Uncertainties in carbon residence time and NPP-driven carbon uptake in terrestrial ecosystems of the conterminous USA: a Bayesian approach

By XUHUI ZHOU[1,2]*, TAO ZHOU[2,3] and YIQI LUO[1,2], *[1]Coastal Ecosystems Research Station of Yangtze River Estuary, Ministry of Education Key Laboratory for Biodiversity Science and Ecological Engineering, The Institute of Biodiversity Science, Fudan University, 220 Handan Road, Shanghai 200433, People's Republic of China; [2]Department of Microbiology and Plant Biology, University of Oklahoma, Norman, OK 73019, USA; [3]State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing, People's Republic of China*

ABSTRACT

Carbon (C) residence time is one of the key factors that determine the capacity of ecosystem C storage. However, its uncertainties have not been well quantified, especially at regional scales. Assessing uncertainties of C residence time is thus crucial for an improved understanding of terrestrial C sequestration. In this study, the Bayesian inversion and Markov Chain Monte Carlo (MCMC) technique were applied to a regional terrestrial ecosystem (TECO-R) model to quantify C residence times and net primary productivity (NPP)-driven ecosystem C uptake and assess their uncertainties in the conterminous USA. The uncertainty was represented by coefficient of variation (CV). The 13 spatially distributed data sets of C pools and fluxes have been used to constrain TECO-R model for each biome (totally eight biomes). Our results showed that estimated ecosystem C residence times ranged from 16.6 ± 1.8 (cropland) to 85.9 ± 15.3 yr (evergreen needleleaf forest) with an average of 56.8 ± 8.8 yr in the conterminous USA. The ecosystem C residence times and their CV were spatially heterogeneous and varied with vegetation types and climate conditions. Large uncertainties appeared in the southern and eastern USA. Driven by NPP changes from 1982 to 1998, terrestrial ecosystems in the conterminous USA would absorb 0.20 ± 0.06 Pg C yr^{-1}. Their spatial pattern was closely related to the greenness map in the summer with larger uptake in central and southeast regions. The lack of data or timescale mismatching between the available data and the estimated parameters lead to uncertainties in the estimated C residence times, which together with initial NPP resulted in the uncertainties in the estimated NPP-driven C uptake. The Bayesian approach with MCMC inversion provides an effective tool to estimate spatially distributed C residence time and assess their uncertainties in the conterminous USA.

Keywords: carbon residence time, carbon uptake, conterminous USA, inverse analysis, MCMC, terrestrial carbon cycle, uncertainties

1. Introduction

The rising atmospheric CO_2 concentration and the resultant climate change may alter carbon (C) cycles of terrestrial ecosystems (IPCC, 2007), which may, in turn, amplify or dampen climate change (i.e., positive or negative C-climate feedback) (Friedlingstein et al., 2006; Luo, 2007; Heimann and Reichstein, 2008). To better understand responses of terrestrial ecosystems to climate change, it becomes essential to quantify ecosystem C uptake. The C-uptake capacity in plant and soil pools is largely determined by both C residence time and net primary productivity (NPP) changes (Luo et al., 2003). The spatial patterns of NPP changes were relatively well qualified by using ecosystem production models with remote sensing techniques in North America (Hicke et al., 2002). The former, C residence time, is the length of time that a carbon atom can stay in a reservoir from the entrance to the exit (Barrett, 2002; Luo et al., 2003). Several methods have been used to estimate the C residence time, such as C balance method via calculating

*Corresponding author.
email: zxuhui14@fudan.edu.cn

1

ratios of C pools divided by fluxes (Vogt et al., 1996), time series decomposition experiments (Powers et al., 2009), C isotope tracing (Gaudinski et al., 2001) and inverse analysis (Barrett, 2002; Luo et al., 2003; Xu et al., 2006). However, uncertainties in estimated C residence time have not been well quantified. If the issue of uncertainty is not adequately addressed, C sink potentials cannot be fully understood and policies to mitigate climate change based on the current understanding of terrestrial C sequestration will fall short in meeting targets of the Kyoto protocol. Therefore, quantifying uncertainties of C residence times, especially at regional scales, is crucial for an improved understanding of terrestrial C uptake (Randerson et al., 1999; Barrett, 2002).

Most of the uncertainties studies have been conducted by sensitivity analysis with a local design via changing one parameter at a time while keeping the other parameters constant (Kremer, 1983; Knorr, 2000; El Maayar et al., 2002; Larocque et al., 2008) or using minimum and maximum values from the literature (Hallgren and Pitman, 2000; Zaehle et al., 2005) in ecological modelling. Although these approaches can identify uncertainty of a single parameter (Kleijnen, 1998), they neglect possible interactions between parameters and can not specify the confidence intervals for parameter distributions, which are very important in complex ecosystem models (Saltelli et al., 2000). Highly correlated parameter sets and large confidence intervals for parameter estimates would tend to indicate equifinality (i.e., the optimal parameter set is not uniquely defined) (Beven, 2006), and may increase the uncertainties in model prediction as a result of high variation in parameterisation (Schulz and Beven, 2003). In addition, non-linearity of ecosystem models may introduce considerable errors in the flux and pool estimates (Raupach et al., 2005; Zhou et al., 2008).

Only a few studies have so far assessed stochasticity and uncertainties of C residence time with inverse analysis (Wang et al., 2001; Xu et al., 2006; Larocque et al., 2008). For example, Xu et al. (2006) found that only three or four of seven parameters were well constrained by six ecological data sets collected from the Duke Free-Air CO_2 Enrichment (FACE) experiment while the other parameters were unconstrained with large uncertainties. The uncertainties in estimated parameters were propagated to model projections of C sink sizes. On regional and global scales, uncertainties have usually been analysed by inversion of measured atmospheric CO_2 concentration in conjunction with atmospheric transport models (Kaminski et al., 2002; Rayner et al., 2005; Prihodko et al., 2008). The atmospheric inversion can estimate regional C source/sink but cannot resolve the spatial distribution of terrestrial C sink. This is mainly due to the lack of information to enable an orthogonal decomposition of the different types of C source/sink and the uneven distribution of the measure-

ments for atmospheric transport models (Wang et al., 2001; Wang and Barrett, 2003; Ahmadov et al., 2009; Pillai et al., 2010, 2011). Inclusion of multiple sources of information has the potential to improve parameter constraint and reduce uncertainties (Barrett, 2002; Wang and Barrett, 2003; Xu et al., 2006).

Among various inverse techniques, Bayesian inversion may be currently the most common approach to parameter estimation and uncertainties analysis (Enting, 2002; Larocque et al., 2008). It combines information from prior knowledge, model structure and data to allow the simultaneous determination of all parameters. The Markov chain-based Bayesian method can not only better account for non-linearity in the model and correlations among parameters, but also provide more robust estimates of uncertainties in parameters and model predictions by generating a probability distribution (Braswell et al., 2005; Xu et al., 2006; Prihodko et al., 2008). The Bayesian paradigm starts with *prior* probability distributions of parameters and then assimilates measurements into models to generate *posteriori* probability density functions (PDFs) of model parameters and forecasted state variables (Braswell et al., 2005; Knorr and Kattge, 2005; Xu et al., 2006). PDFs may contain information on key uncertain properties of climate change simulations (mainly CO_2 concentration and temperature, Murphy et al., 2004) and ecosystem responses to elevated CO_2 (Xu et al., 2006). A cumulative distribution function (CDF) has also been presented to assess the uncertainties of dangerous anthropogenic interference in model predictions of future climate change (Mastrandrea and Schneider, 2004). On regional and global scales, the Bayesian method, to our knowledge, has not been used for uncertainty analysis with multiple ecological data sets, although it has been applied to constrain parameters and for uncertainties analysis in biosphere models against atmospheric CO_2 concentration data (Kaminski et al., 2002; Rayner et al., 2005; Prihodko et al., 2008).

The objectives of this study were to quantify C residence time and NPP-driven ecosystem C uptake, and their uncertainties in the conterminous USA. We first applied the Bayesian probability inversion and Markov Chain Monte Carlo (MCMC) technique to a terrestrial ecosystem regional (TECO-R) model. The 13 spatially distributed observational data sets of C pools and fluxes have been used to constrain TECO-R model for each biome (totally eight biomes) to generate *posterior* PDFs of 22 parameters. C residence times of ecosystem and each pool were estimated from these parameters for each of the eight biomes. Second, standard deviation (SD) and coefficient of variation (CV = SD/mean) of all estimated parameters and C residence times were calculated from *posterior* PDFs to represent their uncertainties. Finally, the propagation of

parameter uncertainties to modelled ecosystem C uptake was assessed.

2. Materials and methods

2.1. Model description

In this study, a TECO-R model was developed by combining the Carnegie-Ames-Stanford-Approach (CASA) model (Potter et al., 1993; Field et al., 1995) with the Vegetation-And-Soil-carbon-Transfer (VAST) model (Barrett, 2002) to estimate the spatial patterns of C residence times and their uncertainties as measured by SD and coefficient of variation (CV = SD/mean) in the conterminous USA. The details of the TECO-R model are described in Zhou and Luo (2008). Here we provide only a brief overview.

In the TECO-R model, NPP was estimated by the CASA algorithms from satellite observation and ground measurements as below (Potter et al., 1993).

$$\mathrm{NPP} = f_{\mathrm{APAR}} \cdot \mathrm{PAR} \cdot \varepsilon^* \cdot T_\varepsilon \cdot W_\varepsilon \qquad (1)$$

where f_{APAR} is a fraction of photosynthetically active radiation (PAR) that is absorbed by vegetation and determined by using a linear relationship between f_{APAR} and satellite-data derived normalised difference vegetation index (NDVI, Potter et al., 1993). Solar radiation is converted to PAR by multiplying 0.5. ε^* is maximum potential light-use efficiency, and T_ε and W_ε are temperature and moisture scalars, respectively (see Supplementary material). The estimated NPP is allocated to plant tissues of leaves, stems and roots. Dead leaf, stem and root materials are transferred to respective litter compartments. Part of the litter C is respired and the remainder is converted into soil organic matter (SOM) pools. The relationships of C transfer among pools were simulated by the VAST algorithms. The first-order ordinary differential equations were applied to model C dynamics in the litter and soil organic C (SOC) pools (see Supplementary material, Barrett, 2002; Zhou and Luo, 2008). The model consists of two aboveground biomass pools, two litter pools, three root biomass pools and three soil C pools. Root biomass and soil C pools were divided into three soil layers (0–20, 20–50 and 50–100 cm) as in Barrett (2002) and Zhou and Luo (2008), instead of compartmentalised SOC according to decomposition rates as in most of the models (e.g., CENTURY, Parton et al., 1987; TECO, Luo et al., 2003). In this way, a one-to-one relationship was ensured between model state variables (especially for root biomass and SOC) and the respective observations obtained from the literature, which do not need extra mapping function as in Luo et al. (2003). We assumed that impacts of disturbances on C residence time were reflected by remote sensing data in NDVI. A schematic diagram of the TECO-R model is shown in Fig. 1.

