The use of forest stand age information in an atmospheric CO$_2$ inversion applied to North America

F. Deng$^{1,*}$, J. M. Chen$^1$, Y. Pan$^2$, W. Peters$^3$, R. Birdsey$^2$, K. McCullough$^2$, and J. Xiao$^4$

$^1$Department of Geography and Program in Planning, University of Toronto, Toronto, ON, Canada
$^2$US Forest Service, Northern Global Change Program, Newtown Square, PA, USA
$^3$Wageningen University and Research Centre, Wageningen, the Netherlands
$^4$Earth Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH, USA
* now at: Department of Physics, University of Toronto, Toronto, ON, Canada

Correspondence to: F. Deng (dengf@geog.utoronto.ca)

Received: 30 January 2013 – Published in Biogeosciences Discuss.: 11 March 2013
Revised: 2 July 2013 – Accepted: 3 July 2013 – Published: 8 August 2013

Abstract. Atmospheric inversions have become an important tool in quantifying carbon dioxide (CO$_2$) sinks and sources at a variety of spatiotemporal scales, but associated large uncertainties restrain the inversion research community from reaching agreement on many important subjects. We enhanced an atmospheric inversion of the CO$_2$ flux for North America by introducing spatially explicit information on forest stand age for US and Canada as an additional constraint, since forest carbon dynamics are closely related to time since disturbance. To use stand age information in the inversion, we converted stand age into an age factor, and inverted without the age constraint. While the carbon flux at the continental and subcontinental scales is predominantly determined by atmospheric CO$_2$ observations, the age constraint is shown to have potential to improve the inversion of the carbon flux distribution among subcontinental regions, especially for regions lacking atmospheric CO$_2$ observations.

1 Introduction

The atmospheric concentration of carbon dioxide (CO$_2$) is rising, leading to a positive radiative forcing of climate and a systematic increase in surface air temperature. Observations of global atmospheric CO$_2$ concentration have been one of the most important datasets for quantifying and understanding the global carbon cycle. Before the 1990s, scientists used CO$_2$ measurements from a limited number of sites to analyze the temporal variability of the global carbon cycle. As more sites were added to the global observation network, inverse techniques were developed to understand both temporal and spatial distributions of carbon exchange across the globe. However, current observation sites are still sparse relative to the size of the global surface, limiting the number of regions that can be robustly inverted. Most inverse studies solved for $\sim$20 source regions (e.g., Enting et al., 1995; Rayner et al., 1999; Fan et al., 1998; Gurney et al., 2003, 2004; Bousquet et al., 2000; Baker et al., 2006; Law et al., 2003b; Patra et al., 2005; Bruhwiler et al., 2011), but these large-region inversions could lead to large spatial aggregation errors due to surface heterogeneity (Kaminski et al., 2001), except for several inversions conducted on systems...
with many more regions (Houweling et al., 1999; Kaminski et al., 1999; Rödenbeck et al., 2003; Michalak et al., 2005; Peylin et al., 2005; Peters et al., 2007). Although solving for more regions potentially allows more information to be extracted from the CO$_2$ concentration observations, it can also introduce large biases if individual regions remain unconstrained. To reduce these biases, a prescribed spatial relationship in flux error covariance is often used as an extra constraint in solving the large number of unknown fluxes (e.g., Rödenbeck et al., 2003; Michalak et al., 2005; Peylin et al., 2005; Peters et al., 2007, 2005; Zupanski et al., 2007; Schuh et al., 2009).

Deng et al. (2007) used several continental sites to pursue a detailed CO$_2$ inversion for North America, and experimented with the use of two meteorological parameters, temperature and precipitation, to further constrain the spatial patterns of CO$_2$ fluxes across the continent. As a result, the inverted fluxes of a local region containing observation site(s) and a region without observation sites under similar climate are constrained with imposed spatial links; the deduced fluxes and their spatial variations will be closer to our intuitive expectation. This effect is particularly apparent when the simulated concentration deviates significantly from the local CO$_2$ concentration observations. The inferred fluxes, however, could be biased due to the fact that large diurnal variations of the planetary boundary layer (PBL) at continental sites cause large diurnal variations of CO$_2$ concentrations, while a transport model without considering the diurnal variations is used to simulate the atmospheric CO$_2$ concentrations. To reduce these biases, we implemented an inversion considering the diurnal variations of the surface CO$_2$ exchange of the terrestrial ecosystems and the atmospheric boundary layer dynamics (Deng and Chen, 2011). However, the traditional diagonal a priori uncertainty matrix was used in these schemes, and the spatial correlation between regions was not considered.

