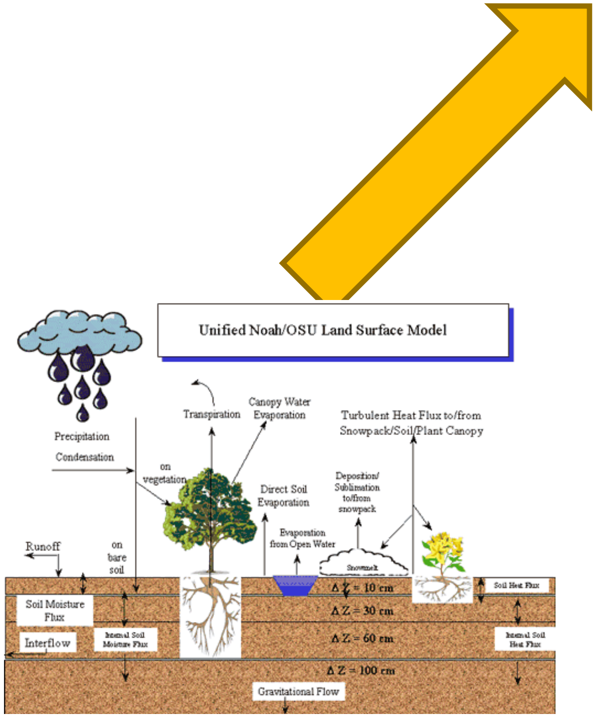
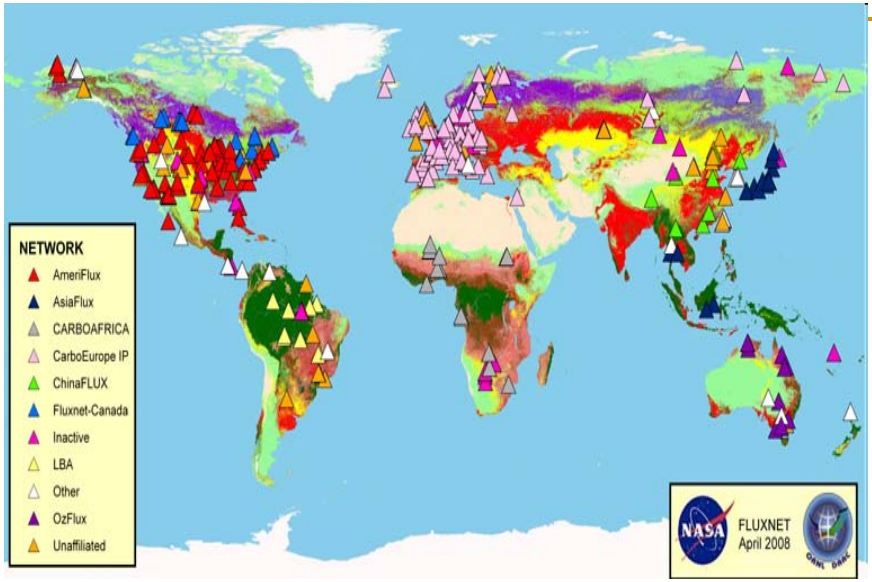
Ecological Data Assimilation to Constrain Regional and Global Models

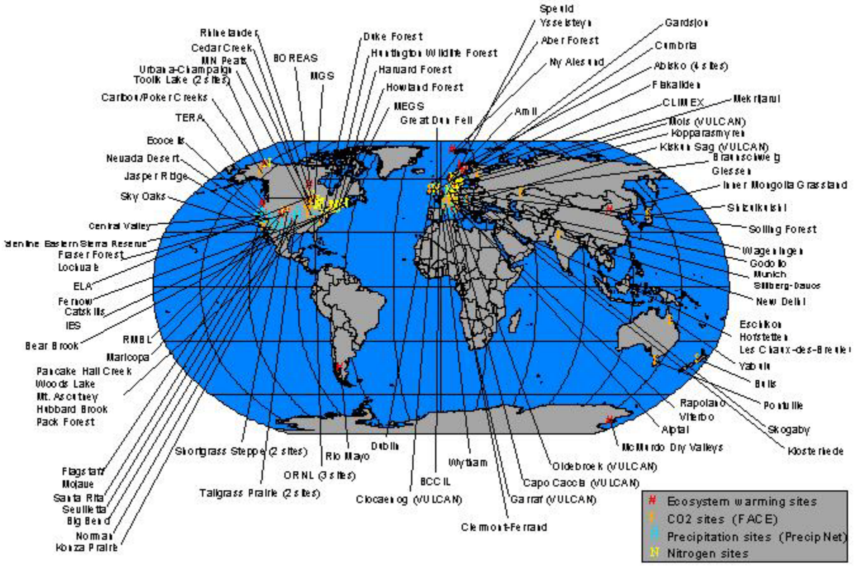
Yiqi Luo, Tao Zhou, Xuhui Zhou, Ensheng Weng

University of Oklahoma

Future states of climate and ecosystem

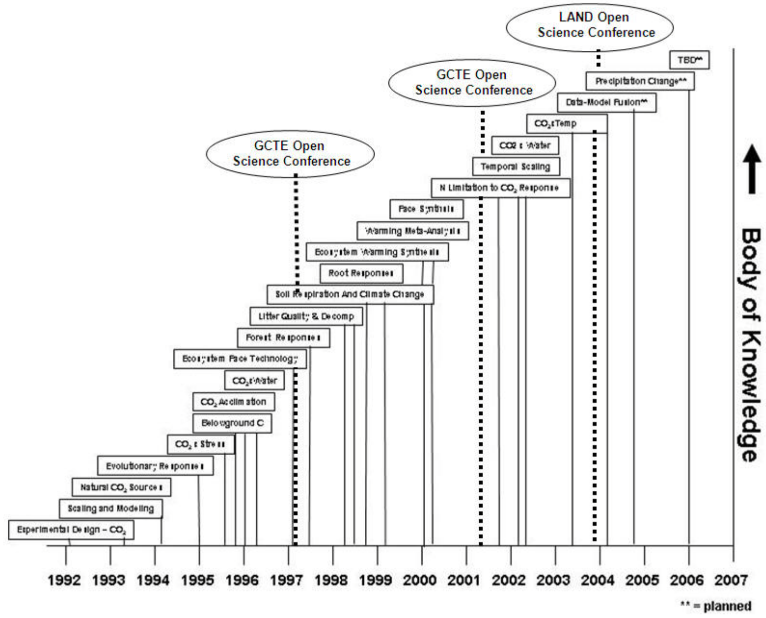




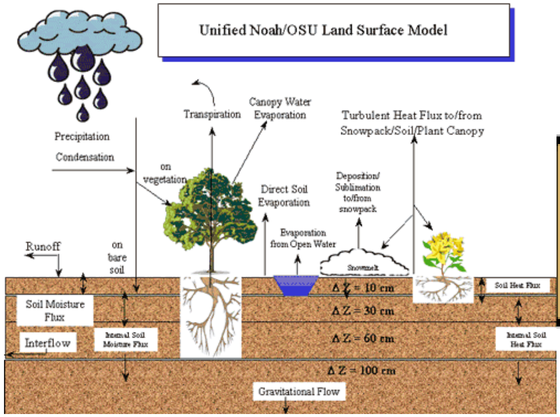


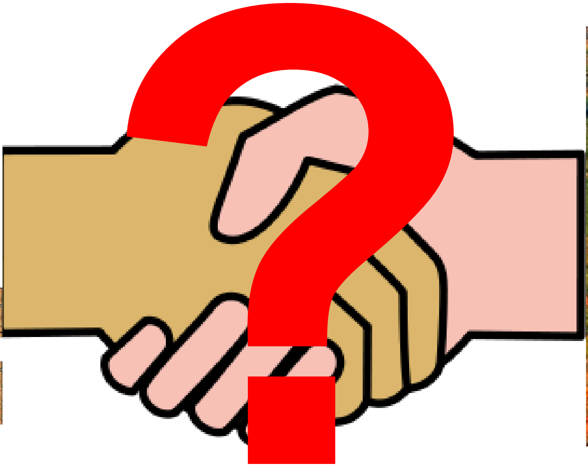
Terrestrial Ecosystem Response To Atmospheric and Climate Change

(GCTE Sponsored and Co-Sponsored Workshops)



Future states of climate and ecosystem





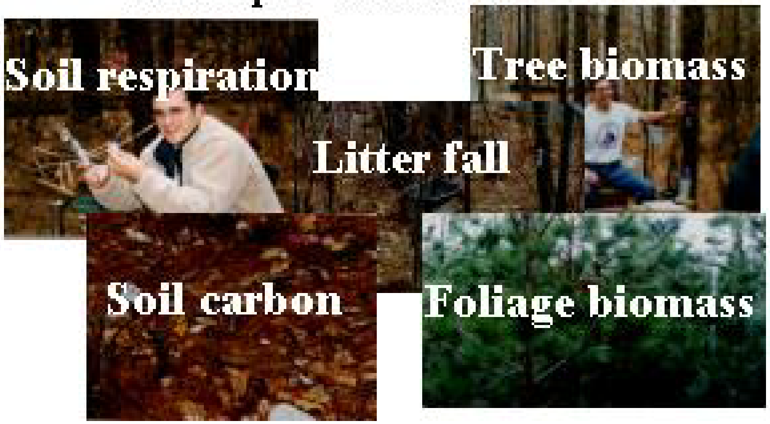
Data from experiments and observations

How to fuse data with models towards prediction?