The C residence times in litter and SOC pools $\left(\tau_F,\ \tau_C,\ \tau_{S_1},\ \tau_{S_2} \text{ and } \tau_{S_3}\right)$ vary largely with both climatic and biotic factors (Schimel et al., 1994). To specify the biotic effects on the C residence times, eight vegetation types (ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; MF, mixed forest; W, woodland; WG, wooded grassland; S, shrubland; G, grassland; and C, cropland) were set up in the conterminous USA based on the 1-km land cover classification by Hansen et al. (2000), and then the TECO-R model was used to estimate the parameters for each biome, separately. To specify the effects caused by spatial heterogeneity of climatic factors, temperature (T_ε) and moisture (W_ε) scalars were used to correct the site-specific C residence times $\left(\tau_F,\ \tau_C,\ \tau_{S_1},\ \tau_{S_2} \text{ and } \tau_{S_3}\right)$. The corrected residence times were called moisture and temperature corrected C residence times $\left(\tau_F^*,\ \tau_C^*,\ \tau_{S_1}^*,\ \tau_{S_2}^* \text{ and } \tau_{S_3}^*\right)$ as follows:

$$\tau_k^* = \tau_k \times (W_s \cdot T_s),\ k = F,\ C,\ S_1,\ S_2,\ S_3 \qquad (2)$$

The moisture scalar was estimated by monthly precipitation, potential evapotranspiration and soil moisture (Randerson et al., 1999). The temperature scalar was obtained directly from temperature data, as shown in the CENTURY soil-carbon model (see Supplementary material, Parton et al., 1987).

The estimated parameters by MCMC technique in this study included the maximum potential light-use efficiency (ε^*), C allocation coefficients among pools and C residence times in individual plant and soil pools, totalling 22 parameters for each of the eight biomes (see Table 1 for the definition of all 22 parameters). Uncertainties of individual parameters and ecosystem C residence times were represented by SD and CV from Metropolis–Hastings (M–H) simulation (see Appendix S1 in Supplementary material). Ecosystem C residence times (τ_E) and their SD were aggregated from the allocation of coefficients and C residence time in the individual plant and soil pools over the conterminous USA using the following formula (Barrett, 2002):

$$\begin{aligned}
\tau_E = {} & \alpha_L(\tau_L + \tau_F) + \eta\alpha_W\tau_F + \alpha_W(\tau_W + \tau_C) \\
& + \alpha_R[\varsigma_{R_1}(\tau_{R_1} + \tau_{S_1}) + \varsigma_{R_2}(\tau_{R_2} + \tau_{S_2}) + \varsigma_{R_3}(\tau_{R_3} + \tau_{S_3})] \\
& + F_1\tau_{S_1} + F_2\theta_{S_1}\tau_{S_2} + F_3\theta_{S_2}\tau_{S_3}
\end{aligned}$$

$$(3)$$

where

$$\begin{aligned}
F_1 &= \theta_F(\alpha_L + \eta\alpha_W) + \theta_C\alpha_W \\
F_2 &= \alpha_R\varsigma_{R_1} + F_1 \qquad\qquad (4) \\
F_3 &= \alpha_R\varsigma_{R_2} + \theta_{S_1}F_2
\end{aligned}$$

Ecosystem C uptake in the conterminous USA was estimated in forward modelling by using maximum likelihood
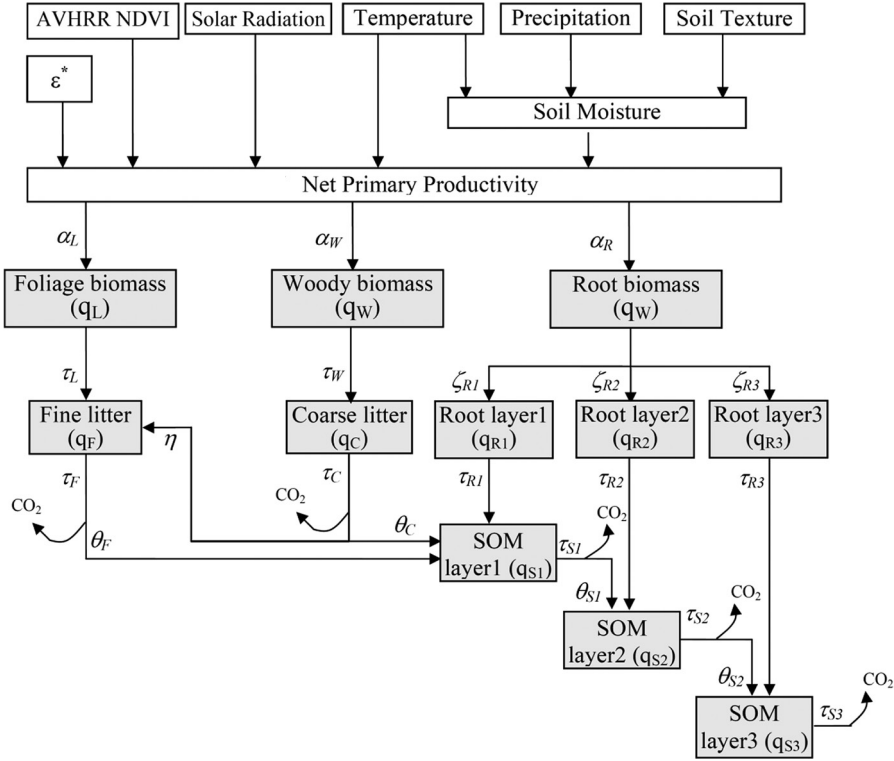
Fig. 1. Schematic diagram of the regional terrestrial ecosystem (TECO-R) model for inversion analysis of carbon residence time. NPP, which is modelled by maximum light-use efficiency (ε^*) (CASA model), is allocated to plant tissues (leaf q_L, stem q_W and root q_{R_j}) based on allocation coefficients (α_L, α_W, α_R). Plant tissues enter into fine litter (q_F), coarse litter (q_C) and soil organic carbon pools (q_S) through litterfall. Decomposed litter releases part of the carbon to the atmosphere, and the rest transfers into the soil (θ_C, θ_F). Through mechanical breakdown, part of the coarse litter becomes fine litter (η). To reflect the differences in soil profile, roots and soil organic carbon pools are divided into three layers (0–20, 20–50 and 50–100 cm). Each pool has an associated carbon residence time, τ_k ($k = 1, 2, 3$).

estimates (MLEs) of parameters and mean changes in NPP from 1982 to 1998 as a C input by Hicke et al. (2002). Propagation of parameter uncertainties to modelled ecosystem C uptake was assessed through simulating C uptake for all 40 000 samples of C residence times and allocation coefficients from MCMC simulation and then calculating SDs and CVs in the conterminous USA.

2.2. Data sources

In this study, 13 observed data sets were used for the parameter estimation, including three NPP data sets (NPP in leaves, stems and roots), five biomass data sets (one for leaves, one for stems and three for roots in three soil layers; 0–20, 20–50 and 50–100 cm), two litter mass data set (fine and coarse litter mass) and three SOC data sets in the three soil layers (0–20, 20–50 and 50–100 cm) for each biome if any. Compared to the study by Zhou and Luo (2008), we added coarse litter and more fine litter data points. There were a total of 7828 observed data points, which contained 175 data points for litter (vs. seven data

points in Zhou and Luo, 2008), 468 data points for NPP, 316 data points for biomass and 6869 data points for SOC. Spatial distribution and the detailed information of data points and their sources were documented in Appendix S2 in Supplementary material. Auxiliary data used in this study included: (1) the AVHRR-NDVI continental subsets of 8 km spatial resolution from 1982 to 1999 available from the Data and Information Services Center of Goddard Earth Science; (2) annual solar radiation produced by the NASA/Global Energy and Water Cycle Experiment with one-by-one degree spatial resolution; (3) monthly precipitation and temperature data sets with 4 km spatial resolution offered by the Spatial Climate Analysis Service; (4) soil texture data from State Soil Geographic Database (STATSGO) compiled by USDA Natural Resources Conservation Service; and (5) 1-km spatial resolution land cover data, containing eight vegetation types in the conterminous USA, derived from AVHRR using a decision tree classifier (Hansen et al., 2000). All those auxiliary data sets were resampled by ERDAS software with cubic convolution

Table 1. Symbol and definition of parameters, their lower (LL) and upper limits (UL) and other constraints for inverse analysis. Unit is gC MJ_{PAR}^{-1} for ε^* and years for carbon residence times. Allocation and partitioning coefficients are dimensionless.

Symbol	Definition	Unit	LL	UL	Other constraint
ε^*	Maximum light-use efficiency	g C MJ_{PAR}^{-1}	0.0	2.76	
α_L	Allocation of NPP to leaves	Dimensionless	0.0	1.0	$\alpha_L > \alpha_W$, not for grassland and cropland
α_W	Allocation of NPP to wood	Dimensionless	0.0	1.0	$\alpha_W = 0$ for grassland and cropland
α_R	Allocation of NPP to roots	Dimensionless	0.0	1.0	$\alpha_L + \alpha_W + \alpha_R = 1$
ξ_{R_1}	Allocation proportion of NPP for roots (0–20 cm)	Dimensionless	0.0	1.0	$\xi_{R_1} > \xi_{R_2} > \xi_{R_3}$
ξ_{R_2}	Allocation proportion of NPP for roots (20–50 cm)	Dimensionless	0.0	1.0	
ξ_{R_3}	Allocation proportion of NPP for roots (50–100 cm)	Dimensionless	0.0	1.0	$\xi_{R_1} + \xi_{R_2} + \xi_{R_3} = 1$
θ_F	Carbon partitioning coefficient of the fine litter pool	Dimensionless	0.0	0.5	
θ_C	Carbon partitioning coefficient of coarse litter pool	Dimensionless	0.0	0.1	$\theta_C = 0$ for grassland and cropland
θ_{S_1}	Carbon partitioning coefficient of SOC (0–20 cm)	Dimensionless	0.0	0.1	
θ_{S_2}	Carbon partitioning coefficient of SOC (20–50 cm)	Dimensionless	0.0	0.1	
η	Fraction of mechanical breakdown for coarse litter pool	Dimensionless	0.0	0.1	
τ_L	Site specific carbon residence time of leaves	Year	0.0	5.0	$0 \leq \tau_L \leq 1$ for deciduous broadleaf forest, grasslands and cropland
τ_W	Site specific carbon residence time of wood	Year	0.0	200.0	$\tau_W > \tau_L$, not for grassland and cropland
τ_{R_1}	Site specific carbon residence time of roots (0–20 cm)	Year	0.0	10.0	$\tau_{R_1} < \tau_{R_2} < \tau_{R_3}$, $\tau_{R_1} \leq 5$ for grassland and cropland
τ_{R_2}	Site specific carbon residence time of roots (20–50 cm)	Year	0.0	20.0	$\tau_{R_2} \leq 5$ for grassland and cropland
τ_{R_3}	Site specific carbon residence time of roots (50–100 cm)	Year	0.0	50.0	$\tau_{R_3} \leq 10$ for grassland and cropland
τ_F^*	Moisture and temperature corrected residence time of fine litter	Year	0.0	2.0	
τ_C^*	Moisture and temperature corrected residence time of coarse litter	Year	0.0	50.0	$\tau_C^* > \tau_F^*$, not for grassland and cropland
$\tau_{S_1}^*$	Moisture and temperature corrected residence time of SOC (0–20 cm)	Year	0.0	50.0	$\tau_{S_1}^* < \tau_{S_2}^* < \tau_{S_3}^*$
$\tau_{S_2}^*$	Moisture and temperature corrected residence time of SOC (20–50 cm)	Year	0.0	50.0	
$\tau_{S_3}^*$	Moisture and temperature corrected residence time of SOC (50–100 cm)	Year	0.0	100.0	

methods to a common projection (Lat–Long Projection) and spatial resolution (0.04°) (Smith and Brown, 1997).

2.3. Parameter estimation and uncertainties analysis

In the study, a Bayesian probabilistic inversion approach was employed in the TECO-R model to estimate C allocation coefficients and C residence times for each biome (a total of eight sets of parameters). A detailed description of the Bayes' theorem was given by Xu et al. (2006) and McCarthy (2007). Below is a brief overview.