Carbon sinks and sources on land have been attributed to two types of mechanisms (Houghton, 2007): physiological (or metabolic) and demographic (or disturbance). Physiological mechanisms driven by variations in climate, elevated atmospheric CO$_2$, and increased nutrient inputs from air pollution have been intensively studied to estimate the possible responses of terrestrial ecosystems to global climate change and to attribute responses to specific factors. Over the short term, climate is the main driving force of the diurnal, seasonal and interannual variations of the terrestrial carbon cycle (Richardson et al., 2007; Dragoni et al., 2011), and the observed variability in carbon sink strength can often be explained by the anomalies in climatic conditions (Dunn et al., 2007; Desai, 2010). These anomalies tend to affect photosynthesis and respiration through physiological mechanisms (Luyssaert et al., 2007; Richardson et al., 2007), and hence change the net carbon exchange between terrestrial ecosystems and the atmosphere. These physiological mechanisms controlling the carbon cycle have been considered in a global atmospheric inversion by using a priori carbon flux with diurnal and seasonal variations simulated by an ecosystem model that includes these mechanisms (Deng and Chen, 2011).

In addition to climate, forest disturbances in the form of discrete events greatly alter the carbon cycle at various spatial and temporal scales (Liu et al., 2011). These discrete events, however, have not been described appropriately in a mathematical model. The forest carbon cycle is closely associated with the forest life cycle, which can last for several hundred years. Stand age or successional stage is a critical factor in determining forest carbon storage and fluxes (Turner et al., 1995; CasperSEN et al., 2000; Schulze et al., 2000; Law et al., 1999, 2003a, 2001; Chen et al., 2003, 2002; Thornton et al., 2002; Pregitzer and Euskirchen, 2004; Bond-Lamberty et al., 2004; Kurz and Apps, 1999). At the regional scale, forests include stands of trees in various stages of development as a direct reflection of the history of past disturbances and management practices, and thus the carbon balance of a region is the aggregation of the carbon balance of the individual forest stands and non-forest land covers. Apps et al. (2006) pointed out that the change in ecosystem structures, especially the age-class structure of forests, is at least as important as the change driven by physiological mechanisms. Recently, observations of carbon fluxes derived from the Fluxnet-Canada network using the eddy covariance technique have also reinforced the strong relationships between forest age and carbon fluxes (e.g., Coursolle et al., 2006; Zha et al., 2009).

The first continental forest stand age product for North America derived from forest inventory data, large fire polygons, and remotely sensed data (Pan et al., 2011) provides a new source of information to re-evaluate the carbon sources and sinks across the continent as affected by age-class distributions and ecosystem productivity. Forest stand age information has never been incorporated into atmospheric CO$_2$ inversions due to lack of global or continental-scale data, and thus using this continental stand age map provides an opportunity to add an additional constraint to atmospheric CO$_2$ inversions. Our research questions are as follows. (1) How can we reasonably integrate stand age information into an atmospheric inversion scheme? (2) Is the stand age information useful to improve atmospheric CO$_2$ inversion? We approach these questions using the spatially explicit information of forest stand age (Pan et al., 2011) to construct the a priori carbon flux covariance matrix to account for the spatial correlation of prior flux errors between regions. Inversions are conducted with and without considering stand age in order to examine the effects of stand age on inverted fluxes. The usefulness of including forest age information in our inversion is explored through analyzing the spatial correlation of the inverted fluxes with an independent flux field over temperate North America derived from site-level eddy covariance flux data and coast-to-coast satellite data from the Moderate Resolution Imaging Spectroradiometer (MODIS) (Xiao et al., 2008, 2011).
2 Transport model and inversion technique

2.1 Forward modeling framework

2.1.1 Global surface fluxes

There are four major reservoirs of carbon that affect the global carbon cycle on the timescale of seconds to millennia: the atmosphere, oceans, reserves of fossil fuels, and terrestrial ecosystems including vegetation and soil. On the timescale of years to centuries, the carbon exchanges between three of the reservoirs (oceans, fossil fuels, and terrestrial ecosystems) and the atmosphere determine the CO$_2$ content in the air. We use four surface flux datasets to simulate the atmospheric CO$_2$ concentration in a forward modeling framework: (i) the fossil fuel emission field (http://carbontracker.noaa.gov), which is constructed based on the global, regional and national fossil fuel CO$_2$ emission inventory from 1871 to 2006 (CDIAC) (Marland et al., 2009) and the EDGAR 4 database for the global annual CO$_2$ emission on a 1° × 1° grid (Olivier et al., 2005); (ii) the hourly terrestrial ecosystem exchange produced by the Boreal Ecosystem Productivity Simulator (BEPS) (Chen et al., 1999), which is driven by NCEP/NCAR reanalysis data (Kalnay et al., 1996) and remotely sensed leaf area index (LAI) (Deng et al., 2006), and for which a special treatment is implemented to neutralize the annual flux at each grid cell; (iii) the flux of CO$_2$ across the air–water interface based on the results of daily CO$_2$ fluxes using the OPA-PISCES-T model forced by daily wind stress and heat and water fluxes from the NCEP reanalyzed data for 2000 to 2007 (Buitenhuis et al., 2006); and (iv) the monthly mean fire emission available from the Global Fire Emissions Database version 2 (GFEDv2) (Randerson et al., 2007; van der Werf et al., 2006).

