach to more complex ecosystem models, assess the usefulness, validity, and alternatives of the PFT and vegetation type concepts, and to fully quantify the uncertainty of regional carbon fluxes by incorporating other sources of uncertainty.

1. Introduction

Terrestrial ecosystems play an important role in regulating carbon dioxide (CO_2) concentrations in the atmosphere through photosynthesis and respiration. The quantification of ecosystem carbon fluxes over regions can improve our understanding of the feedbacks between the terrestrial biosphere and the atmosphere in the context of global change. Ecosystem models have been widely used to estimate carbon fluxes over various spatial and temporal scales (McGuire et al., 2001; Nemani et al., 2003; Bond-Lamberty et al., 2005; Xiao et al., 2009). Despite numerous modeling studies, the assessment of the uncertainty in modeled fluxes has been overlooked. Uncertainty in future terrestrial ecosystem carbon exchange also remains one of the dominant uncertainties in prediction of future climatic change (Friedlingstein et al., 2006; IPCC, 2007). A better understanding of the uncertainty in carbon dynamics is essential for sound climate and decision-making (Ascough et al., 2008).

The overall uncertainty of modeled fluxes is from three sources of uncertainty (Beck, 1987; Verbeeck et al., 2006): uncertainty of input variables (e.g., imperfect land cover or climate data), uncertainty of model structure (e.g., incomplete or flawed underlying processes and assumptions), and uncertainty of model parameters (e.g., imperfectly or poorly defined parameters due to lack of information). For example, the uncertainty in land cover maps can

* Corresponding author. Tel.: +1 603 862 1873.
E-mail address: j.xiao@unh.edu (J. Xiao).

0168-1923/$ – see front matter
http://dx.doi.org/10.1016/j.agrformet.2014.01.022

lead to significant uncertainty in modeled carbon fluxes at regional scales (Quaife et al., 2008; Xiao et al., 2011). The uncertainty of model parameters can also lead to significant uncertainty in carbon fluxes. Model-data fusion (or data assimilation) approaches are increasingly used to estimate or optimize the parameters of carbon cycle models (Braswell et al., 2005; Knorr and Kattge, 2005; Mahadevan et al., 2008; Ricciuto et al., 2008a; Wang et al., 2009; Peng et al., 2011; Xiao et al., 2011). These approaches typically use carbon flux observations from eddy covariance (EC) towers to estimate parameters.

Model-data fusion approaches can be grouped to two categories: non-Bayesian and Bayesian methods. Non-Bayesian methods typically produce a best estimate for each parameter, but do not quantify its uncertainty (Mahadevan et al., 2008; Xiao et al., 2011). By contrast, Bayesian methods such as Markov chain Monte Carlo (MCMC) generate a probability density function (PDF) for each parameter, and the spread of the distribution as measured by standard deviation provides an estimate of the parameter uncertainty (Braswell et al., 2005; Ricciuto et al., 2008a). The ranges and uncertainty estimates resulting from MCMC will not only facilitate the comparison of parameter values across both plant functional types (PFTs) and sites but also lead to probabilistic estimates of carbon fluxes. Few studies, however, have propagated parameter uncertainty through model simulations and quantified the uncertainty of regional carbon fluxes (Rayner et al., 2005).

Previous model-data fusion studies typically use data from a single site to constrain parameters for a given PFT (Braswell et al., 2005; Knorr and Kattge, 2005; Mahadevan et al., 2008). This is similar to the conventional parameterization of ecosystem models that use a single value for each parameter and do not consider the uncertainty and/or variability of the parameter. Despite numerous studies on parameter estimation, few studies have assessed the variability of parameters within a given PFT and across PFTs (Groenendijk et al., 2011; Xiao et al., 2011).

The Upper Midwest region of northern Wisconsin and Michigan, USA, is a highly heterogeneous mixture of upland forests and lowland wetlands (Fig. 1). This region has, to our knowledge, the highest density of EC flux sites of any region in the world (Desai et al., 2008) as a result of the Chequamegon Ecosystem-Atmosphere Study (ChEAS) and related projects (http://www.cheas.psu.edu). The ChEAS began with flux tower measurements collected at the WLEF tall tower (Davis et al., 2003). Over the period 1998–2006, eddy flux tower systems were deployed in 18 different sites spanning a range of ecosystem types and stand ages, including the regionally representative 447-m tall tower. These landscape-scale flux measurements have spawned a variety of complementary research, including the initiation of three additional long-term flux tower records in the region (Cook et al., 2004; Desai et al., 2005; Sulman et al., 2009), deployment of nearly ten shorter-term flux measurements (Noormets et al., 2008; Xiao et al., 2011), and incorporation of a nearby AmeriFlux tower (University of Michigan Biological station, or UMBS) into this regional network (Gough et al., 2008).