In the context of this study, we applied Bayes' theorem to the C cycling. Specifically, the *posterior* PDF $p(c|Z)$ of C allocation coefficients and C residence times can be obtained from prior knowledge of parameters represented by a prior PDF $p(c)$ and information contained in 13 data sets represented by a likelihood function $p(Z|c)$. To specify the biotic influences, Bayes' theorem was applied for each of the eight biomes (ENF, DBF, MF, W, WG, S, G and C). Given one biome, we first specified the prior PDF $p(c)$ by giving a set of limiting intervals for all parameters with

uniform distribution (Table 1), and then constructed the likelihood function $p(Z|c)$ on the basis of the assumption that errors in the observed data followed Gaussian distributions. Parameter ranges were determined through a combination of literature values, knowledge of the system and prior information from Zhou and Luo's (2008) study for the optimal parameter values (Table 1). Initial C pools were set in 1987 and estimated by the method used in Luo et al. (2001) based on an initial steady-state C balance in the TECO model and experimental data at the start of this study. A partial likelihood function $p_i(Z|c)$ was specified according to distributions of observation errors ($e(t)$) for data set m:

$$p_i(Z|c) = \exp\left\{-\frac{1}{2\sigma_m^2} \sum_{n=1}^{N_m} [y_{nm} - \hat{y}_{nm}(x_n, \mathbf{a})]^2\right\} \quad (5)$$

where constant σ_m^2 is the error variance of the mth data set, y_{nm} is the nth observed data point in the mth data set, $\hat{y}_{nm}(x_n, \mathbf{a})$ is the modelled value that corresponds to the observation y_{nm}, N_m is the total data points in the mth

data set, x_n is an auxiliary forcing vector that includes NDVI, solar radiation, air temperature, precipitation and soil texture in a spatial grid where the nth observation was made; and $\mathbf{a}$ is a vector consisting of 22 parameters: $\mathbf{a} = \{\varepsilon^*,$ $\alpha_L,\ \alpha_W,\ \alpha_R,\ \xi_{R_1},\ \xi_{R_2},\ \xi_{R_3},\ \tau_L,\ \tau_W,\ \theta_F,\ \theta_C,\ \eta,\ \tau_{R_1},\ \tau_{R_2},\ \tau_{R_3},\ \tau_F^*,$ $\tau_C^*,\ \tau_{S_1}^*,\ \tau_{S_2}^*,\ \tau_{S_3}^*,\ \theta_{S_1},\ \theta_{S_2}\}$, each of the parameters is defined in Table 1. An integrated likelihood function $p(Z|c)$, which consists of $M\ (=13)$ partial likelihood functions $p_i(Z|c)$, is defined to measure the deviations between modelled and observed values for all the data points in the 13 data sets as follows:

$$p(Z|c) = \exp\left\{ -\sum_{m=1}^{M} \frac{1}{2\sigma_m^2} \sum_{n=1}^{N_m} [y_{nm} - \hat{y}_{nm}(x_n,\ \mathbf{a})]^2 \right\} \quad (6)$$

Therefore, with Bayes' theorem, the PPDF of parameters is given by

$$p(c|Z) \propto p(Z|c)p(c) \quad (7)$$

Since Bayes' theorem was applied to each of the eight biomes, eight sets of biome-specific values of parameter vector, $\mathbf{a}$, were obtained in the study.

To draw samples from $p(c|Z)$, a M–H algorithm was employed to construct a PPDF of C residence time and C allocation coefficients on the basis of prior information of model parameters, model structure and the data sets. The M–H algorithm is a MCMC technique, which could obtain high-dimensional PDFs of random variables via a sampling procedure (Metropolis et al., 1953; Hastings, 1970; Gelfand and Smith, 1990). For each biome, 40 000 samples were generated by the five runs from the M–H simulation with an acceptance rate of about 30–40% for the newly generated samples when the five runs converged as indicated by the G-R statistics (see Supplementary material). Marginal distribution, MLEs, means, SD and CV of C residence time and allocation coefficients were calculated from simulated 40 000 samples to assess the uncertainties of parameters (see Appendix S1 for a detailed description of the M–H algorithm, as well as an estimate of maximum likelihood estimators (MLEs), means and cross-correlations between parameters in Supplementary material).

2.4. Carbon uptake and uncertainties analysis

The MLEs of parameter values estimated from M–H simulation together with the corresponding C pool sizes in the inverse analysis were used in forward modelling to simulate ecosystem C uptake. In addition to C residence time, NPP is another key factor that determines the capacity of an ecosystem to absorb C (Luo et al., 2003). Hicke et al. (2002) estimated that the averaged NPP increase was 1.83 g C m^{-2} yr^{-1} (i.e., 0.5% per year) from 1982 to 1998 by

remote sensing in the conterminous USA but the trend in NPP changes varied with ecosystems. We applied the actual spatial pattern of NPP changes from Hicke et al. (2002) to quantify C uptake for each spatial grid at a yearly time step. Therefore, the spatial difference of C uptake potential was caused by both the spatial pattern of C residence times and NPP changes. In our study, to focus on these two factors (i.e., NPP changes and C residence times) in influencing C uptake, we assumed that impacts of disturbance and environmental factors (e.g., temperature) on C uptake were reflected by remote sensing data in NDVI.

Uncertainties of terrestrial C uptake were assessed by the SDs and coefficients of variation (CVs). We used 40 000 samples of C residence times and allocation coefficients estimated from M–H simulation as expressed by PPDFs $p(c|Z)$ for each biome to simulate ecosystem C uptake and assess propagation of the parameter uncertainties in forward modelling as described above. The forward model simulation for ecosystem C uptake was made over 50 yr from 1987 to 2036. The SDs and CVs were calculated from the 40 000 data sets of ecosystem C uptake in the conterminous USA.

3. Results

3.1. Comparison between modelled and observed data sets

To evaluate the validity of the inverse analysis method, it is important to compare the values between modelled and observed data sets. This study used the 13 observed multi-site data sets and multiple sources of information to constrain the parameters for each biome (totally eight biomes), which may reduce uncertainty. The TECO-R model with the Bayesian inversion reasonably simulated C fluxes and pools in the conterminous USA, which were closely related to the corresponding observed data sets (Fig. 2). There were a few cases where modelled values deviated from observations. For example, the simulated NPP in roots and stems, and root biomass in 0–20 cm soil layer were largely underestimated compared to the observed data (Fig. 2).

3.2. Estimated C residence times and allocation coefficients

The histograms of 40 000 samples generated by the five runs from the M–H simulation are shown in Fig. 3 for ENF and in Fig. 4 for G, representing forest and non-forest biomes, respectively, when the five runs converged as indicated by the G-R statistics ($=1$, data not shown). For all eight biomes, most of the parameters are nearly Gaussian distributed but with considerable differences in
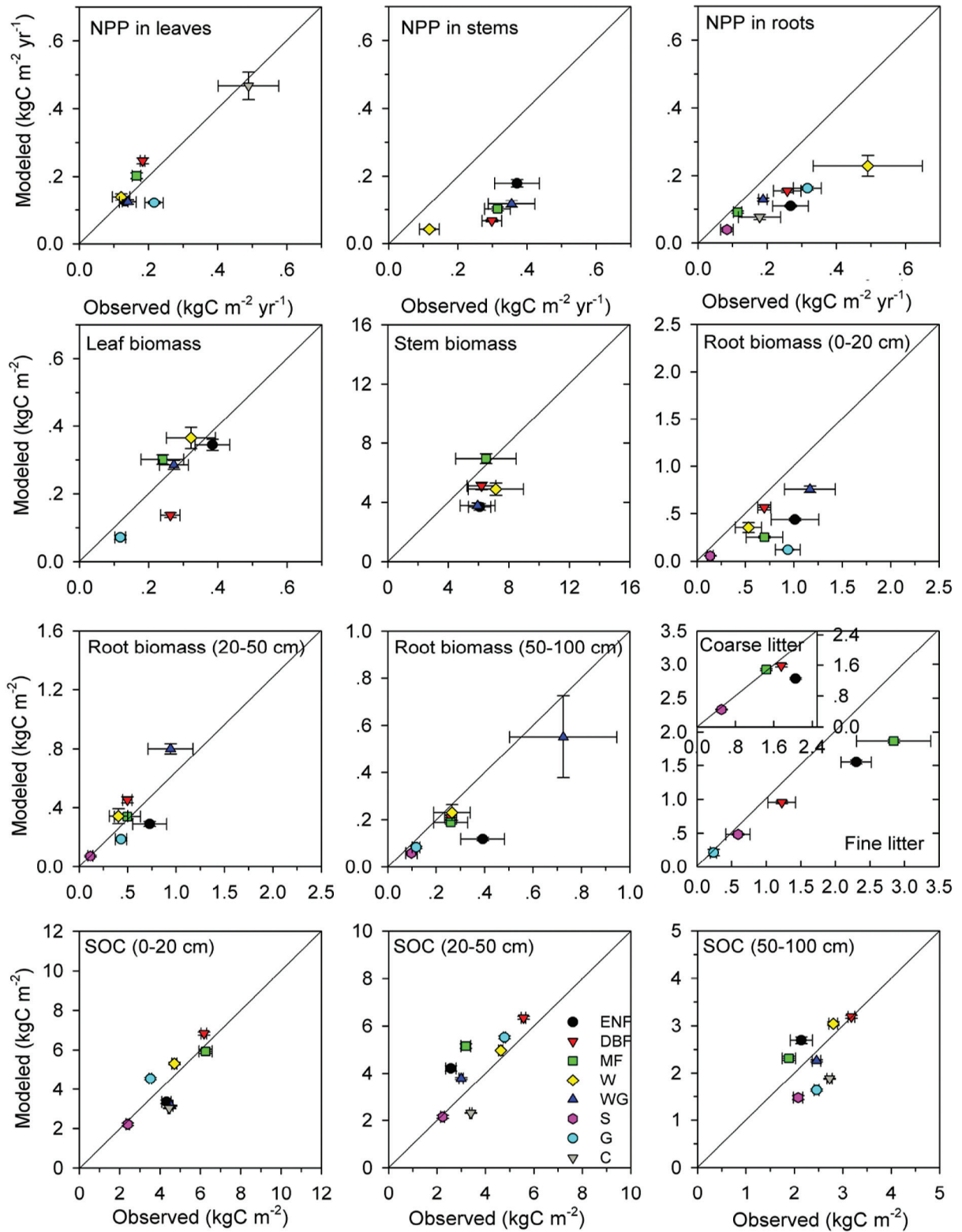
Fig. 2. Comparisons between modelled and observed data lumped by each biome for 13 data sets. Biomes: ENF, evergreen needleleaf forest; DBF, deciduous broadleaf forest; MF, mixed forest; W, woodland; WG, wooded grassland; S, shrubland; G, grassland; and C, cropland. Unit with dimensions is not shown.
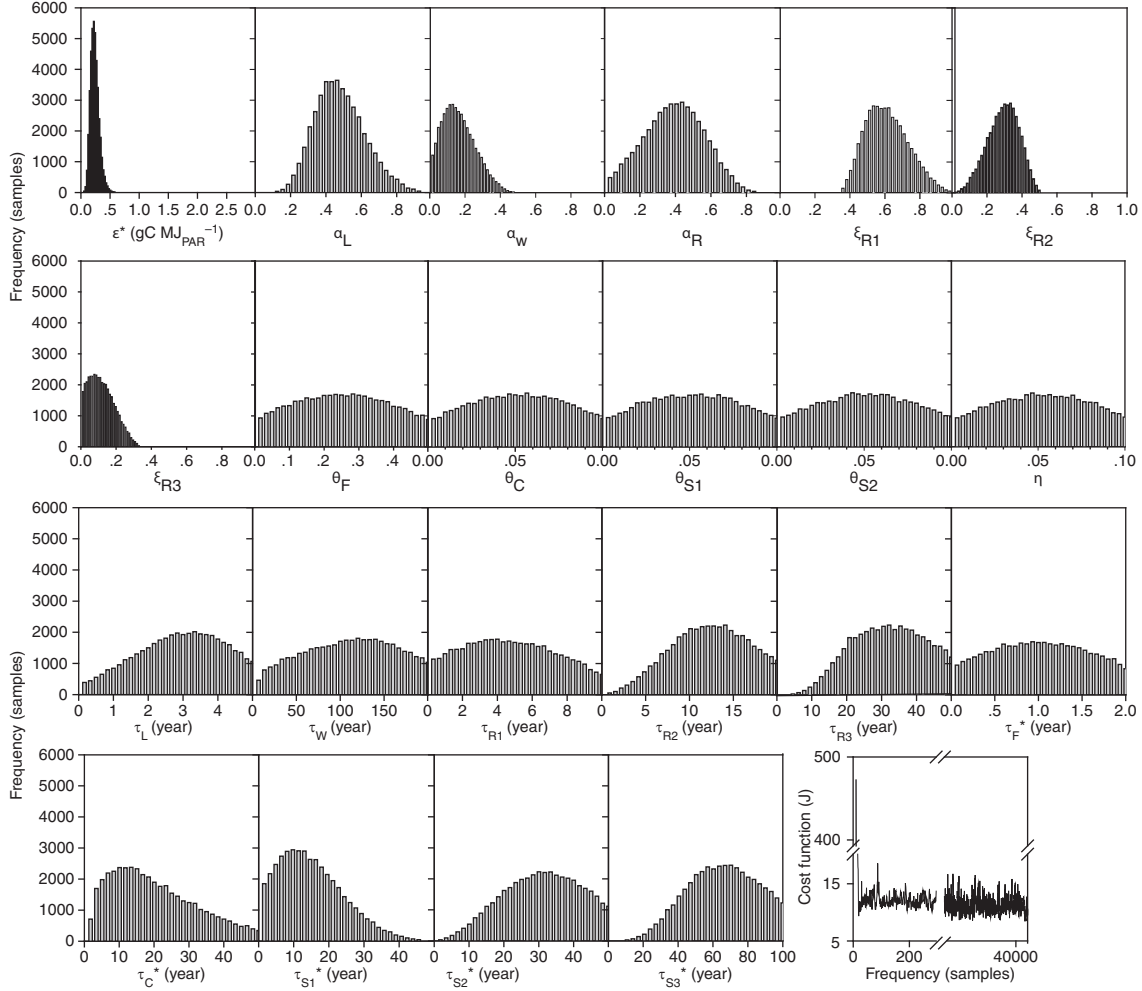
Fig. 3.   Inversion results showing the histograms of 22 estimated parameters and cost function with 40 000 samples from M–H simulation for evergreen needleleaf forest (ENF). See Table 1 for parameter abbreviation. Unit with dimensions is not shown.