2.1.2 Atmospheric transport modeling

The terrestrial carbon flux exhibits strong seasonal and diurnal variations. The atmospheric CO$_2$ concentration over land is characterized by substantial seasonal and diurnal variations as a result of the temporal covariances between the atmospheric transport and the surface flux, producing a rectifier effect (Denning et al., 1996a, 1995, 1996b). In order to include the diurnal CO$_2$ variations in the simulations, an atmospheric transport model that can simulate the diurnal cycle of the atmosphere must be used. TM5 (global chemistry-transport model) (Krol et al., 2005, 2003) is a model that is able to accommodate diurnal variations in PBL. TM5 used in this study is a transport-only version of the TM5 chemistry transport model and is a fully linear operator on CO$_2$ fluxes. Tracer transport (advection, vertical diffusion, cloud convection) in TM5 is driven by offline meteorological fields taken from the European Centre for Medium Range Weather Forecast (ECMWF) model. All physical parameterizations in TM5 are kept as close as possible to the ECMWF formulation to achieve compatibility between them. TM5 offers the possibility to use online two-way nested grids, giving it regional-scale capabilities in a global framework. In this project, we define a global 6° × 4° grid with a nested grid within that focuses on North America at 3° × 2° based on Peters et al. (2004, 2005). The model transport of TM5 has been extensively evaluated using $^{222}$Rn and SF$_6$ (Krol et al., 2005; Peters et al., 2004), and TM5 performs consistently well in atmospheric transport model intercomparisons (Law et al., 2008; Patra et al., 2008).

We use the four surface fluxes introduced above from the terrestrial ecosystem, ocean, fossil fuel burning, and biomass burning as input to the forward transport model to simulate atmospheric CO$_2$ concentration. However, there are substantial differences between the simulated concentration and the observed atmospheric CO$_2$ concentration. These differences indicate that large biases exist in the input surface fluxes, and they often serve as the driving forces to inversely calculate optimal surface fluxes.

2.2 Inverse modeling framework

Inverse modeling techniques compare simulated CO$_2$ concentrations in the atmosphere with observations at discrete sites over the globe. The spatial and temporal distributions of the difference between the simulated and observed values are used to infer the spatial and temporal patterns of the carbon flux. In Bayesian synthesis inversion (Tarantola, 1987), the following objective function is employed:

\[
J = \frac{1}{2} (\textbf{Ms} - \textbf{c})^T \textbf{R}^{-1} (\textbf{Ms} - \textbf{c}) + \frac{1}{2} (\textbf{s} - \textbf{s}_p)^T \textbf{Q}^{-1} (\textbf{s} - \textbf{s}_p),
\]

where \(\textbf{M}\) is a matrix representing a transport (observation) operator; \(\textbf{c}\) is a vector of the observations; \(\textbf{s}\) is an unknown vector of the carbon fluxes of all regions to be inverted at monthly time steps combined with the assumed initial well-mixed atmospheric CO$_2$ concentrations; \(\textbf{s}_p\) is the a priori estimate of \(\textbf{s}\); and uncertainties of \(\textbf{c}\) and \(\textbf{s}_p\) are expressed in covariance matrices \(\textbf{R}\) and \(\textbf{Q}\).

A total of 3000 forward simulations were conducted in this study to form the transport (observation) operator (\(\textbf{M}\)) of five years (2000–2004) for the 50 regions (Fig. 1). For each month and region, a flux of 1 Pg C is prescribed in the TM5 model for forward transport computation to determine the contribution of each region to the CO$_2$ concentration at each observation site since January of 2000. There are strong diurnal variations in boundary layer dynamics, surface fluxes, and observed CO$_2$ concentrations over land, and the correlation among them could have a considerable effect on the inversion result (Denning et al., 1996a). To handle these diurnal variations appropriately in a monthly inversion, we sampled the simulations to construct a transport (observation) operator for continental sites. For tower sites, afternoon hours were sampled to avoid the error-prone simulations under stable and stratified nocturnal atmospheric conditions.
Fig. 1. An inversion scheme: 30 regions in North America and 20 regions for the rest of the globe. Locations of 209 CO$_2$ observational sites are also indicated. Numbers and green dot circles stand for regions and observational sites, respectively.

near the ground. For non-tower continental sites, the sample collection times for discrete observation records associated with the CO$_2$ dataset described in the next section were used in a weighted fashion to modify the monthly transport (observation) operator ($M$). The same weighting approach was also used to sample simulated CO$_2$ concentrations from prior fluxes of fossil fuel combustion, biosphere, ocean, and biomass burning, respectively, as used in Eq. (2) below.