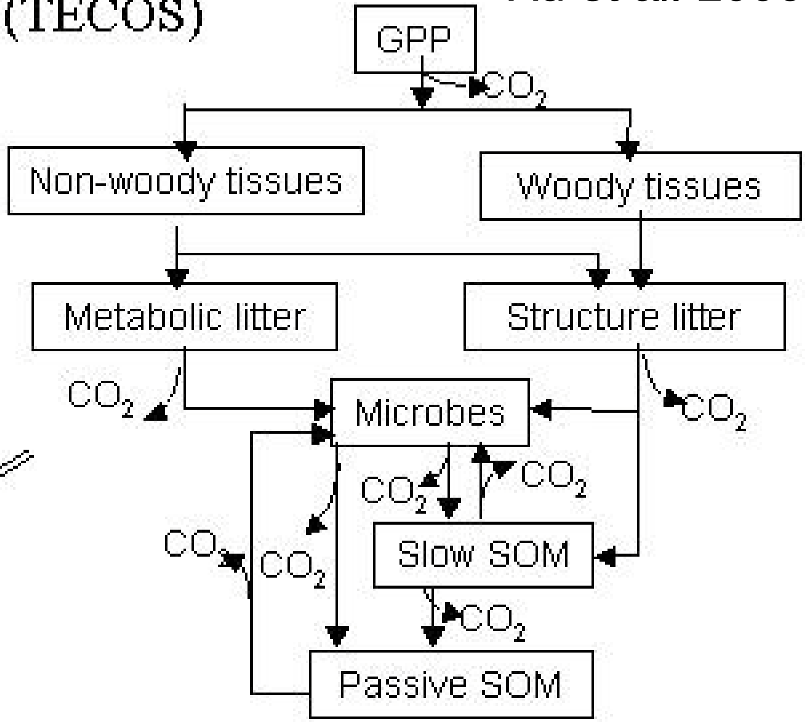
Data assimilation techniques applied to multiple, spatially distributed ecological measurements

Luo et al. 2003
Xu et al. 2006

Multiple data sets

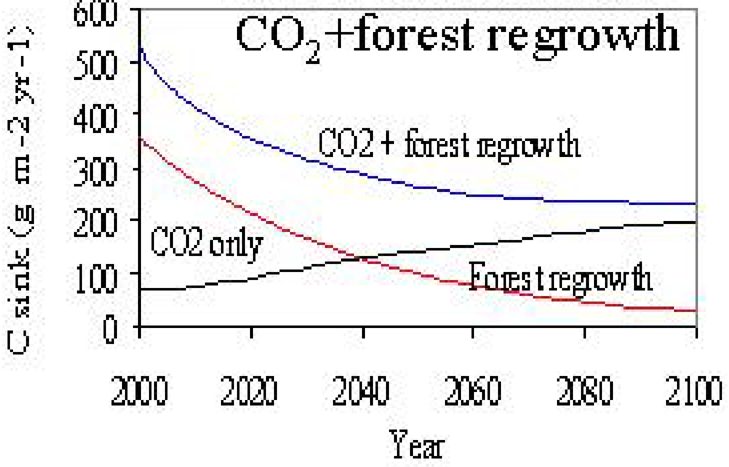


Multi-pool model (TECOS)



Multiple parameter values

Pool	Residence time (yr)	
	$-CO_2$	$+CO_2$
Nonwoody biomass	1.557	1.263
Woody biomass	27.369	19.429
Metabolic litter	0.128	0.121
Structural litter	3.242	2.839
Microbial biomass	0.321	1.081
Slow SOM	30.523	49.134
Passive SOM	885.027	1031.368


Future C sinks due to CO_2+forest regrowth

Techniques

Levenburg-Marquardt minimization with quasi-Monte Carlo algorithm (White and Luo 2002)

MCMC Metropolis-Hastings algorithms (Xu et al. 2006)

Genetic algorithms (Zhou and Luo 2008)

Ensemble Kalman Filter (Gao et al. in revision)

Applications

Deconvolution of soil respiration (Luo et al. 2001)

CO_2 effects on carbon residence times (Luo et al. 2003)

Uncertainty analysis of residence times (Xu et al. 2006; Zhou et al. in revision)

Propagation of measurement errors (Weng et al. in review)

Ecological forecasting (Gao et al. in revision)

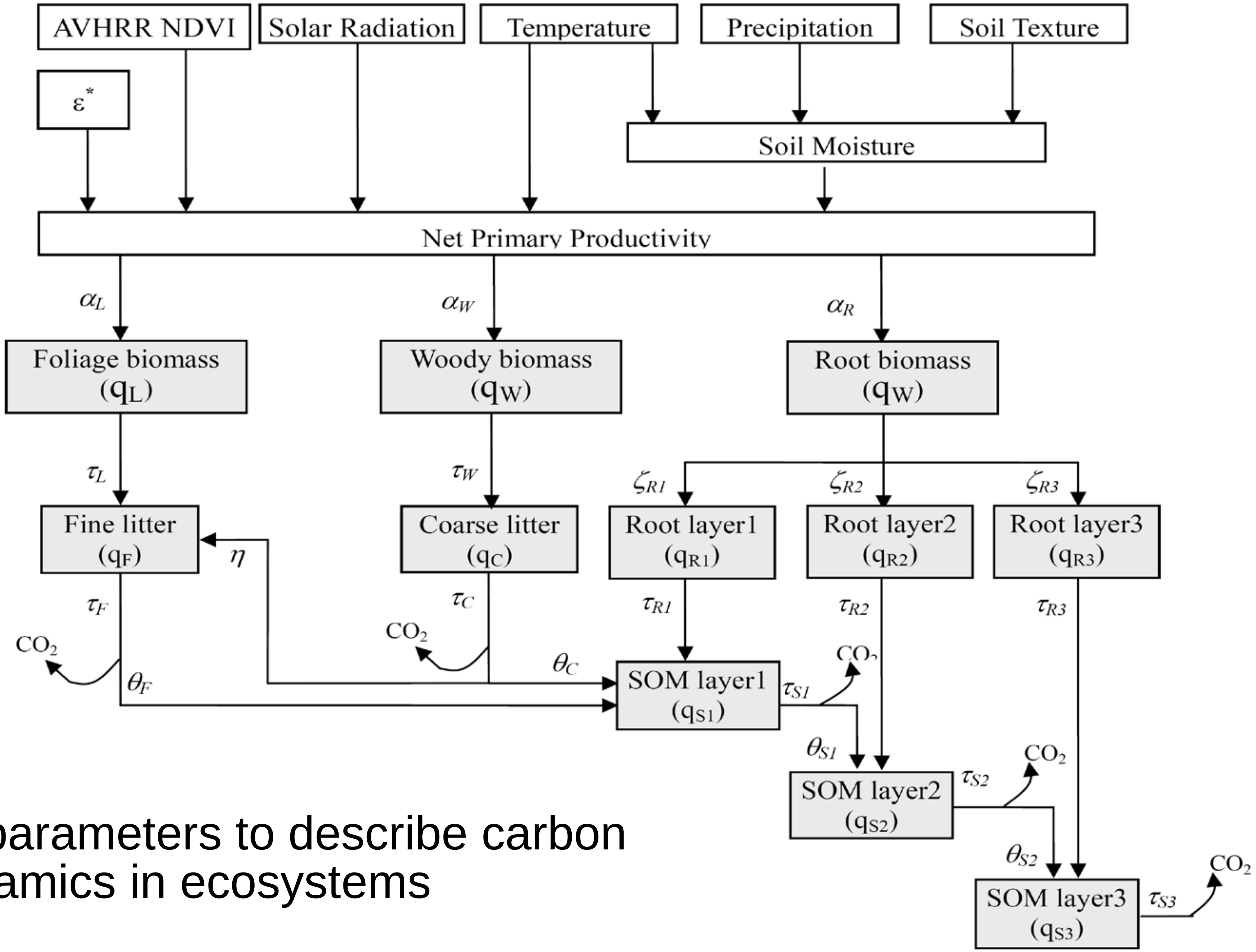
Information contents of model structure, data, and parameters (Weng et al. to be submitted)

Regional carbon residence times and sequestration (Zhou and Luo 2008; Zhou et al. in revision)