The high density of EC flux sites in the ChEAS region makes it a unique test bed for the assessment of parameter uncertainty and the associated uncertainty in regional carbon fluxes. Previous studies have used model-data fusion techniques to estimate model parameters in the region (Ricciuto et al., 2008a; Desai, 2010; Xiao et al., 2011). One of the studies upscaled carbon fluxes from EC towers to the regional scale, and showed that the uncertainty in land cover maps could lead to significant uncertainty in regional carbon fluxes (Xiao et al., 2011). However, no study has quantified the uncertainty of regional carbon fluxes by propagating parameter uncertainty through simulations.

Here we used net ecosystem exchange (NEE) observations from EC flux sites across the ChEAS region, a simple diagnostic carbon flux model, and a model-data fusion approach to quantify parameter uncertainty and the associated uncertainty of modeled carbon fluxes. We used a Markov chain Monte Carlo (MCMC) approach to quantify parameter uncertainty, and propagated the uncertainty through the regional prediction to produce probabilistic estimates of carbon fluxes and to quantify the associated uncertainties. This work contributes to the development of methods for obtaining regional CO_2 flux estimates with uncertainty bounds.

2. Data

2.1. Flux tower data

Northern Wisconsin and Michigan, USA (Fig. 1) is an area of temperate/sub-boreal forests and glaciated landforms with many small glacial lakes and wetlands. The majority of upland forests consist of mature northern hardwood forests (e.g., maple, basswood, birch, and ash) and younger fast-growing aspen (*Populus termulouides*) forests; coniferous species include red pine, jack pine, eastern hemlock and white pine forests cover smaller areas (Desai et al., 2008). Around 1/3 of the region is lowland wetlands, including forested wetlands (e.g., black spruce, white cedar or tamarack), shrub wetlands (alder or willow species), and open meadows (Desai et al., 2008). The land cover is highly heterogeneous and the region has been relatively densely sampled with EC flux towers. The land area of our study region (Fig. 1) is approximately 1.13×10^5 km^2.

We used NEE data from 13 EC flux sites in the ChEAS region, with 12 sites in northern Wisconsin and the Upper Peninsula of Michigan, and 1 site in the Lower Peninsula of Michigan (Table 1; Fig. 1). These sites involve different PFTs: deciduous forests (DF, 3 sites), evergreen forests (EF, 6 sites), mixed forests (MF, 3 sites), and woody wetlands (WW, 1 site). We used these data to optimize the parameters of a simple diagnostic model. These data have been described in previous publications (Cook et al., 2004; Desai et al., 2005; Gough et al., 2008; Noormets et al., 2008; Sulman et al., 2009; Xiao et al., 2011). We used the half-hourly data for all sites, and only days with no less than 75% of original half-hourly measurements were used in our analysis. Measurements of aboveground biomass and micrometeorological data including air temperature and photosynthetically active radiation (PAR) were also obtained for each site.

2.2. MODIS data

We used vegetation indices (MOD13A2) (Huete et al., 2002) and surface reflectance (MOD09A1) (Vermote and Vermeulen, 1999) derived from the Moderate Resolution Imaging Spectroradiometer (MODIS). For each EC site, MODIS ASCII subsets (Collection 5) for both products were obtained from the Oak Ridge National Laboratory's Distributed Active Archive Center (ORNL DAAC). The subsets consist of 7 km × 7 km regions centered on the flux tower. For each variable, we extracted average values for the central 3 km × 3 km area to better represent the flux tower footprint (Xiao et al., 2008). For each variable, the quality of the value of each pixel within the area was determined using the quality assurance (QA) flags included in the product. At each time step, we averaged the values of each variable using the pixels with good quality within the area to represent the values at the flux site. If none of the values within the 3 km × 3 km area was of good quality, the period was treated as missing.

For regional prediction of NEE, we obtained vegetation indices and surface reflectance from the Earth Observing System (EOS) Data Gateway for each 8-day interval over the period 2000–2007. For each variable and for each 8- or 16-day period, two tiles (1200 km × 1200 km) were needed to cover the ChEAS region. For