their variability. Parameters ε^*, α_L, α_W, α_R, ξ_{R_1}, ξ_{R_2} and ξ_{R_3} were well constrained with steep peaks of frequencies (Figs. 3 and 4). In contrast, parameters θ_F, θ_C, θ_{S_1}, θ_{S_2}, η and τ_F^* were constrained to some degree but with considerably spread *posterior* PDFs. Other parameters contain intermediate information on parameter constraint compared to those well and loosely constrained (Figs. 3 and 4). Several parameters (e.g., τ_W, τ_C^*) were well constrained by some biomes but not by others, probably due to the limitation of available data.

In performing a simultaneous optimisation on 22 parameters, which provides an ensemble of parameter sets, it is likely to have parameter correlations due to tradeoffs between the parameters pairs. Our cross-correlation analysis showed that more than 85% of correlations between parameters, |corr|, were below 0.2 and >90% were below 0.3 (Appendix S3 in Supplementary material). High |corr| values (>0.4) mainly resulted from equality constraints

(e.g., $\alpha_L + \alpha_W + \alpha_R = 1$). In addition, allocation of NPP to wood (α_W) was directly related to C residence times of wood (τ_W) and coarse litter (τ_C), resulting in high |corr| values with a range of 0.4–0.7 depending on the biomes.

The MLEs are the parameter values corresponding to the peaks of their marginal distributions (Figs. 3 and 4). The estimated maximum potential light-use efficiencies (ε^*) varied largely with vegetation types (Fig. 5a), being high in cropland (0.71 ± 0.20 g C MJ_{PAR}^{-1}) and being low in ENF (0.23 ± 0.08 g C MJ_{PAR}^{-1}). The S has the largest variability (CV = 119%) with a ε^* value of 0.47 ± 0.56 g C MJ_{PAR}^{-1}. The ε^* values for DBF, G, MF, W and WG were similar, ranging from 0.30 to 0.41 g C MJ_{PAR}^{-1} (Fig. 5a). Our estimated maximum potential light-use efficiency of cropland was within the range from 0.41 to 0.94 g C MJ_{PAR}^{-1} given by Lobell et al. (2002).

Allocation coefficients of NPP to leaves, stems and roots also varied with biomes (Fig. 5b). C allocation to leaves
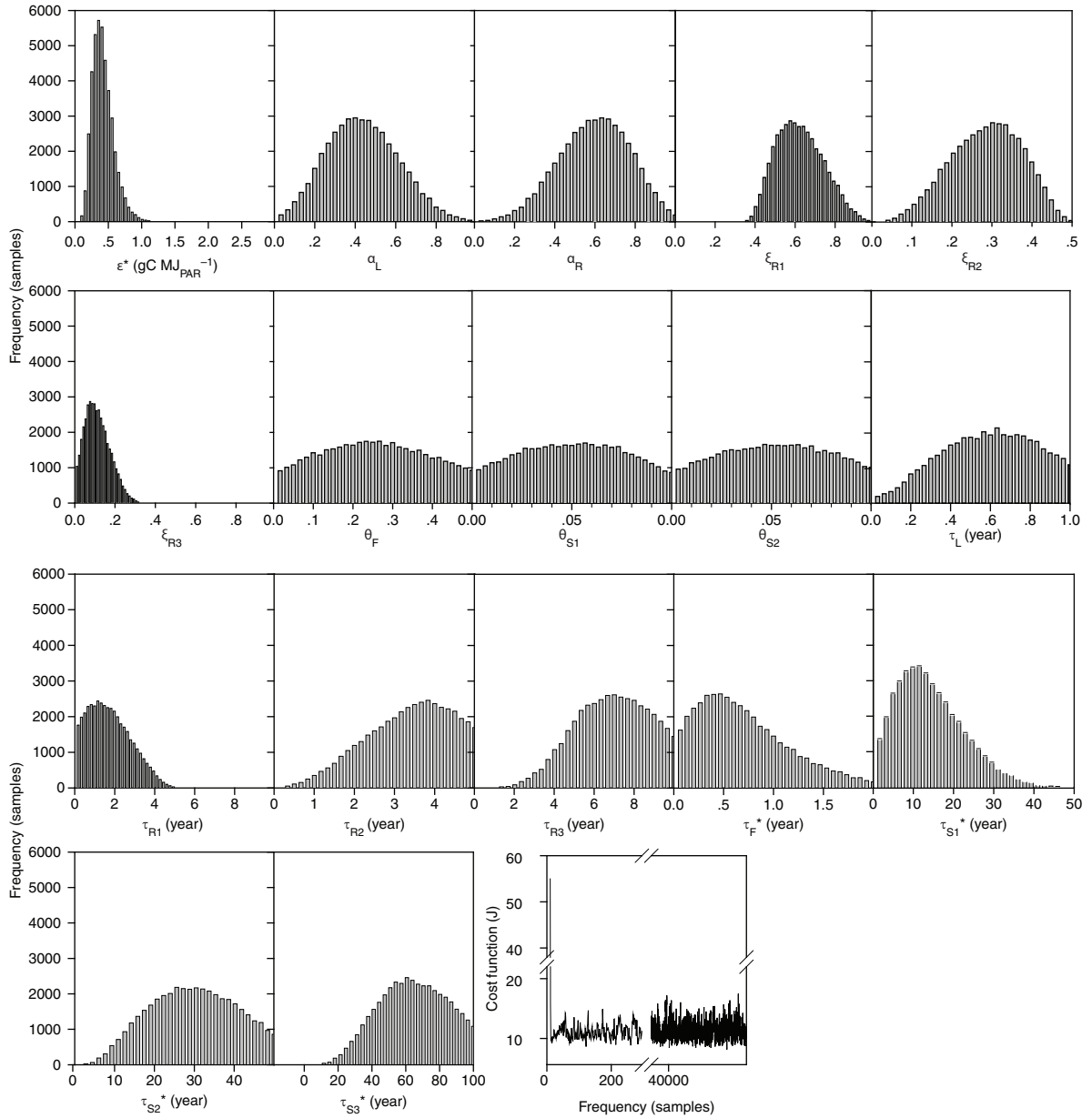
Fig. 4. Inversion results showing the histograms of 17 estimated parameters and cost function with 40 000 samples from M–H simulation for Grassland. See Table 1 for parameter abbreviation. Unit with dimensions is not shown.

was relatively higher for the cropland ($79 \pm 11\%$) and S ($60 \pm 14\%$) than the other biomes, which had similar C allocation coefficients from $42 \pm 18\%$ (G) to $52 \pm 10\%$ (MF). The C allocation of NPP to stems ranged from $12 \pm 8\%$ (WG) to $23 \pm 10\%$ (MF). Accordingly, there were relatively lower C allocation to roots for MF, S, and cropland (20 ± 13 to $24 \pm 10\%$) and higher C allocation for G ($58 \pm 18\%$) and WG ($44 \pm 7\%$). For the two non-forest biomes (G and cropland), which allocated C only to leaves and roots, G allocated much more C into roots ($58 \pm 18\%$) than cropland did ($21 \pm 11\%$). The results were similar to

the findings from Bradford et al. (2005), showing cropland allocated 28% of NPP to roots, much lower than G. Vegetation types did not significantly affect the allocation coefficients of roots to three soil layers (0–20, 20–50 and 50–100 cm) (Fig. 5c).