Global distributions of Q_{10} (Zhou et al. 2009)

Data assimilation with spatially distributed ecological measurements

Regional Terrestrial Ecosystem (TECO) Model



22 parameters to describe carbon dynamics in ecosystems

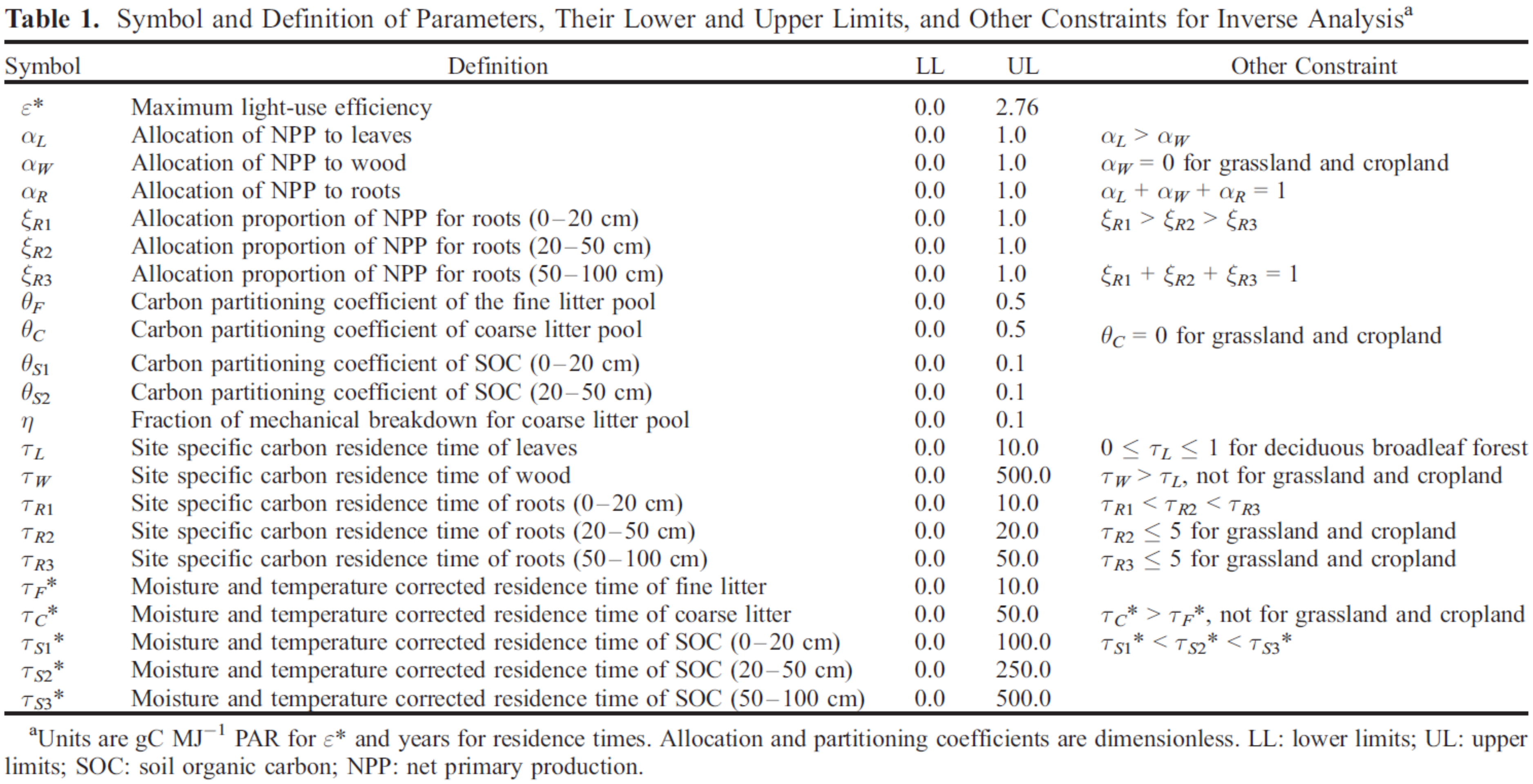
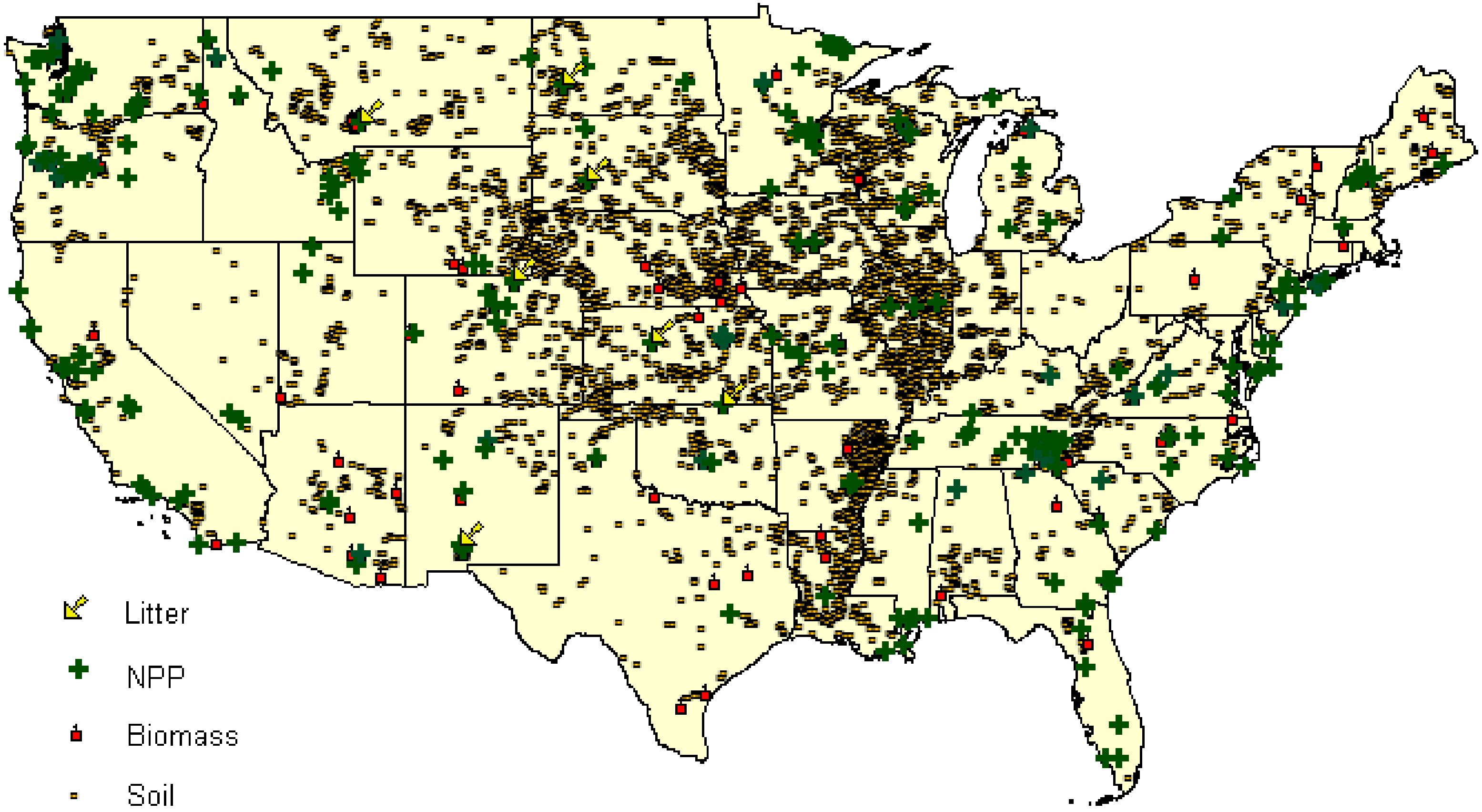


Table 1. Symbol and Definition of Parameters, Their Lower and Upper Limits, and Other Constraints for Inverse Analysis[a]