3 Data and methods

As mentioned earlier, currently, the number of atmospheric CO$_2$ observation sites is still small relative to the size of the global surface. Moreover, these observation sites are unevenly distributed over the globe. Therefore, the inversion of CO$_2$ flux is underdetermined (Deng et al., 2007). Additional constraints, such as a priori flux uncertainties, are needed to further constrain the inverse problem. In this section, we will first describe all the inputs used in this study, and then devise a new a priori uncertainty matrix based on the forest age map for North America (Pan et al., 2011).
3.1 CO₂ concentration data and model–data mismatch error

The GLOBALVIEW-CO₂ data (GLOBALVIEW-CO₂, 2008) were used to produce the monthly CO₂ concentration data. The GLOBALVIEW-CO₂ consists of both extrapolated and interpolated data. We selected the synchronized and smoothed values of actual observations to compile our CO₂ concentration dataset from 209 sites across the globe (Fig. 1) over the period from July 2001 to December 2004. Based on this CO₂ concentration dataset and the forward modeling discussed in Sect. 2.1, the vector $c$ in Eq. (1) can be further expressed as

$$c = c_{\text{obs}} - c_{\text{ff}} - c_{\text{bio}} - c_{\text{ocn}} - c_{\text{fire}} ,$$

where $c_{\text{obs}}$ is the monthly CO₂ concentration derived from GLOBALVIEW-CO₂; $c_{\text{ff}}$, $c_{\text{bio}}$, $c_{\text{ocn}}$, and $c_{\text{fire}}$ are simulated CO₂ concentrations from fluxes of fossil fuel combustion, biosphere, ocean, and biomass burning, respectively.

The model–data mismatch includes errors associated with observations (instrument errors) and errors from modeling the observations. Various approaches (Rayner et al., 1999; Rödenbeck et al., 2003; Baker et al., 2006; Gurney et al., 2004; Michalak et al., 2005; Bruhwiler et al., 2011) have been used to determine the model–data mismatch. We used a method based on categorization of sites similar to Baker et al. (2006). We divided the observation sites into five categories, each with its own assigned constant ($c_{\text{const}}$) and variable portion (GVsd) that is computed monthly from the standard deviation of the residual distribution of the average monthly variability (var) files of GLOBALVIEW-CO₂ 2008. Then the mismatch error covariance as a diagonal matrix, $R$, can be defined with the error variance of month $i$ by

$$R_{ii} = \sigma_{\text{const}}^2 + \text{GVsd}_i^2 .$$

The categories and respective $\sigma_{\text{const}}$ are Antarctic sites (0.15 ppm), oceanic sites (0.30 ppm), land and tower sites (1.25 ppm), mountain sites (0.90 ppm), and aircraft samples (0.75 ppm).

3.2 Prior fluxes and uncertainties

The prior flux estimates for both oceans and lands in this study were set to zero after we subtracted the modeled contributions to the CO₂ concentration from fossil fuel combustion, biosphere, ocean and biomass burning (see Eq. 2, also Sect. 2.1). The vector $s$ to be optimized is thus the deviation from the prior fluxes.

The sources of uncertainty in the air–sea CO₂ flux are multifold, including the error of reanalyzed meteorological data, the error of the estimated parameters in the OPA-PISCES-T model, and the error caused by the model structure. Estimating the uncertainties of the model outputs is therefore extremely difficult. The spread (0.5 Pg C yr⁻¹) of ocean uptakes in four ocean models (Le Quéré et al., 2007; Rodgers et al., 2008; Lovenduski et al., 2008; Wetzel et al., 2005) provides a good reference of uncertainty in bottom-up modeling. Rödenbeck et al. (2003) also used 0.5 Pg C yr⁻¹ as his prior global ocean flux uncertainty. Considering that the a priori used in this study is at the upper bound of those four ocean uptakes, we used an uncertainty of 0.67 Pg C yr⁻¹, distributed among 11 ocean regions according to Baker et al. (2006).

The sources of uncertainty in the land surface flux are more complicated than those of the air–sea surface exchange. Fossil fuel emissions have small uncertainties (+6 % of the global fossil fuel emissions; Marland, 2008) that are often not considered in atmospheric inversions. Gurney et al. (2005) demonstrated that ignoring the uncertainties of fossil fuel emissions in time and space could effectively bias the seasonal pattern of surface fluxes and significantly change the distribution of inverse fluxes in a high resolution inversion. The estimates of vegetation fire emissions are uncertain to about 20 % ($\sigma$) on global, annual scales (van der Werf et al., 2010). Similar to fossil fuel emissions, the uncertainties of vegetation fires could not be handled separately in this inversion study. It should be noted that, however, any inverse estimate of the terrestrial carbon flux excluding fire emissions and fossil fuel emissions could be biased by these uncertainties. Compared with the error in the spatiotemporal distribution of emissions from fossil fuel combustion and vegetation fires discussed above, the heterogeneous distributions of carbon assimilation and respiration of terrestrial ecosystems in both space and time are even more error-prone. With these considerations in mind, we used an uncertainty of 2.0 Pg C yr⁻¹ (based on a similar regional scheme of TRANSCOM 3) for the global land surface that is spatially distributed based on the annual NPP (net primary productivity) distribution simulated by BEPS. These are used to specify the diagonal elements of the a priori uncertainty covariance matrix ($Q$), and the off-diagonal elements that provide information about the spatial correlation of the a priori errors between regions are all set to 0.