In DBF, G and cropland (0.50 ± 0.27 to 0.61 ± 0.02 yr), the C residence times of leave (τ_L) were much lower than those in other biomes (1.58 ± 0.97 to 2.82 ± 1.22 yr, Fig. 5e). However, the C residence times of wood (τ_W) did not vary largely with biomes (78.8 ± 48.8 to 109.3 ± 52.4 yr) but they appeared in high variability (CV $= 46$–62%, Fig. 5g).
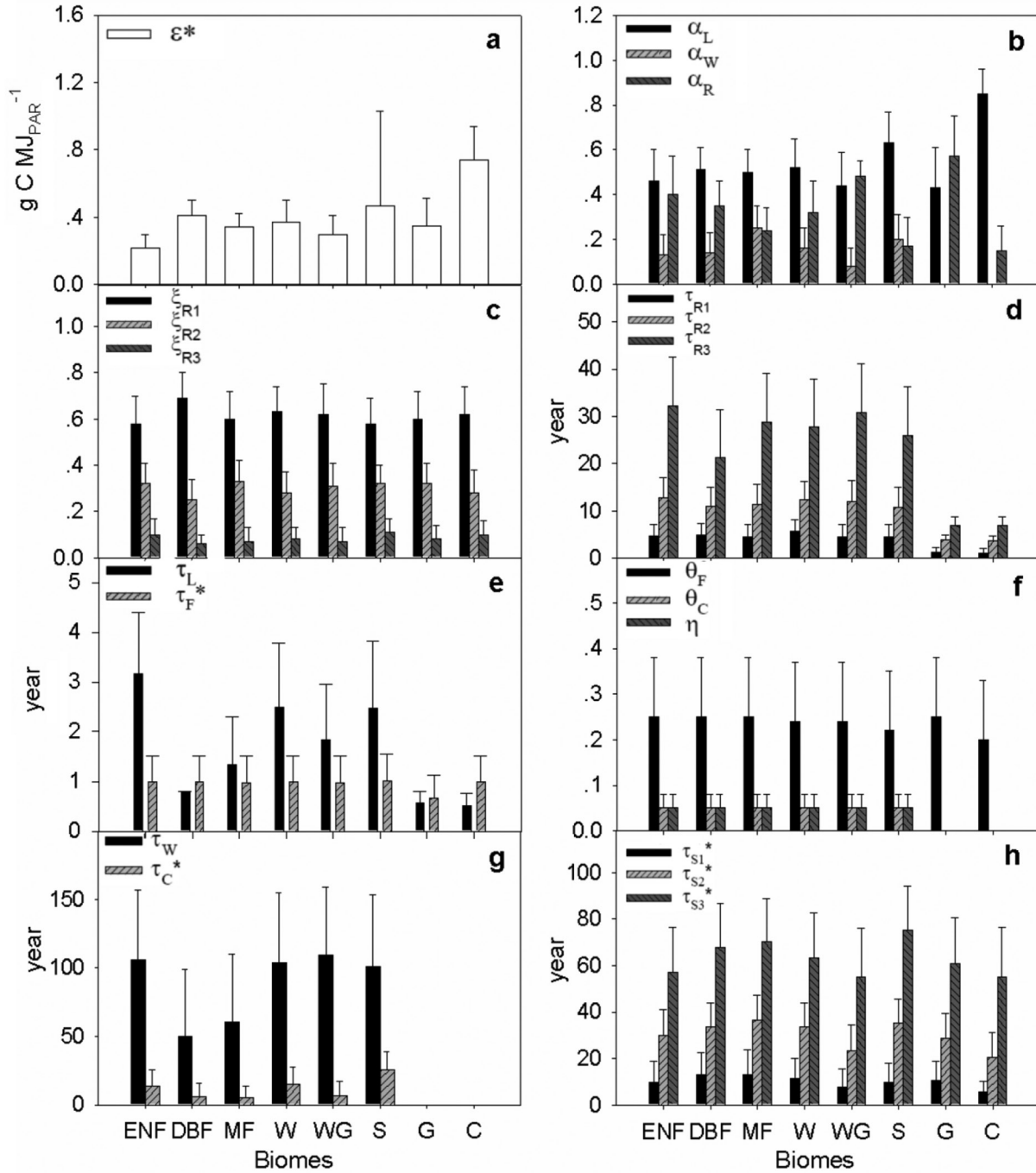
Fig. 5. MLEs or means of estimated parameters for eight biomes. Error bars represent standard deviations (SDs) of parameters calculated from 40 000 samples of M–H simulation. Parameters θ_{S1} and θ_{S2} were not shown due to their consistence between biomes (0.05 ± 0.03). See Fig. 2 for biome abbreviations and Table 1 for parameter abbreviations.

The root residence times varied from 4.46 ± 2.60 to 5.37 ± 2.38 yr in top soil layer, 11.44 ± 1.23 to 12.38 ± 3.92 yr in middle soil layer and 27.12 ± 10.1 to 31.1 ± 10.3 yr in deep soil layer for the forests, W and S (Fig. 5d). The G and

cropland had much lower root residence times for all three layers than other biomes (Fig. 5d). The temperature- and moisture-corrected C residence times of litter and SOC showed considerable uncertainties with large SDs

(Fig. 5e, g, h). The C residence times of fine litter (τ_F*) were only well constrained for G, showing the lowest value (Fig. 5e). The C residence times of coarse litter (τ_C*) were relatively lower for DBF, MF and WG with a range from 11.1 ± 8.8 to 13.9 ± 10.4 yr than those for ENF, W and S (19.3 ± 11.9 to 25.7 ± 13.1 yr). The values of SOC residence times ($\tau_{S_1}^*$, $\tau_{S_2}^*$, $\tau_{S_3}^*$) were much higher than those of litter, ranging from 7.48 ± 4.86 to 17.3 ± 10.3 yr for the top soil layer, from 23.5 ± 10.7 to 32.8 ± 10.4 yr for the middle soil layer, and from 56.7 ± 21.4 to 67.3 ± 18.8 yr for the bottom layer, but they did not vary largely with biomes. C partitioning coefficients (θ_F, θ_C, θ_{S_1}, θ_{S_2}, and η) also did not vary with biomes but with large variability (CV $> 50\%$).

3.3. Spatial patterns of ecosystem C residence times

The C residence time of the whole ecosystem varied highly with locations (or ecosystems) in the conterminous USA, ranging from about 10 yr in some cropland grids to 180 yr in some S grids (Fig. 6a). However, most values fell
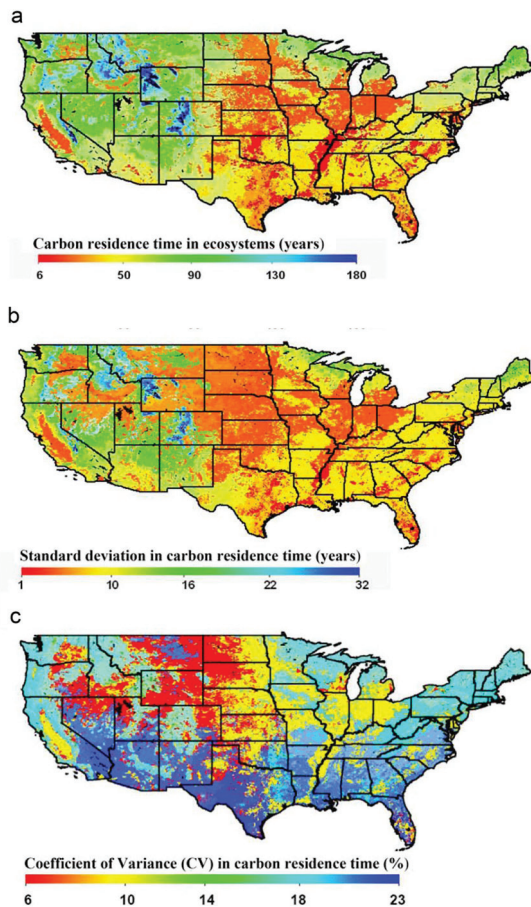
between 15 and 90 yr. The averaged C residence time of the whole conterminous USA was 56.8 yr. The west regions had higher C residence time, while the central and east regions showed lower values. The SDs of the ecosystem C residence time for each grid were shown in Fig. 6b, ranging from 1 to 32 yr, with most values between 2 and 15 yr. The averaged SD of ecosystem C residence time was 8.8 yr for the conterminous USA. As shown by the coefficients of variation (CV = SD/mean), large uncertainties appeared in southern USA (CV = 14–22%) and small uncertainties in the central (CV = 5–10%, Fig. 6c).

Among all eight biomes, the cropland had the lowest C residence time (16.6 ± 1.8, Fig. 7b) due to the lack of long residence wood tissues and coarse litter. WG also showed the relative low C residence time because of the low allocation of wood tissue from NPP (40.5 ± 8.5, Fig. 7a). The ENF showed the highest C residence times of 85.9 years, probably due to the low temperature in this area. The other biomes had C residence times between 55.9 ± 9.7 and 72.1 ± 4.1 yr. Along the latitude, C residence time displayed a strong positive correlation (Appendix S4 in Supplementary material, $R^2 = 0.91$), suggesting that

Fig. 6.   Spatial patterns of ecosystem carbon residence times (a), its standard deviation (SD, b) and coefficient of variation (CV = SD/mean, c) in terrestrial ecosystems of the conterminous USA.
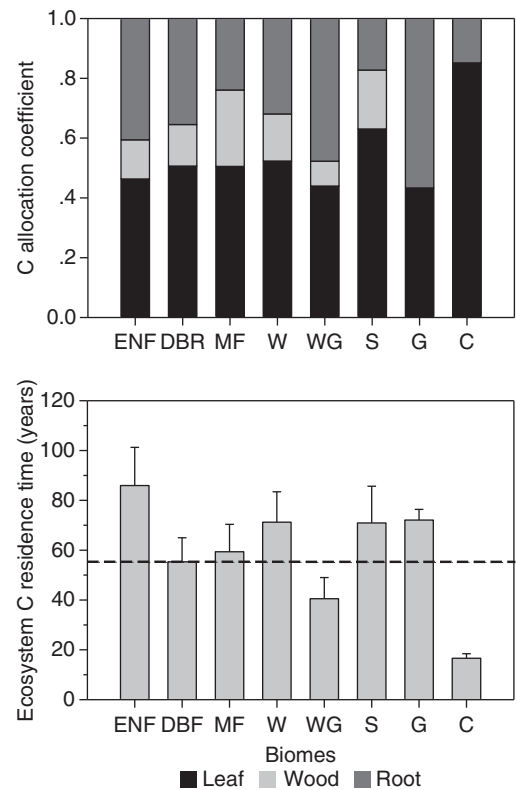

Fig. 7.   Carbon allocation coefficients (a) and the averaged ecosystem C residence time (b) for eight biomes. See Fig. 2 for biomes abbreviation. Horizontal dash line represents the average ecosystem C residence time in the conterminous USA. Error bars represent standard deviation.

temperature could be one of the most important factors in influencing C residence times.

3.4. Carbon uptake in soil and ecosystems

When the actual, spatially heterogeneous NPP changes from Hicke et al. (2002) was applied to each spatial grid in the forwarding model, the estimated ecosystem C uptake showed large variability with a range from -60 to 140 g C m^{-2} yr^{-1} within the 50 yr. The southeast regions had higher C uptake, possibly due to the relatively higher values of both NPP increases and C residence times (Fig. 8a and Appendix S5). The Central Plains also had high C uptake, which may mainly result from the highest NPP increase in the conterminous USA. But ecosystem C release occurred in some southwest regions because of decreased NPP (Fig. 8a and Appendix S5). The estimated total C uptake in the conterminous USA was 0.20 ± 0.06 Pg C yr^{-1}.

Ecosystem C sequestration (g C m^{-2} yr^{-1})

-60 0   40   80   140

Coefficient of variance of ecosystem C sequestration (%)

0   20   40   60   80

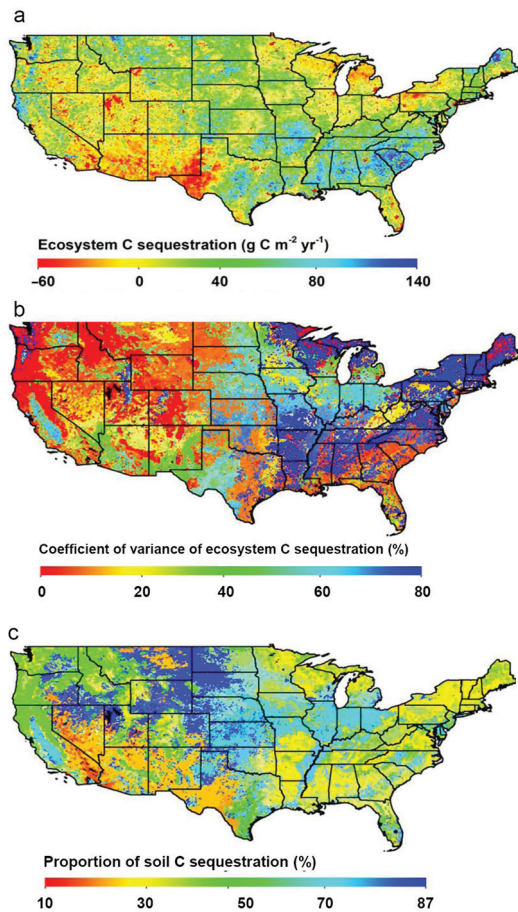**Proportion of soil C sequestration (%)**

10   30   50   70   87

Fig. 8. The potential of ecosystems C uptakes (a), their coefficient of variance (b) and proportion of soil to ecosystem C uptake in the conterminous USA under the actual NPP increases from Hicke et al. (2002). The positive value means C sink while the negative value mean C source.