Symbol	Definition	LL	UL	Other Constraint
ε^*	Maximum light-use efficiency	0.0	2.76	
α_L	Allocation of NPP to leaves	0.0	1.0	$\alpha_L > \alpha_W$
α_W	Allocation of NPP to wood	0.0	1.0	$\alpha_W = 0$ for grassland and cropland
α_R	Allocation of NPP to roots	0.0	1.0	$\alpha_L + \alpha_W + \alpha_R = 1$
ξ_{R1}	Allocation proportion of NPP for roots (0–20 cm)	0.0	1.0	$\xi_{R1} > \xi_{R2} > \xi_{R3}$
ξ_{R2}	Allocation proportion of NPP for roots (20–50 cm)	0.0	1.0	
ξ_{R3}	Allocation proportion of NPP for roots (50–100 cm)	0.0	1.0	$\xi_{R1} + \xi_{R2} + \xi_{R3} = 1$
θ_F	Carbon partitioning coefficient of the fine litter pool	0.0	0.5	
θ_C	Carbon partitioning coefficient of coarse litter pool	0.0	0.5	$\theta_C = 0$ for grassland and cropland
θ_{S1}	Carbon partitioning coefficient of SOC (0–20 cm)	0.0	0.1	
θ_{S2}	Carbon partitioning coefficient of SOC (20–50 cm)	0.0	0.1	
η	Fraction of mechanical breakdown for coarse litter pool	0.0	0.1	
τ_L	Site specific carbon residence time of leaves	0.0	10.0	$0 \leq \tau_L \leq 1$ for deciduous broadleaf forest
τ_W	Site specific carbon residence time of wood	0.0	500.0	$\tau_W > \tau_L$, not for grassland and cropland
τ_{R1}	Site specific carbon residence time of roots (0–20 cm)	0.0	10.0	$\tau_{R1} < \tau_{R2} < \tau_{R3}$
τ_{R2}	Site specific carbon residence time of roots (20–50 cm)	0.0	20.0	$\tau_{R2} \leq 5$ for grassland and cropland
τ_{R3}	Site specific carbon residence time of roots (50–100 cm)	0.0	50.0	$\tau_{R3} \leq 5$ for grassland and cropland
$\tau_F{}^*$	Moisture and temperature corrected residence time of fine litter	0.0	10.0	
$\tau_C{}^*$	Moisture and temperature corrected residence time of coarse litter	0.0	50.0	$\tau_C{}^* > \tau_F{}^*$, not for grassland and cropland
$\tau_{S1}{}^*$	Moisture and temperature corrected residence time of SOC (0–20 cm)	0.0	100.0	$\tau_{S1}{}^* < \tau_{S2}{}^* < \tau_{S3}{}^*$
$\tau_{S2}{}^*$	Moisture and temperature corrected residence time of SOC (20–50 cm)	0.0	250.0	
$\tau_{S3}{}^*$	Moisture and temperature corrected residence time of SOC (50–100 cm)	0.0	500.0	

[a] Units are gC MJ^{-1} PAR for ε^* and years for residence times. Allocation and partitioning coefficients are dimensionless. LL: lower limits; UL: upper limits; SOC: soil organic carbon; NPP: net primary production.

Data sets: NPP in leaves, stems, and roots, biomass in leaves, stems, fine litter, and roots and SOC in the three soil layers.



Biome types

- ENF – Evergreen needleleaf forest
- DBF – Deciduous broadleaf forest
- MF – Mixed forest
- W – Woodland
- WG – Wooded grassland
- S – Shrubland
- G – Grassland
- C – Cropland

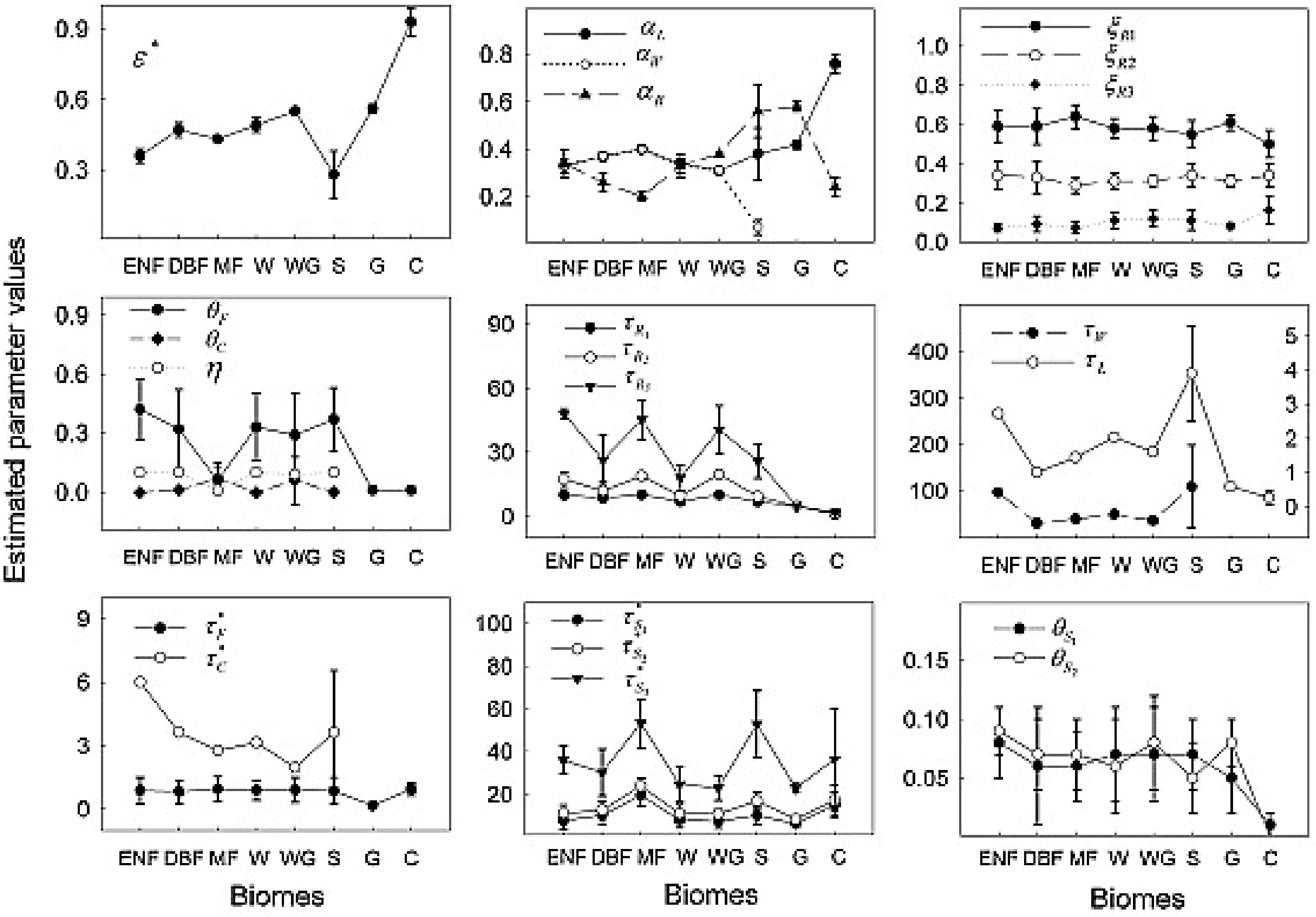
Parameter estimation - weighed least squares principle

Cost function

$$J = \sum_{m=1}^{M} \lambda_m \left\{ \sum_{n=1}^{N_m} [y_{nm} - \hat{y}_{nm}(x_n, \mathbf{a})]^2 \right\}, \qquad m = 1, 2\ , \ldots ,\ M$$