3.3 Forest stand age information and a priori uncertainties

Forest stand age is a fundamental factor that affects net carbon accumulation by forest ecosystems (Chapin et al., 2002; Bradford et al., 2008; Apps et al., 2006). Forest age is associated with the various forest stand development stages, i.e., young stands shortly after disturbance are likely to be sources of carbon while middle-aged stands are likely to be sinks. Forest stand age information has not yet been explicitly used in global forest carbon cycle studies due to lack of spatially explicit information. As mentioned earlier, a map of stand age for US and Canada has recently been produced (Pan et al., 2011). We used the stand age map to further constrain CO₂ fluxes over North America. The functional relationship between the forest age and NEP (net ecosystem production)
is exceedingly complicated (Chen et al., 2003; Turner et al., 1995; Law et al., 2003a), and therefore we did not directly estimate NEP from the age information. Instead, we used the forest age map as a covariate to describe the relative spatial distribution of carbon sinks and sources among the 30 regions in North America. Chen et al. (2003) developed a general semiempirical mathematical function to describe the relation of NPP with age over large areas as

\[
\text{NPP(age)} = a \left(1 + \frac{b \left(\frac{\text{age}}{c}\right)^d - 1}{\exp\left(\frac{40}{c}\right)} \right),
\]

(4)

where the parameters \(a\), \(b\), \(c\), and \(d\) are functions of the mean annual air temperature. To extract information from stand age while avoiding the calculation of absolute NPP values, we normalized the chronosequence curve against the maximum NPP (i.e., \(\text{NPP}_{\text{max}}\)).

\[
f_s = \frac{\text{NPP(age)}}{\text{NPP}_{\text{max}}},
\]

(5)

where \(\text{NPP}_{\text{max}}\) is the maximum value of NPP(age) for age = 1 to 250 yr. \(f_s\) ranges between 0 and 1, with large values representing mature forests and small values representing young or old forest in terms of NPP value. This normalization accentuates the general pattern of NPP variation with age while the absolute values of NPP under the influence of meteorological conditions (precipitation, temperature and radiation) are modeled by BEPS and included as part of the prior carbon flux. Figure 2 shows the distribution of \(f_s\) for North America.

This ratio was then scaled to a quantity \((f_{ri})\) for each region \(i\) within North America as an area-weighted average of \(f_s\), and for each region \(i\) we defined an age factor as

\[
f(\text{age})_i = f_{ri} - f_c,
\]

(6)

where \(f_c\) is the mean of \(f_i\) for all North American regions. Similar to \(f_s\), the first term \((f_{ri})\) in Eq. (6) also ranges between 0 and 1 with small, medium and large values representing young, old, and mid-aged forests, respectively. The purpose of subtracting \(f_c\) from \(f_{ri}\) is to balance the influences of this consideration between regions with forest age information and regions without forest age information. We introduced this NPP-based age factor into our atmospheric inversion for the net carbon flux because net ecosystem carbon exchange (NEE) variations with age are mostly determined by NPP variations with age (Amiro et al., 2010). Heterotrophic
respiration, which depends not only on the new organic matter transferred to the soil (proportional to current NPP) but also on the amount of soil carbon accumulated over thousands of years, varies much less than NPP with forest stand age (Amiro et al., 2010), although there are some higher values at younger sites.

To combine the regional age factor into the a priori flux uncertainty, we constructed the following exponential covariance matrix for a specific month:

\[
Q = \begin{bmatrix}
\sigma_1^2 & \cdots & \sigma_1 \sigma_2 & \cdots & \sigma_1 \sigma_n \\
\cdots & \ddots & \cdots & \cdots & \cdots \\
\sigma_2 \sigma_1 & \cdots & \sigma_2^2 & \cdots & \sigma_2 \sigma_n \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\sigma_n \sigma_1 & \cdots & \sigma_n \sigma_2 & \cdots & \sigma_n^2
\end{bmatrix}
\]

where \(\sigma_i\), \(f_i\), and \(r_i\) are the standard deviation, the regional age factor, and the percentage of forest cover for region \(i\), respectively; \(l\) is the integral scale, or decorrelation length scale.

The guiding principle in constructing this matrix is that in the nested inversion there are insufficient atmospheric CO\(_2\) stations over North America, and the spatial patterns of the prior surface fluxes could be used to constrain the inversion for small regions. This constraint may be accomplished by using the covariance of the errors between various small regions. Errors in flux estimates for regions with similar age factors are expected to be better correlated than those for regions with dissimilar age factors. This approach has been applied to regions 4 to 28 across North America (Fig. 1) based on the availability of stand age information.