For each biome, the large C uptake appeared in ENF, MF, cropland and G, while S did not contribute C uptake in the conterminous USA (Fig. 9). Among the ecosystem C uptake, soil contributed about 50% of C uptake. G and cropland had the relatively higher proportions of soil C uptake (51–82%) than forests and Ws (25–49%) (Figs. 8c and 9).

The coefficients of variance (CVs) of ecosystem C uptake were higher in central and eastern USA with a range from 50 to 80% than other regions from 0 to 40% (Fig. 8b). Both high C residence times and large increases in NPP may mainly lead to high variability in ecosystem C uptake in these regions. The ENF, MF and cropland displayed large uncertainties with high SDs (Fig. 9).

4. Discussion

The results from atmospheric inversion methods have substantially improved our understanding on the dynamics of the global and regional C cycle over several decades (Kaminski et al., 2002; Rayner et al., 2005; Prihodko et al., 2008). Meanwhile, the atmospheric inversion can be greatly complemented by the application of multiple-constraint data assimilation with ecological observations in predictive understanding of ecosystem C uptake (Barrett, 2002; Wang and Barrett, 2003). In this study, the multiple-constraint approach with Bayesian inversion of a regional terrestrial ecosystem model (TECO-R) worked well to optimise C residence times and assess their uncertainties with SD and CV. The method also provided information on parameter correlations. Such information on uncertainties of C residence time is crucial for understanding the degree of confidence we can have on model predictions of the global terrestrial C balance.
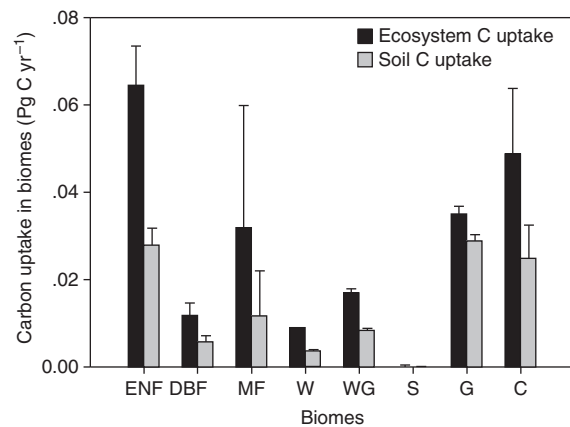
Fig. 9. Yearly ecosystem and soil C uptake in different biomes under the actual NPP increases from Hicke et al. (2002). Error bars represent standard deviation (SD).

4.1. Method comparisons on estimated carbon residence times

Comparisons between modelled and observed data are essential to evaluate the validity of the inverse analysis method. When most of the modelled values were closely related to the corresponding observed data, the TECO-R model simulated the lower values for the NPP in roots and stems, and root biomass in 0–20 cm soil layer compared to the observed ones (Fig. 2). The large underestimation is probably due to both model assumptions and observation errors. The model assumed that stem and root allocation coefficients depended on leaf allocation coefficients ($\alpha_L + \alpha_W + \alpha_R = 1$), and the residence times of roots were on a scale of years. We allowed the model to search for root residence times within ranges of 0–10, 0–20 and 0–50 yr in three soil layers (0–20, 20–50 and 50–100 cm), respectively. Generally, the root residence times of deep soil layers (20–50 and 50–100) may be larger than those of soil surface. The wide lower ranges (close to 0) for the two deep soil layers may underestimate root residence times. On the other hand, the observed root NPP and biomass were estimated from a variety of methods (e.g., ingrowth cores, minirhizotron and allometry; Jackson et al., 1996) and stem NPP was usually indirectly estimated from diameters at the breast height (DBH, DeLucia et al., 1999), possibly resulting in large errors in different locations. In addition, the tradeoff of estimated parameter values under the integrated cost function for all data sets minimised the total deviation possibly at expense of stem and root NPP.

Using Bayesian inversion technique in this study, most of the estimated optimal C residence times were similar to those from Zhou and Luo (2008) using generic algorithm (GA), although there were the less observed data sets in Zhou and Luo (2008) (i.e., no coarse litter and only seven fine litter data points). C residence times of leaves for forests and Ws ranged from 0.78 to 3.17 yr, which were comparable to those estimated by observed stocks and fluxes with 1.64 yr for Ws and 2.32 yr for forests (Cannell, 1982; DeAngelis et al., 1997). The C residence times of wood ranged from 78.8 to 109.3 yr similar to those estimated from C balance and radiocarbon dating methods (Cannell, 1982; Turner, 1984; DeAngelis et al., 1997). The estimated temperature- and moisture-corrected residence times of fine litter were from 0.67 to 1.0 yr (Fig. 5), which were very close to the means from meta-analysis by Silver and Miya (2001) and Zhang et al. (2008). The estimated residence times of coarse litter were from 5.0 to 25.7 yr (Fig. 5), well within the range (2.3–33 yr) summarised from coarse-wood decomposition experiments for deciduous and coniferous species in the eastern USA (Todd and Hanson, 2003). The estimated C residence times of roots were similar to those from the other two inverse analysis by

Genetic Algorithm (GA) in Australia and USA (Barrett, 2002; Zhou and Luo, 2008), but were much higher than the experimental estimations (Gill and Jackson, 2000). The discrepancy between modelling and experiments may result from the differences of root representatives for soil layers, time scales and the methods (Jackson et al., 1996). The estimated residence times of SOC increased with soil depth, having a range of 5.53–13.32 yr for the top soil layer, 20.35–36.73 yr for the middle soil and 55.01–75.34 yr for the bottom soil (Fig. 5), which were comparable from only a few studies at different soil layers (Jenkinson et al., 1999; Fontaine et al., 2007). However, most of the experimental studies showed larger ranges from several years to thousands of years when only the surface soil or one combined soil layer was used (Jenkinson et al., 1999; Garten and Hanson, 2006).

To facilitate parameter estimation in this study, we divided SOC and root biomass into three soil layers (0–20, 20–50 and 50–100 cm) as by Barrett (2002) and Zhou and Luo (2008), instead of compartmentalised SOC (i.e., fast, slow and passive) according to decomposition rates as well as fine root and coarse root compartments in most of the models (e.g., CENTURY, Parton et al., 1987; TECO, Luo et al., 2003). In this way, a one-to-one relationship was ensured between modelled and observed root biomass and SOC, and the models do not need extra mapping functions as by Luo et al. (2003). In the process-based models, the conceptual compartmentalised SOC generally cannot be measured directly. The experimentally extractable soil C pools (e.g., labile and the remaining recalcitrant pools) cannot simply represent the conceptual C pools in the mechanistic models. This highlights the need for a comprehensive representation of belowground C-pools in the process models (including deep soil C) in order to achieve unbiased parameter estimates (Barrett, 2002).

The relatively fast turnover of ecosystem C appeared in the east and central regions, largely resulting from the cropland and G due to the lack of long-residence wood tissues and coarse litter (Figs. 6a and 7b). In the west regions, forests and Ws are in the mountainous regions with high residence times (Fig. 7b). In addition, our results that C residence times correlated positively with the latitude (Appendix S4, $R^2 = 0.91$) suggested that temperature may be a key factor in influencing C residence time. The estimated aggregate ecosystem C residence time in the conterminous USA was 56.8 yr, which represented an upper limit of the global average C residence time (26–60 yr) as estimated by various experimental and modelling approaches (Raich and Schlesinger, 1992; Randerson et al., 1999; Thompson and Randerson, 1999), but was lower than the estimation by Barrett (2002) in Australia (78 yr). The relatively high ecosystem C residence time in the conterminous USA compared to the global average one

mainly resulted from the inclusion of deep C in the soil profile (0–100 cm).

4.2. Uncertainties in C residence times

Unlike the methods of uncertainties analysis by varying one parameter at a time (Knorr, 2000; El Maayar et al., 2002), the probabilistic inversion can estimate the probability of uncertain parameters and state variables (McCarthy, 2007). The Bayesian approach constructs parameter distributions and assesses parameter uncertainties by quantifying MLEs, means, SDs, and CVs, and offers much richer information contained in data, model structure and prior knowledge on parameters than the deterministic inversion (Raupach et al., 2005; Xu et al., 2006). Our results showed that *posterior* PDFs of most of the parameters are nearly Gaussian distributed but with considerably different variability, which were largely improved in comparison with the prior uniform distribution (Figs. 3 and 4). The MLEs of relatively well constrained parameters ε^*, α_L, α_W, α_R, ξ_{R_1}, ξ_{R_2} and ξ_{R_3} were equivalent to optimal estimates from deterministic inversion using the least squares method (Tarantola, 1987; Raupach et al., 2005), which are generally in good agreement with those by GA in Zhou and Luo (2008). A very large set of various observations, especially NPP and biomass in leaves, stems and roots, is advantageous to constrain these allocation coefficients of C from NPP.

The parameters θ_F, θ_C, θ_{S_1}, θ_{S_2}, η and τ_F^* were constrained to some degree but with considerably spread *posterior* PDFs. Their large uncertainties were probably due to either the lack of experimental data or the mismatch of timescales between the available data and the parameters to be estimated. For example, the lack of microbial biomass, respiration and other data probably resulted in low constraints of parameters related to litter and SOC dynamics (Xu et al., 2006). Because of the mismatch between timescales for observation of vegetation change and parameters, little is known about the dynamics of terrestrial ecosystems over long temporal scales (Wiegand and Milton, 1996). In addition, parameter correlations (e.g., α_W–τ_W or τ_C) may also result in large uncertainties in C residence times (Appendix S3). Xu et al. (2006) found that reduced measurement errors led to substantially reduced uncertainties of estimated parameters, suggesting that error magnitudes in observations may be one of the factors in determining parameter uncertainties. Prior information about parameter space has been suggested to be another source of parameter uncertainties (Spear et al., 1994). When we expanded *prior* ranges of these parameters (i.e., increased upper limit), their histograms, however, did not show statistically significant changes, similar to the finding from Xu et al. (2006).

Parameters τ_{R_2}, τ_{R_3} $\tau_{S_1}^*$, $\tau_{S_2}^*$ and $\tau_{S_3}^*$ were relatively well constrained by 13 data sets (Figs. 3 and 4), probably due to a large number of data sets on roots and soil C (Appendix 2–3). The parameters representing long C residence times from deep old SOC pools (i.e., $\tau_{S_2}^*$ and $\tau_{S_3}^*$) can be further constrained by long-term observations or data sets that characterise long-term processes, such as those from C isotope studies (Randerson et al., 2002; Scholze et al., 2008).

Ecosystem C residence times in the conterminous USA showed large spatial variability and uncertainties (Figs. 6b, c). Uncertainties of ecosystem C residence times as indicated by CV were high in southern USA with a range of 14–22% and low in the central area (6–12%, Fig. 6c). The spatial distribution of uncertainties in the conterminous USA may be attributable to either the lack of experimental data or the mismatch of timescales between the available data and the estimated parameters as well as environmental conditions. CVs of ecosystem C residence times were similar with a range from 17.2 to 21.0% except G (6.0%) and cropland (10.9%, Fig. 7b). Low uncertainties of C residence times in the G and cropland greatly resulted from availability of large volumes of data and relatively evenly distributed environmental conditions over space, particularly in croplands, compared to those in forests and W.