a is a vector of 22 parameters

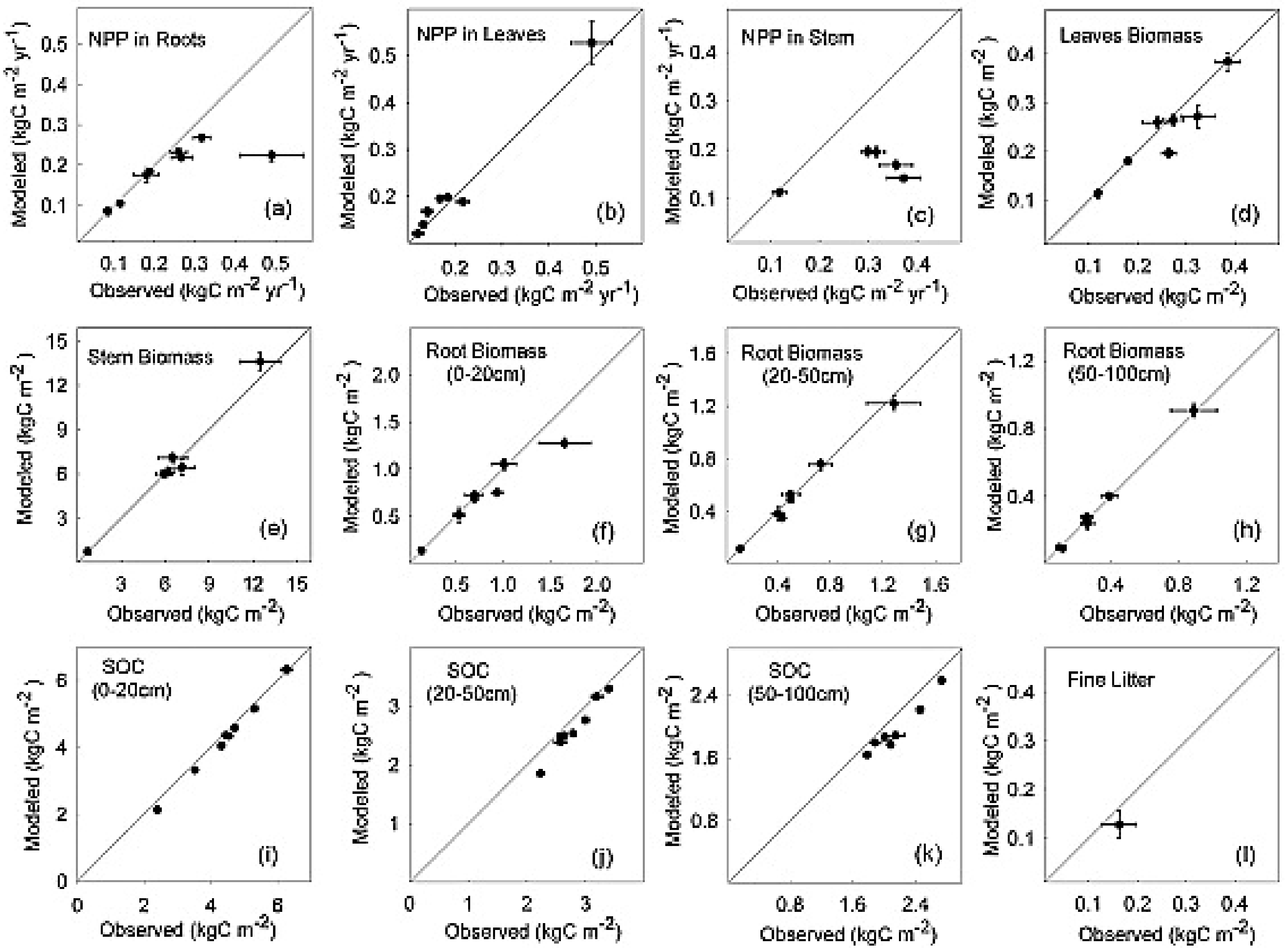
x_n is an auxiliary forcing vector



Optimized values of 22 parameters for eight biomes

with mean ± one standard deviation of optimized values from 30 runs of genetic algorithm