4 Results and discussion

4.1 Inverted annual fluxes

To demonstrate how the integration of the information on forest stand age could affect our inversion results, we ran the inverse modeling with different sets of a priori uncertainty matrixes, i.e., using an a priori diagonal uncertainty matrix without considering any correlation between regions or an a priori uncertainty matrix with off-diagonal covariances based on the similarity of the age factor between regions. Figure 3 shows two sets of the inverted CO\(_2\) fluxes and uncertainties for the 30 North America regions in 2003 using both a priori uncertainty matrixes. The general spatial pattern of the inverted CO\(_2\) flux over the continent is not altered much by the use of the forest age information, as the highest uptake areas are found in the Eastern US from both inversions. There are, however, adjustments in the distribution of sources and sinks induced by the age constraint. These adjustments are not enough to overturn the general spatial pattern driven by the atmospheric constraint. For example, Canada region 10, the forested region of Alberta, and regions 14, 15, 16, and 17 with middle-aged productive forests in southern Ontario and Quebec were inverted with larger carbon uptake. In contrast, regions 11, 12, and 13, where fire disturbances are frequent in recent decades (Flannigan et al., 2006), resulted in smaller sinks when the forest age constraint was applied. The inverted sink size is reduced by about one-fifth in region 9 (British Columbia, Canada), where a large proportion of the forest is old-growth. These changes in the regional sinks are not forced by the a priori fluxes that could be formed as a function of the age distribution. Instead, the sources/sinks are forced to change in the same direction (either increase or decrease) for areas with similar age factors, or change in opposite directions (one increase and another decrease) for areas with dissimilar age factors. The relative changes of inverted fluxes in region 9 and adjacent region 10 are good examples to illustrate this mechanism. As mentioned above, a lot of old forests grow in region 9, while more middle-aged productive forests proliferate in region 10. Accordingly, negative and positive regional age factors are derived for region 9 and region 10, respectively. These age factors transform into the a priori covariance matrix, and consequently, the negative correlation between region 9 and 10 arose from the covariance matrix design (Eq. 7), and this leads region 10 to become a larger sink and region 9 to become a smaller sink. The resulting sink size is reduced in region 9, but the sink strength is still strong because of the constraint of atmospheric CO\(_2\) concentration measurements around the region in the inversion process. Moderate sinks in British Columbia were previously estimated through ecosystem modeling, and climate warming and CO\(_2\) fertilization had positive effects on the forest growth in this region (Chen et al., 2003). Based on the same principle, in the northern USA the sink sizes are reduced in regions with recent disturbances or old-growth forests (e.g., Region 19), or with low forest cover (e.g., Region 20) (Pan et al., 2011); the uptake in the southeastern USA is strengthened in the highly managed forest regions where forests in predominantly productive ages could actively sequester CO\(_2\) from the atmosphere.

Since we have plausible explanations for the changes revealed by applying the forest stand age information to our
inversion results, and these changes conform to our understanding of the relationship between forest stand age and carbon balance, we examined these changes in more detail. We considered separately the physiological mechanisms that could improve the modeling of the seasonal and diurnal cycle, and the demographic mechanisms that could be used to further define the spatial distribution of carbon sinks and sources. The relevant questions are (1) whether the modeled changes alter our understanding of the role of North America in the global carbon budget, and (2) how large are the changes in the spatial distribution of the terrestrial carbon exchange over North America?

A paired t test shows that considering the stand age factor did not significantly change the total flux of all 30 North American regions in 2003 (p = 0.9866). It is understandable that the total annual carbon flux for the whole region, inferred with the stand age factor, does not change significantly from that inferred without considering the stand age factor in the US and Canada. Because the forest stand age information is only available for North America in this application, we centralize $f_t$ by removing $f_c$ (the mean of $f_t$ for all regions) in Eq. (6) as the age factor, to cancel the possible influences on other regions outside of North America where forest age information is not available. In a balanced global carbon cycle, this leads to an insignificant change in the annual total flux for North America as shown in the paired t test.

However, we found that the change in the spatial distribution of the inverted fluxes is much stronger than the change in the annual total carbon uptake. The root-mean-square of the differences of inverted fluxes for the 20 North American regions (from region 9 to region 28) between inversions with and without considering the forest stand age factor is calculated to be $10^{10}$ $\text{Tg yr}^{-1}$. This is 7 times the magnitude of the change in annual mean fluxes, equivalent to 23% of the mean absolute flux of these 20 regions. We also note that despite being one of the most densely observed regions, the problem is still strongly underdetermined when the inversion is conducted at a higher spatial resolution with the same set of observations. The added value of stand age information might increase as the spatial resolution increases. At higher resolutions, it would be more difficult to extract carbon source and sink information from CO$_2$ concentrations alone, and the forest stand age should be seen as complimentary to the CO$_2$ observations to reduce uncertainties in the inverse estimates. However, we need to ensure that the addition of this new dataset keeps our inversion results consistent with other sources of information, and does not introduce irreconcilable NEE estimates. Therefore, we compared them with another set of independent regional estimates of NEE upscaled from eddy covariance flux measurements that were neither used in the inversions nor in the construction of the stand age map.