4.3. Estimated ecosystem C uptake and its uncertainties

Driven by NPP changes from 1982 to 1998 as estimated by Hicke et al. (2002), total C uptake estimated by the Bayesian approach was 0.20 ± 0.06 Pg C yr^{-1} in the conterminous USA. The estimate is comparable to the one by Houghton et al. (1999) with a value of 0.15–0.35 Pg C yr^{-1} based on the analysis of historical data on land use and land cover changes as well as inventory estimate (0.30 Pg C yr^{-1}, Birdsey and Heath, 1995). However, this estimate is larger than an ensemble of three models (Biome-BGC, Century and TEM) with a value of 0.08 Pg C yr^{-1} (Schimel et al., 2000) but smaller than C sink from atmospheric inversion with a range from 0.30 to 0.58 Pg C yr^{-1} (Pacala et al., 2001).

Ecosystem C uptake is largely determined by both ecosystem C residence time and C input (Luo et al., 2003). Our modelling results show that both forest and non-forest ecosystems are potentially important to sequester C in the conterminous USA (Fig. 8a and 9). The spatial pattern of ecosystem C uptake was closely related to the greenness map from U.S. Geological Survey (USGS) in the summer (http://ivm.cr.usgs.gov/). The most important forests for C uptake were ENF and MF, which were located in the southeast regions. This may be attributed to the young

stand age of pine plantations (Sheffield and Dickson, 1998) and the immature hardwood forests (Brown et al., 1997). Currently, many of the pine stands in southern states have large potentials to increase productivity with intensive management (Allen et al., 1990). Interestingly, our results suggested that the substantial amounts of carbon (42% of the total) were sequestered in the G and cropland, which was comparable to the study by Pacala et al. (2001) with atmospheric inverse analysis. The substantial C uptake in the G and cropland largely resulted from their large areas, the Conservation Reserve Program (i.e., conversion of unproductive croplands to perennial G), expanded uses of no-till agriculture and improved productivity due to new plant varieties and increased fertiliser inputs (Houghton et al., 1999; Pacala et al., 2001; Lobell et al., 2002). In addition, disturbance and land management are important components of terrestrial C balance. In this study, we simply assumed that impacts of disturbance, land management and environmental factors on C residence time and uptake were reflected by remote sensing data in NDVI.

It is evident that large uncertainties remain in our assessment of terrestrial C uptake on the regional and global scales (Schimel et al., 2000; Pacala et al., 2001). Our results show large uncertainties with high CV in continental C storage, especially in central and eastern USA (Fig. 8b). The large uncertainties in C uptake could result from high spatial variability in initial NPP values, C residence times, or both (Fig. 6a and Appendix S5). To separate sources of uncertainties, we applied a uniformed NPP increase of 1.83 g C m^{-2} yr^{-1} (i.e., the averaged NPP increase in the conterminous USA from Hicke et al. (2002)) to quantify uncertainties of ecosystem C uptake. Our results suggest that the main source of uncertainties in C uptake was from spatial variability in C residence times (data not shown). In addition, the initial condition of NPP was another source of uncertainties. In the central and eastern USA, high variability in initial NPP and median variability in ecosystem residence time were combined to contribute to the large uncertainties (Fig. 7a). Therefore, uncertainties of ecosystem C uptake were strongly influenced by variability in C residence times and initial NPP values. Uncertainties in C uptake also varied with biomes (Fig. 9). The ENF, MF and cropland thus displayed larger uncertainties than the other biomes (Fig. 9). Besides those reasons discussed above, one additional cause of the high uncertainties in forest C uptake is scarce observations in mountainous regions where forests grow.

In this study, the 40 000 samples of posterior parameters were used to assess the regional uncertainty in terrestrial C uptake for the conterminous US. However, it is impractical with such a large size of ensemble simulations to assess the uncertainty in the global models (e.g., Community Land Model - CLM), especially with 3-D earth system models.

To reduce the computational cost for assessing the regional uncertainty, we randomly selected 5000 samples to test whether a subset of posterior parameters could represent all parameters or not. We found that the results from these two sets of posterior parameters were similar (data not shown). We can thus use a subset of posterior parameters to reduce the computational cost in assessing the uncertainty for global models.

This study focuses on the uncertainties in process-based terrestrial ecosystem models due to parameter uncertainties, and aims at providing quantitative information about the confidence in model results. Future steps toward a quantitative assessment of uncertainties in modelling terrestrial ecosystem dynamics will need to account for uncertainties related to model assumptions and to consider driving factors of climate change such as warming, elevated CO_2, altered precipitation, and land use change and N deposition. In addition, investigation and mapping of disturbances are important to quantify C uptake in the conterminous USA, particularly in the human-impacted ecosystems. Long-term remote-sensing data sets may be important to provide estimates of disturbance that could also be assimilated into a model. An important improvement in our understanding of uncertainties might be obtained by the combination of long-term experiments with well-documented time series data, such as net ecosystem exchange by eddy covariance, NPP, plant biomass, SOC and by the integration of the results with ecosystem modelling in a multiple-constraint framework.

5. Conclusions

Terrestrial ecosystems play a significant role for CO_2 uptake in global C cycling budget. The capacity of ecosystem C storage is largely determined by C residence time in plants and soil. However, C residence time is a large source of uncertainties in the present understanding of the global C cycle. To reduce its uncertainties and improve prediction from C-cycle models, quantifying C residence times and assessing its uncertainties becomes paramount. In this study, we applied Bayesian approach with a M–H algorithm to a regional model (TECO-R) to quantify optimal C residence times and assess their uncertainties with marginal distribution, SDs and CVs in the conterminous USA based on 13 data sets. Our results showed that most of the parameters are nearly Gaussian distributed but with considerably different variability. The parameters with large variability may be related to low availability of experimental data or a mismatch of timescales. The cross-correlations of parameters (corr) illustrated that more than 85% of the |corr| values were below 0.20, and the high |corr| values mainly resulted from equality constraints. With MLEs of estimated parameters, the estimated average

ecosystem C residence time was 56.8 ± 8.8 yr with a range from 16.6 ± 1.8 (cropland) to 85.9 ± 15.3 yr (ENF). Ecosystem C residence times and SDs have highly spatial heterogeneity and their values depend on vegetation types and climate conditions. Large uncertainties (i.e., high CV) in estimated C residence times appeared in the southern and eastern USA. Under the driving force of NPP changes from 1982 to 1998, the forward modelling estimated that terrestrial ecosystems in the conterminous USA would absorb about 0.20 ± 0.06 Pg C yr^{-1}. Uncertainties of ecosystem C uptake were strongly influenced by variability in C residence times and initial NPP values. The Bayesian approach developed here is powerful for assimilating multiple source of information to constrain parameter estimation and assess uncertainties in model prediction.

6. Acknowledgements

References

Ahmadov, R., Gerbig, C., Kretschmer, R., Koerner, S., Roedenbeck, C. and co-authors. 2009. Comparing high resolution WRF-VPRM simulations and two global CO_2 transport models with coastal tower measurements of CO_2. *Biogeosciences* **6**, 807–817.

Allen, H. L., Dougherty, P. M. and Campbell, R. G. 1990. Manipulation of water and nutrients – practice and opportunity in Southern United-States pine forests. *For. Ecol. Manage.* **30**, 437–453.

Barrett, D. J. 2002. Steady state turnover time of carbon in the Australian terrestrial biosphere. *Global Biogeochem. Cycles* **16**, 1108. DOI: 1110.1029/2002GB001860.

Beven, K. 2006. A manifesto for the equifinality thesis. *J. Hydrol.* **320**, 18–36.

Birdsey, R. A. and Heath, L. S. 1995. Carbon changes in US forests. In: *Climate Change and the Productivity of America's Forest: General Technical Report* (ed. L. A. Joyce). USDA Forest Service, Rocky Mountain Forest and Range Experiment Station, Ft. Collins, CO, pp. 56–70.

Bradford, J. B., Lauenroth, W. K. and Burke, I. C. 2005. The impact of cropping on primary production in the U.S. great plains. *Ecology* **86**, 1863–1872.

Braswell, B. H., Sacks, W. J., Linder, E. and Schimel, D. S. 2005. Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations. *Global Change Biol.* **11**, 335–355.

Brown, S., Schroeder, P. and Birdsey, R. 1997. Aboveground biomass distribution of US eastern hardwood forests and the use of large trees as an indicator of forest development. *For. Ecol. Manage.* **96**, 37–47.

Cannell, M. G. R. 1982. *World Forest Biomass and Primary Production Data*. Academic Press, London, New York.

DeAngelis, D. L., Gardner, R. H. and Shugart, H. H. 1997. NPP Multi-Biome: Global IBP Woodlands Data, 1955–1975. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, TN, USA. Online at: http://www.daac.ornl.gov

DeLucia, E. H., Hamilton, J. G., Naidu, S. L., Thomas, R. B., Andrews, J. A. and co-authors. 1999. Net primary production of a forest ecosystem with experimental CO_2 enrichment. *Science* **284**, 1177–1179.

El Maayar, M., Price, D. T., Black, T. A., Humphreys, E. R. and Jork, E. M. 2002. Sensitivity tests of the integrated biosphere simulator to soil and vegetation characteristics in a Pacific coastal coniferous forest. *Atmosphere-Ocean* **40**, 313–332.

Enting, I. G. 2002. *Inverse Problems in Atmospheric Constituent Transport*. Cambridge University Press, Cambridge, New York.

Field, C. B., Randerson, J. T. and Malmstrom, C. M. 1995. Global net primary rroduction – combining ecology and remote-sensing. *Remote Sens. Environ.* **51**, 74–88.

Fontaine, S., Barot, S., Barre, P., Bdioui, N., Mary, B. and co-authors. 2007. Stability of organic carbon in deep soil layers controlled by fresh carbon supply. *Nature* **450**, 277–280.

Friedlingstein, P., Cox, P., Betts, R., Bopp, L., Von Bloh, W. and co-authors. 2006. Climate-carbon cycle feedback analysis: results from the (CMIP)-M-4 model intercomparison. *J. Clim.* **19**, 3337–3353.

Garten, C. T. and Hanson, P. J. 2006. Measured forest soil C stocks and estimated turnover times along an elevation gradient. *Geoderma* **136**, 342–352.

Gaudinski, J. B., Trumbore, S. E., Davidson, E. A., Cook, A. C., Markewitz, D. and co-authors. 2001. The age of fine-root carbon in three forests of the eastern United States measured by radiocarbon. *Oecologia* **129**, 420–429.

Gelfand, A. E. and Smith, A. F. M. 1990. Sampling-based approaches to calculating marginal densities. *J. Am. Stat. Assoc.* **85**, 398–409.

Gill, R. A. and Jackson, R. B. 2000. Global patterns of root turnover for terrestrial ecosystems. *New Phytol.* **147**, 13–31.

Hallgren, W. S. and Pitman, A. J. 2000. The uncertainty in simulations by a Global Biome Model (BIOMES) to alternative parameter values. *Global Change Biol.* **6**, 483–495.

Hansen, M. C., Defries, R. S., Townshend, J. R. G. and Sohlberg, R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. *Int. J. Remote Sens.* **21**, 1331–1364.

Hastings, W. K. 1970. Monte-Carlo sampling methods using Markov Chains and their applications. *Biometrika* **57**, 97–109.

Heimann, M. and Reichstein, M. 2008. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**, 289–292.