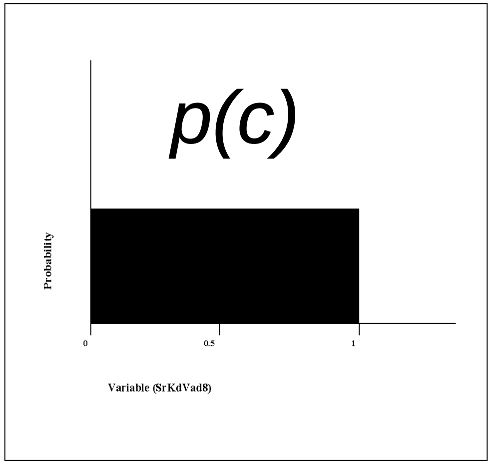
Zhou and Luo 2008 GBC



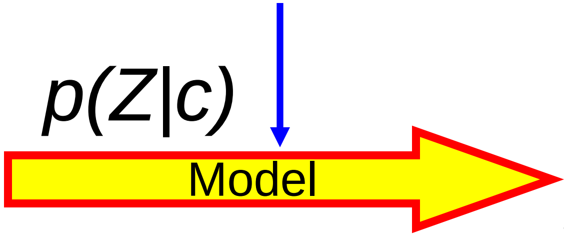
Zhou and Luo 2008 GBC

Stochastic inversion: Bayesian approach

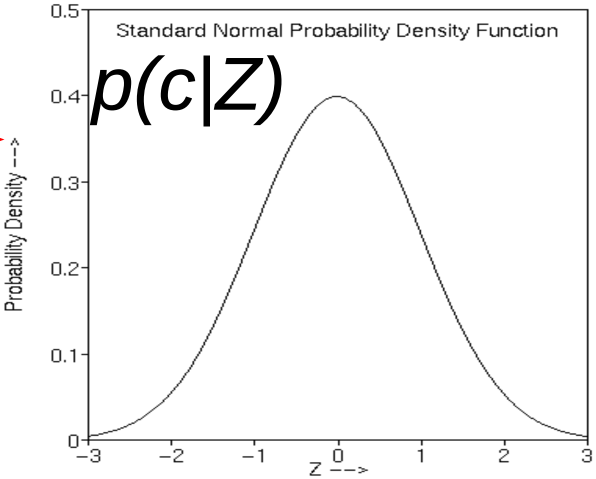
Observed Data



Prior knowledge

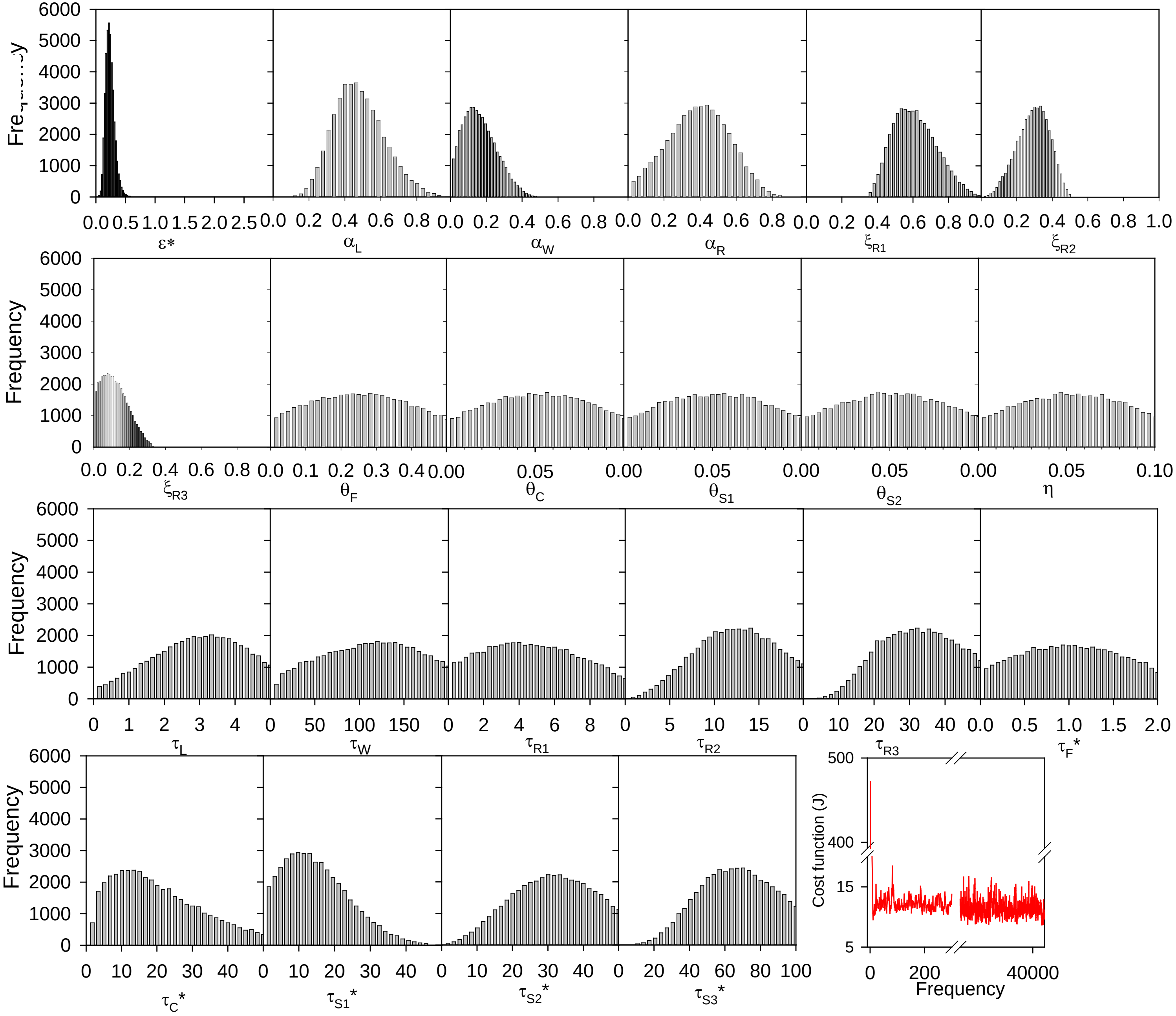


$$p(c|Z) \propto p(Z|c)p(c)$$



Posterior information

Probability density function

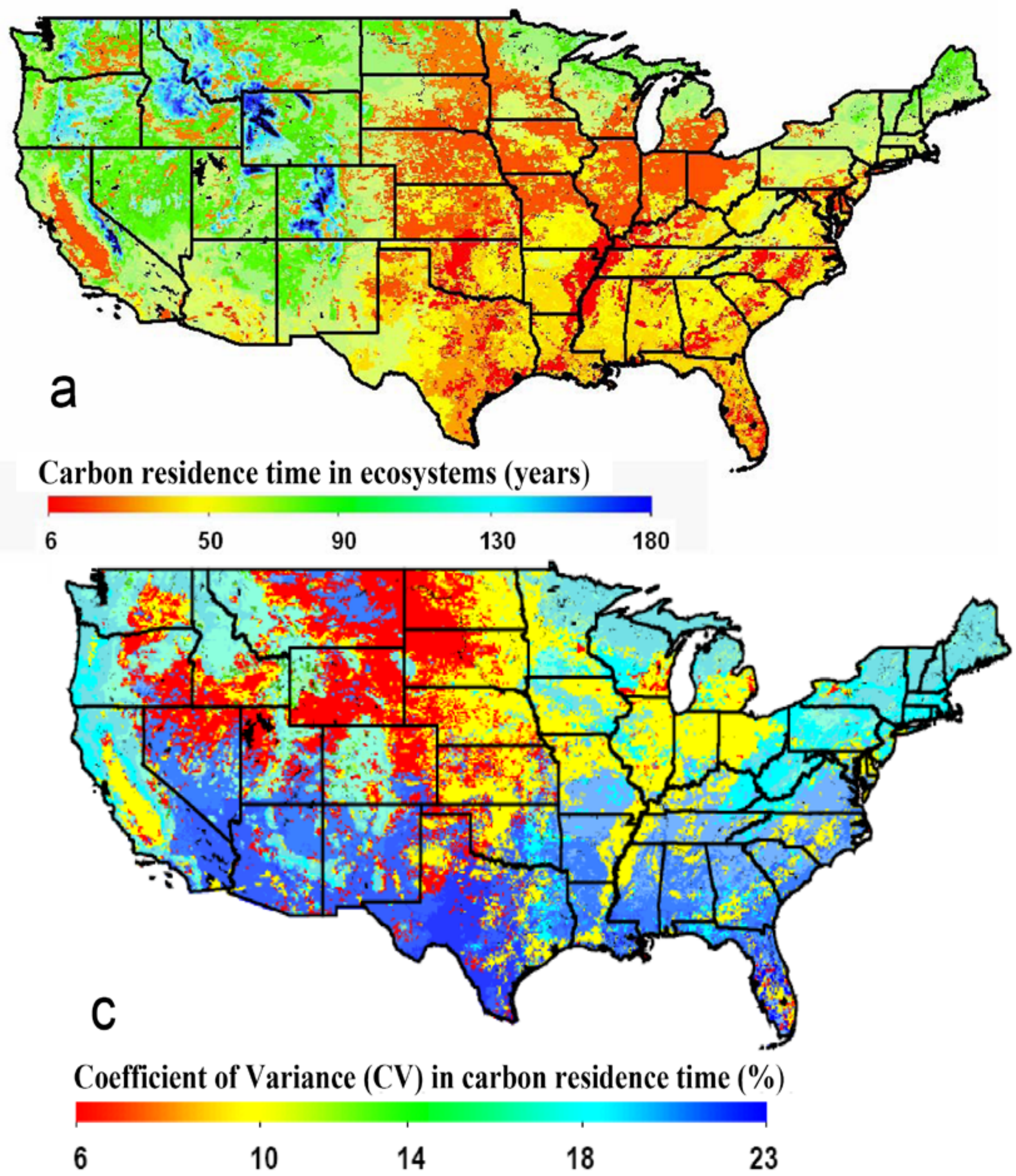


Zhou et al. In revision

Ecosystem carbon residence time

High in cold and forest regions

Low in agricultural areas



Uncertainty

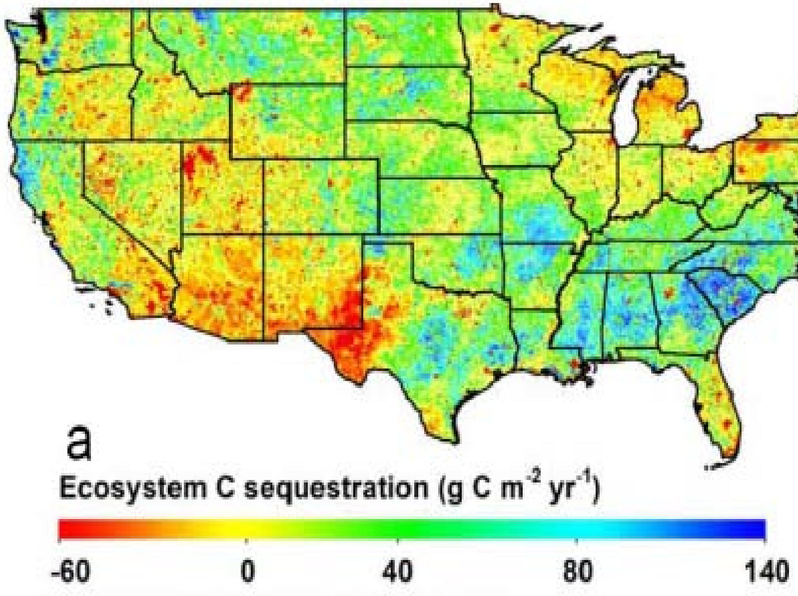
High uncertainty in southwest, south, and southeast is probably due to low data density in those regions

Zhou et al. In revision

Ecosystem carbon sequestration

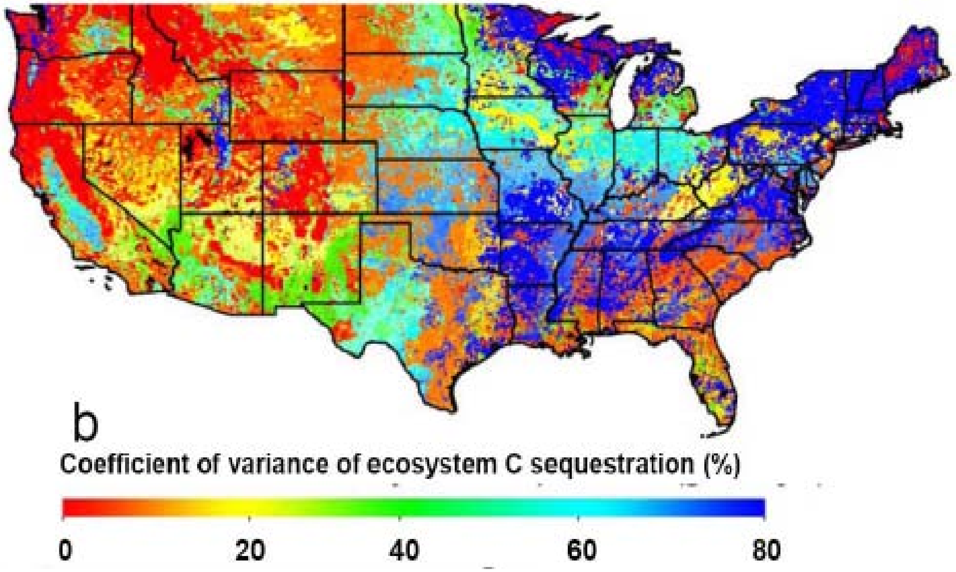
High in southeast forests and central croplands