![Fig. 4. CO$_2$ fluxes and uncertainties for 20 North America regions (region 9 to 28) estimated from three different approaches: blue solid square – the EC-MOD method (Xiao et al., 2008), and no uncertainty information is available; empty square – inversion method without age constraint; red solid square – inversion method with age constraint.](image)

**4.2 Comparison with independent flux estimates**

Eddy covariance flux towers provide continuous measurements of net ecosystem carbon exchange (NEE) for various terrestrial ecosystems, but these measurements only represent the carbon fluxes at the scale of the tower footprint (about 1 km$^2$) and areas that have very similar characteristics. Therefore, the site-specific eddy covariance measurement results are not well suited for validating estimates of regional carbon fluxes. Recently, Xiao et al. (2008, 2011) used a data-driven approach to extrapolate the NEE measurements to large areas using a variety of data streams from MODIS, and developed a gridded NEE dataset with high spatial (1 km) and temporal (8-day) resolutions for temperate North America over the period 2000-2006. This NEE dataset was derived from eddy covariance (EC) and MODIS data, and is therefore referred to as EC-MOD. The EC-MOD dataset covers most of the North American regions (Regions 9–30). At cropland sites where strong CO$_2$ assimilation signals can be captured by the EC measurement but CO$_2$ release from the consumption of the yield off-site would spread to a fairly large area (Ciais et al., 2007; Schuh et al., 2010) that is not measured by EC instruments, we decided to neutralize the croplands grids to avoid overestimation of the carbon uptake. We are aware that this is a drastically simple treatment of a complicated matter, as the percentage of the export of agricultural products out of a region would influence our inversion results for the region. However, for the purpose of this study, this simple treatment provides a useful basis for our analysis of carbon sinks in forested areas.

Figure 4 shows the comparison of the inverted fluxes of North America with and without the age constraint and the EC-MOD flux with neutral croplands of regions 9 to 28 in 2003. It is very encouraging to see that the inversion results and the EC-MOD flux show similar spatial patterns.
F. Deng et al.: The use of forest stand age information in an atmospheric CO₂ inversion

This figure clearly shows that the large sink regions of EC-MOD coincide with our inverted sink centers. These two estimates exhibit similar magnitudes in northern regions (regions 9-17). In the northeastern and southeastern US, however, EC-MOD sinks are much larger than inversion results. These large discrepancies may be caused by the fact that the eddy covariance measurement sites are usually located at productive, undisturbed vegetation areas, and measured carbon fluxes are larger than the average over the landscape. In addition, even though MODIS data are used for the spatial extrapolation that can capture the spatial variability to a certain extent, the carbon dynamics associated with forest age are not explicitly considered in the spatial extrapolation (Xiao et al., 2008, 2011). An old forest stand, for example, would have similar leaf area to a middle-aged stand but have more biomass and higher autotrophic respiration. In other words, the EC-MOD product likely has included some age effect with respect to aboveground productivity, and this makes it worthwhile for comparison with an atmospheric inversion that implements an age constraint.

Further comparisons of our inversions with EC-MOD flux estimates show that the inversion with age constraint is better correlated with EC-MOD fluxes than the inversion without age constraints (Fig. 5). The age constraint increases the \( R^2 \) value by 26\%. The increase in slope (11\%) towards unity and the decrease in the intercept (4\%) of the regression (Fig. 5b) may indicate improvements achieved through using age information in the atmospheric inversion. It should be noted that, however, this may only expose one aspect of a complicated problem. Most of the eddy flux sites may also be selected at an age around their peak carbon uptake. This likely preference in site selection could positively bias the EC-MOD sink estimation when the forest age information is not used in the spatial extrapolation from the site measurements.

Several regions exhibited large discrepancies between our inversion results and EC-MOD fluxes. For example, our inversion results show that drought-dominated regions 23, 24, and 25 are nearly carbon neutral, while EC-MOD flux estimates indicated sources for regions 24 and 25, and a minor sink (0.03 Pg yr\(^{-1}\)) for region 23. These discrepancies may imply that the CO₂ concentration observations used are not sufficient to constrain our inversion results for these regions, and it is possible that mutual compensating effects may exist between regions 23, 24 and 25, and between region 25 and regions to the east. These hypotheses are consistent with the information revealed from the a posteriori error covariance matrix produced in our inversion calculation. We found that a posteriori error for region 23 negatively correlates with those for regions 24 and 25, while regions 24 and 25 are positively correlated and region 25 is negatively correlated with regions 26, 27, and 28.

4.3 Further remarks

The inversion approach to estimate carbon fluxes relies on the observations of CO₂ concentration. The observations integrate information on many aspects, including both temporal and spatial distributions of surface CO₂ exchanges. Therefore, the source and sink problem of CO₂ can be solved easily if we have enough highly precise CO₂ observations and a perfect atmospheric transport model. North America is one of the most densely observed regions, and the spatial distribution of the surface carbon exchange is better constrained there than over other continents where observations are sparse. Though the incorporation of stand age in North America only slightly improves the inversion estimates of net carbon flux, our study suggests that adding forest stand age information could bring more pronounced improvements to the inverse estimates of carbon fluxes over other less observed continents. Further improvement could also be expected if we pursue higher resolution inversions over North America (or Europe) as discussed in Sect. 4.1.