Hicke, J. A., Asner, G. P., Randerson, J. T., Tucker, C., Los, S. and co-authors. 2002. Trends in North American net primary productivity derived from satellite observations, 1982–1998. *Global Biogeochem. Cycles* **16**, 1018. DOI: 1010.1029/2001GB001550.

Houghton, R. A., Hackler, J. L. and Lawrence, K. T. 1999. The US carbon budget: contributions from land-use change. *Science* **285**, 574–578.

IPCC 2007. Climate change 2007: the physical science basis (eds S. Solomon, D. Qin, M. Manning, Z. Chen, M. Marquis, K. B. Averyt, M. Tignor and H. L. Miller). Cambridge University Press, Cambridge, UK and New York, NY, USA.

Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E. and co-authors. 1996. A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**, 389–411.

Jenkinson, D. S., Meredith, J., Kinyamario, J. I., Warren, G. P., Wong, M. T. F. and co-authors. 1999. Estimating net primary production from measurements made on soil organic-matter. *Ecology* **80**, 2762–2773.

Kaminski, T., Knorr, W., Rayner, P. J. and Heimann, M. 2002. Assimilating atmospheric data into a terrestrial biosphere model: a case study of the seasonal cycle. *Global Biogeochem. Cycles* **16**, 1066. DOI: 1010.1029/2001GB001463.

Kleijnen, J. 1998. Experimental design for sensitivity analysis, optimization and validation of simulation models. In: *Handbook of Simulation – Principles, Methodology, Advances, Applications, and Practice* (ed. J. Banks). John Wiley, Hoboken, NJ, pp. 173–224.

Knorr, W. 2000. Annual and interannual CO_2 exchanges of the terrestrial biosphere: process-based simulations and uncertainties. *Global Ecol. Biogeogr.* **9**, 225–252.

Knorr, W. and Kattge, J. 2005. Inversion of terrestrial ecosystem model parameter values against eddy covariance measurements by Monte Carlo sampling. *Global Change Biol.* **11**, 1333–1351.

Kremer, J. N. 1983. Ecological implications of parameter uncertainty in stochastic simulation. *Ecol. Model.* **18**, 187–207.

Larocque, G. R., Bhatti, J. S., Boutin, R. and Chertov, O. 2008. Uncertainty analysis in carbon cycle models of forest ecosystems: research needs and development of a theoretical framework to estimate error propagation. *Ecol. Model.* **219**, 400–412.

Lobell, D. B., Hicke, J. A., Asner, G. P., Field, C. B., Tucker, C. J. and co-authors. 2002. Satellite estimates of productivity and light use efficiency in United States agriculture, 1982–98. *Global Change Biol.* **8**, 722–735.

Luo, Y. Q. 2007. Terrestrial carbon-cycle feedback to climate warming. *Ann. Rev. Ecol. Evol. Syst.* **38**, 683–712.

Luo, Y. Q., White, L. W., Canadell, J. G., DeLucia, E. H., Ellsworth, D. S. and co-authors. 2003. Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with inversion approach. *Global Biogeochem. Cycles* **17**, 1021. DOI: 1010.1029/2002GB001923.

Luo, Y. Q., Wu, L. H., Andrews, J. A., White, L., Matamala, R. and co-authors. 2001. Elevated CO_2 differentiates ecosystem carbon processes: deconvolution analysis of Duke Forest FACE data. *Ecol. Monogr.* **71**, 357–376.

Mastrandrea, M. D. and Schneider, S. H. 2004. Probabilistic integrated assessment of "dangerous" climate change. *Science* **304**, 571–575.

McCarthy, M. A. 2007. *Bayesian Methods for Ecology*. Cambridge University Press, Cambridge, UK and New York, NY.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. 1953. Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**, 1087–1092.

Murphy, J. M., Sexton, D. M. H., Barnett, D. N., Jones, G. S., Webb, M. J. and co-authors. 2004. Quantification of modelling uncertainties in a large ensemble of climate change simulations. *Nature* **430**, 768–772.

Pacala, S. W., Hurtt, G. C., Baker, D., Peylin, P., Houghton, R. A. and co-authors. 2001. Consistent land- and atmosphere-based US carbon sink estimates. *Science* **292**, 2316–2320.

Parton, W. J., Schimel, D. S., Cole, C. V. and Ojima, D. S. 1987. Analysis of factors controlling soil organic-matter levels in great-plains grasslands. *Soil Sci. Soc. Am. J.* **51**, 1173–1179.

Pillai, D., Gerbig, C., Marshall, J., Ahmadov, R., Kretschmer, R. and co-authors. 2010. High resolution modeling of CO_2 over Europe: implications for representation errors of satellite retrievals. *Atmos. Chem. Phys.* **10**, 83–94.

Pillai, D., Gerbig, C., Ahmadov, R., Roedenbeck, C., Kretschmer, R. and co-authors. 2011. High-resolution simulations of atmospheric CO_2 over complex terrain – representing the Ochsenkopf mountain tall tower. *Atmos. Chem. Phys.* **11**, 7445–7464.

Potter, C. S., Randerson, J. T., Field, C. B., Matson, P. A., Vitousek, P. M. and co-authors. 1993. Terrestrial ecosystem production – a process model-based on global satellite and surface data. *Global Biogeochem. Cycles* **7**, 811–841.

Powers, J. S., Montgomery, R. A., Adair, E. C., Brearley, F. Q., DeWalt, S. J. and co-authors. 2009. Decomposition in tropical forests: a pan-tropical study of the effects of litter type, litter placement and mesofaunal exclusion across a precipitation gradient. *J. Ecol.* **97**, 801–811.

Prihodko, L., Denning, A. S., Hanan, N. P., Baker, I. and Davis, K. 2008. Sensitivity, uncertainty and time dependence of parameters in a complex land surface model. *Agric. Forest Meteorol.* **148**, 268–287.

Raich, J. W. and Schlesinger, W. H. 1992. The global carbon-dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus Series B-Chem. Phys. Meteorol.* **44**, 81–99.

Randerson, J. T., Thompson, M. V. and Field, C. B. 1999. Linking C-13-based estimates of land and ocean sinks with predictions of carbon storage from CO_2 fertilization of plant growth. *Tellus Series B-Chem. Phys. Meteorol.* **51**, 668–678.

Randerson, J. T., Still, C. J., Balle, J. J., Fung, I. Y., Doney, S. C. and co-authors. 2002. Carbon isotope discrimination of arctic and boreal biomes inferred from remote atmospheric measurements and a biosphere-atmosphere model. *Global Biogeochem. Cycles* **16**, 1028. DOI: 1010.1029/2001GB001435.

Raupach, M. R., Rayner, P. J., Barrett, D. J., DeFries, R. S., Heimann, M. and co-authors. 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications. *Global Change Biol.* **11**, 378–397.

Rayner, P. J., Scholze, M., Knorr, W., Kaminski, T., Giering, R. and co-authors. 2005. Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS). *Global Biogeochem. Cycles* **19**, GB2026. DOI: 2010.1029/2004GB002254.

Saltelli, A., Chan, K. and Scott, E. M. 2000. *Sensitivity Analysis*. John Wiley, Chichester and New York.

Schimel, D. S., Braswell, B. H., Holland, E. A., Mckeown, R., Ojima, D. S. and co-authors. 1994. Climatic, edaphic, and biotic controls over storage and turnover of carbon in soils. *Global Biogeochem. Cycles* **8**, 279–293.

Schimel, D., Melillo, J., Tian, H. Q., McGuire, A. D., Kicklighter, D. and co-authors. 2000. Contribution of increasing CO_2 and climate to carbon storage by ecosystems in the United States. *Science* **287**, 2004–2006.

Scholze, M., Ciais, P. and Heimann, M. 2008. Modeling terrestrial C-13 cycling: climate, land use and fire. *Global Biogeochem, Cycles* **22**, GB1009. DOI: 1010.1029/2006GB002899.

Schulz, K. and Beven, K. J. 2003. Data-supported robust parameterizations in land surface-atmosphere flux predictions: towards a top-down approach. *Hydrological Processes* **17**, 2259–2277.

Sheffield, R. M. and Dickson, J. G. 1998. The south's forestland: on the hot seat to provide more. In: *Transactions of the 63rd North American Wildlife and Natural Resources Conference*. Wildlife Manage. Inst., Orlando, FL, pp. 316–331.

Silver, W. L. and Miya, R. K. 2001. Global patterns in root decomposition: comparisons of climate and litter quality effects. *Oecologia* **129**, 407–419.

Smith, C. and Brown, N. 1997. *Erdas Field Guide. (Fourth Edition)*. Atlanta, Georgia, USA, ERDAS ®, Inc.

Spear, R. C., Grieb, T. M. and Shang, N. 1994. Parameter uncertainty and interaction in complex environmental-models. *Water Resources Res.* **30**, 3159–3169.

Tarantola, A. 1987. *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, Amsterdam; Oxford; New-York.

Thompson, M. V. and Randerson, J. T. 1999. Impulse response functions of terrestrial carbon cycle models: method and application. *Global Change Biol.* **5**, 371–394.

Todd, D. E. and Hanson, P. J. 2003. Rates of coarse-wood decomposition. In: *North America temperate deciduous forest responses to changing precipitation regimes* (eds. P. J. Hanson and S. D. Wullschleger). Springer-Verbag, New York, pp. 210–214.

Turner, J. 1984. Radiocarbon dating of wood and charcoal in an Australian forest ecosystem. *Australian Foresty* **47**, 79–83.

Vogt, K. A., Vogt, D. J., Palmiotto, P. A., Boon, P., Ohara, J. and co-authors. 1996. Review of root dynamics in forest ecosystems grouped by climate, climatic forest type and species. *Plant Soil* **187**, 159–219.

Wang, Y. P. and Barrett, D. J. 2003. Estimating regional terrestrial carbon fluxes for the Australian continent using a multiple-constraint approach I. Using remotely sensed data and ecological observations of net primary production. *Tellus Series B-Chem. Phys. Meteorol.* **55**, 270–289.

Wang, Y. P., Leuning, R., Cleugh, H. A. and Coppin, P. A. 2001. Parameter estimation in surface exchange models using non-linear inversion: how many parameters can we estimate and which measurements are most useful? *Global Change Biol.* **7**, 495–510.

Wiegand, T. and Milton, S. J. 1996. Vegetation change in semi-arid communities – simulating probabilities and time scales. *Vegetatio* **125**, 169–183.

Xu, T., White, L., Hui, D. F. and Luo, Y. Q. 2006. Probabilistic inversion of a terrestrial ecosystem model: analysis of uncertainty in parameter estimation and model prediction. *Global Biogeochem. Cycles* **20**, GB2007. DOI: 2010.1029/2005gb002468.

Zaehle, S., Sitch, S., Smith, B. and Hatterman, F. 2005. Effects of parameter uncertainties on the modeling of terrestrial biosphere dynamics. *Global Biogeochem. Cycles* **19**, GB3020. DOI: 3010.1029/2004GB002395.

Zhang, D. Q., Hui, D. F., Luo, Y. Q. and Zhou, G. Y. 2008. Rates of litter decomposition in terrestrial ecosystems: global patterns and controlling factors. *J. Plant Ecol.-Uk* **1**, 85–93.

Zhou, T. and Luo, Y. Q. 2008. Spatial patterns of ecosystem carbon residence time and NPP-driven carbon uptake in the conterminous United States. *Global Biogeochem. Cycles* **22**, GB3032. DOI: 3010.1029/2007gb002939.

Zhou, X. H., Weng, E. S. and Luo, Y. Q. 2008. Modeling patterns of nonlinearity in ecosystem responses to temperature, CO_2, and precipitation changes. *Ecol. Appl.* **18**, 453–466.