Low or negative in southwest regions



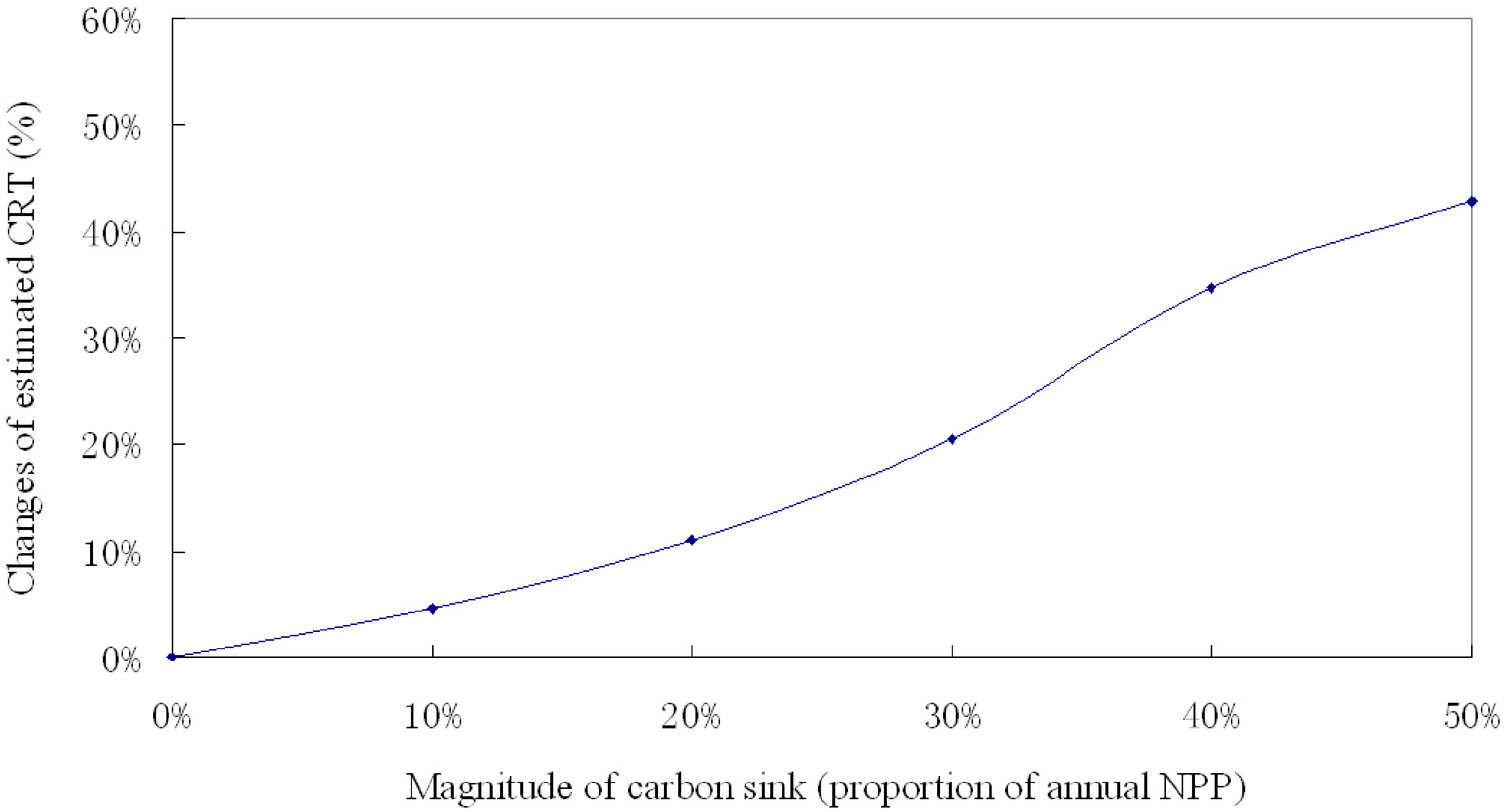
Uncertainty

high in east forest ecosystems due to high uncertainty in residence times and low data density.



Zhou et al. In revision

Legacy effects



Sensitivity analysis of the estimated C residence times in nonsteady state. Estimated C residence times increase as the magnitude of carbon uptake varied from 10 to 50% of the total NPP

Zhou and Luo 2008 GBC

Global pattern of Q_{10} and its implications

Parameter and inversion algorithm

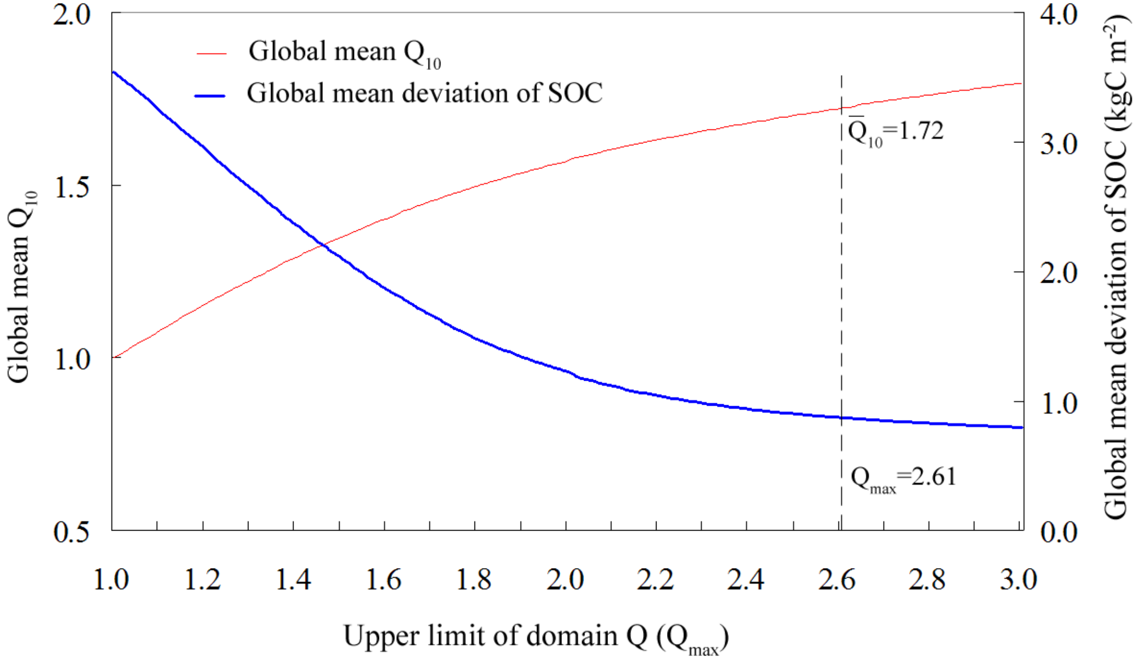
CASA model and the parameter

$$CO_2(x,t)_i = C(x,t)_i \cdot k_i \cdot W_s(x,t) \cdot T_s(x,t) \cdot (1 - M_\varepsilon)$$

$$T_s(x,t) = Q_{10}^{[(T(x,t)-35)/10]}$$

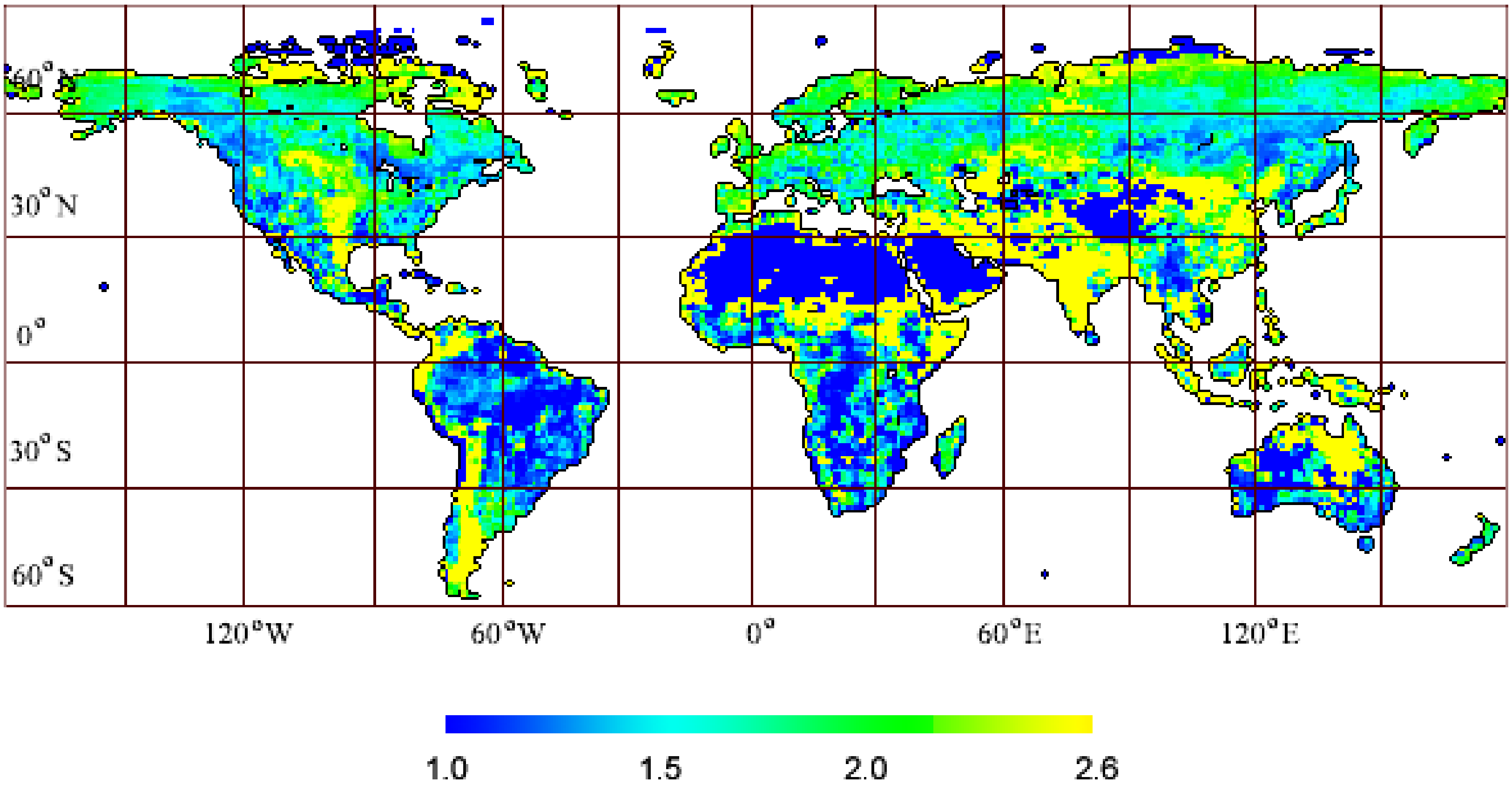
Cost function

$$J(Q) = \frac{\sum_x \left| S_{m,x}(Q_{10}^0(x)) - S_{0,x} \right| \times a(x)}{\sum_x a(x)}$$