This study opens the way for further development of the approach established in this paper. We used a simple function (Eq. 4) to capture the basic relationship between NPP
and forest stand age. Although the normalized NPP–age relationship includes the influence of temperature, other site conditions, such as water availability might also have an influence on the relationship. Further, it is obvious that a NEP function with stand age would be a better covariate of the full forest carbon cycle than the NPP–age relationship used in this study. The correlations we established may only reflect the dominant association among regions, and a general NEP function needs to be developed in the future, along with novel ways to incorporate this relationship into the dual-data inversion approach. The use of forest age information could be extended to the global land surface and to a multi-year inversion as soon as the map covering the global land surface is available. In the meantime, we can examine this dual-data inversion approach at a finer resolution in a region where more CO$_2$ observations and forest stand age information are both available. Although the influence of forest stand age information on the inverted carbon flux distribution could be large or small because it is complimentary information to the atmospheric CO$_2$ observations, the uncertainties of the inverted fluxes would certainly be decreased as more constraints are incorporated. We can, therefore, gain increased confidence from the dual-data inversion approach.

However, it is necessary to point out the limitations of the forest age factor method. The demographic variations associated with forest disturbance and regrowth could be considered as a low-frequency force superimposed on the physiological effects mostly determined by both short-term and long-term variations in climate. Considering the forest age factor itself could help us to resolve the spatial distribution of carbon sources and sinks on average, but hardly improves our knowledge of the seasonal variation and interannual variation of regional carbon fluxes, which are critical for us to understand the relationship between climate and the terrestrial carbon cycle. Therefore, developing and improving a terrestrial ecosystem model containing the physiological mechanisms mostly driven by both short-term and long-term variations in climate, and the demographic mechanisms as reflected by changes in the age structure of forests could improve our bottom-up estimates of the spatial and temporal variations in the terrestrial carbon flux. This is because the seasonal and interannual dynamics we are interested in change as a function of forest age. Not describing the age structure correctly can thus lead to biases in the carbon cycle response to climate, and provide biased a priori information for the inverse modeling of the carbon source and sink distribution. The method discussed above could contribute to limiting such biases.

5 Conclusions

The spatially explicit information on forest stand age (Pan et al., 2011) was used to establish the spatial correlation between regions through an a priori covariance matrix, which further constrained the CO$_2$ flux inversion for the subcontinental regions across North America. Our results show no fundamental changes to the aggregated spatial patterns of CO$_2$ fluxes across North America when the stand age information is included in the inversion as the large-scale sinks and sources are basically driven by atmospheric constraints. The use of stand age information in the inversion brings adjustments in the distribution of the inverted fluxes among regions of dissimilar forest ages representing different stand development stages. In recently disturbed or old-growth forest regions, the inverted fluxes often move towards the source direction, while regions with middle-aged productive forests, especially the highly managed southeastern US forests, were nudged toward larger sinks. The a posteriori uncertainties are large in southeastern US regardless of the inclusion of the stand age information.

The relatively high level of agreement of the spatial patterns between our inversions and independent carbon flux estimates (EC-MOD) derived from eddy covariance flux measurements and remotely sensed data increases our confidence in inferring the surface carbon flux through atmospheric inversion. The inverted spatial distribution of the carbon flux with the forest stand age constraint is better correlated with EC-MOD estimates than that without the age constraint ($R^2$ increased from 0.46 to 0.58). Considering the likely bias induced by preferential EC site selection, we would expect that the EC-MOD product could be improved by using forest age information in the spatial extrapolation from sites to a region. A useful direction to pursue in future work would be to use forest age information in both top-down and bottom-up modeling so that the spatial distribution pattern of the terrestrial sink, especially over less observed continents, can be better resolved, as forest stand age has first order effects on the spatial distribution of forest carbon sources and sinks.

Acknowledgements. This work was made possible through the support from a research grant (2010CB950704) from the Ministry of Science and Technology of China, a Canadian Foundation of Climate and Atmospheric Sciences project (grant no. GR-646), a Natural Sciences and Engineering Research Council of Canada (NSERC) scholarship (2007–2010), a grant from the US Forest Service Northern Research Station (07-JV-11242300-114), and a Meteorological Service Canada (MSC) graduate supplement (2007–2009). W. Peters was partly supported by a Dutch Science Foundation (NWO) VIDI grant (no. 864.08.012). J. Xiao was supported by the National Science Foundation (NSF) through MacroSystems Biology (award number 1065777) and the National Aeronautics and Space Administration (NASA) through Carbon Monitoring System (award number NNX11AL32G). The authors thank D. Harvey for his comments and useful discussions. Anonymous referees provided constructive comments that greatly improved this paper.

Edited by: S. Liu
References


www.biogeosciences.net/10/5335/2013/ Biogeosciences, 10, 5335–5348, 2013