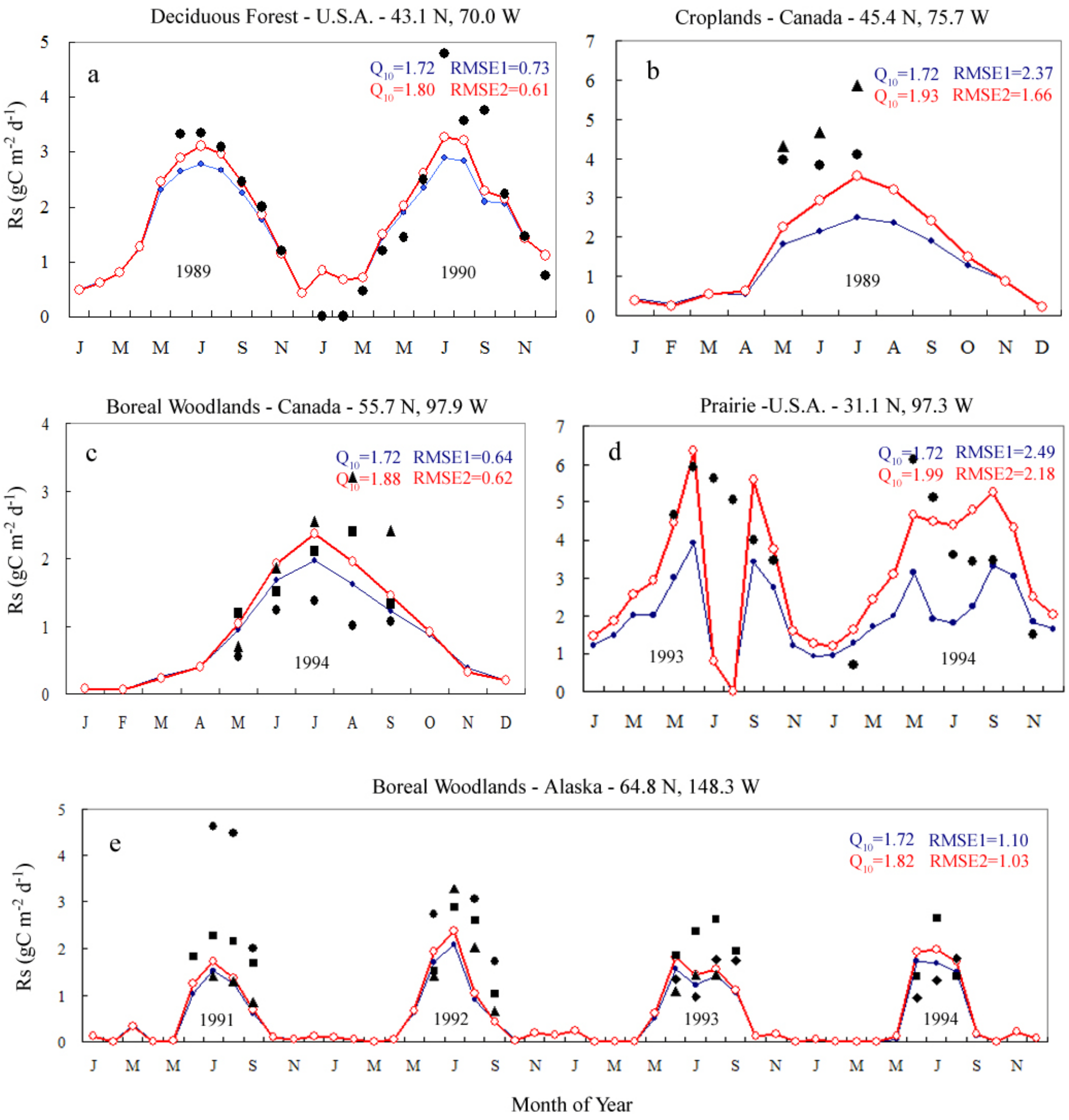
Relationship between optimal global mean Q10 value and upper limit of domain Q

Zhou et al. 2009 JGR-Biogeosciences



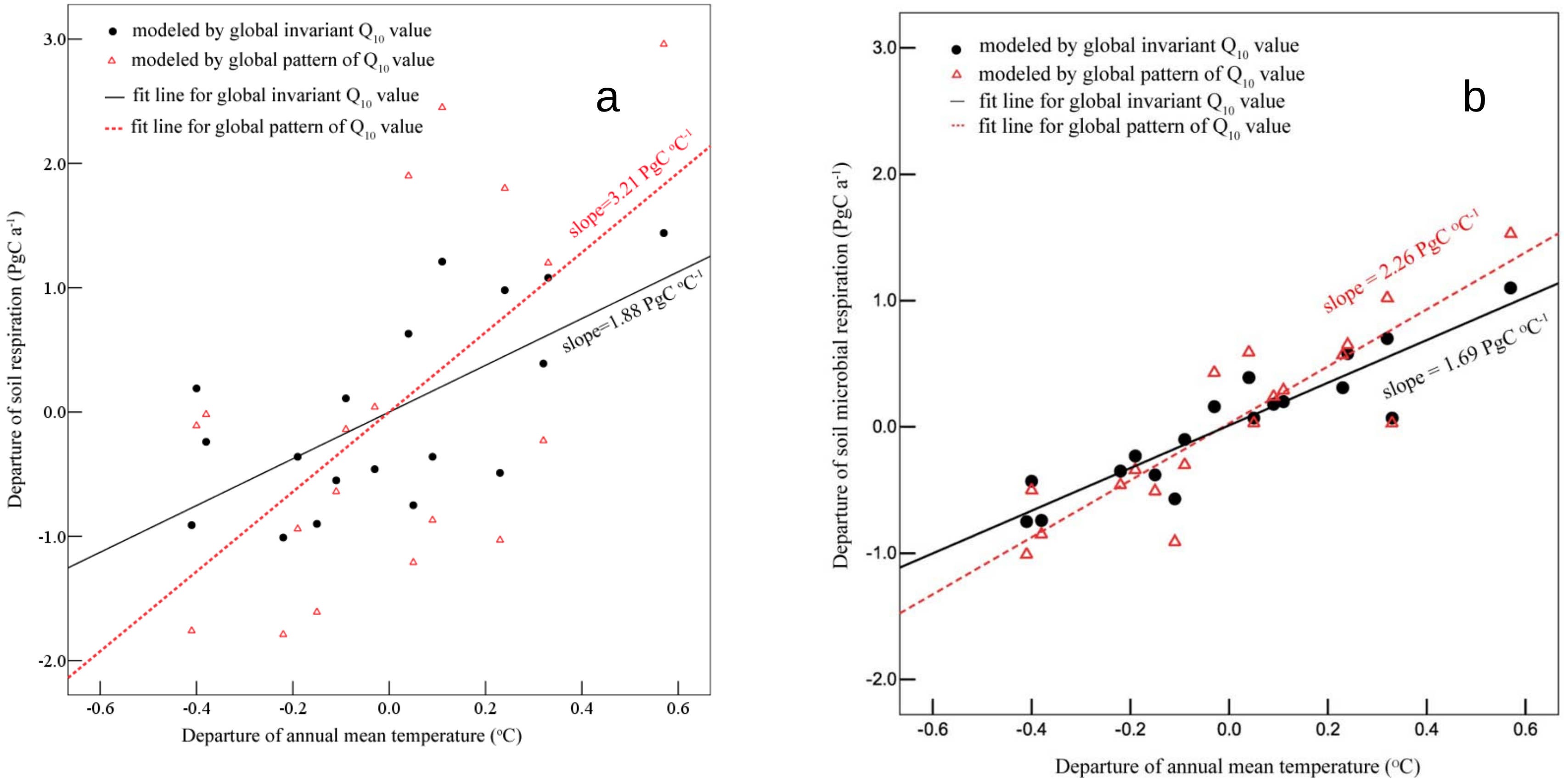
Spatial pattern of the optimal Q10 values. In general, tundra, C3 and C4 grasslands, shrublands, and croplands have higher Q10 values than deserts, bare grounds, broadleaf deciduous forests, and woodlands.

Zhou et al. 2009 JGR-Biogeosciences



Zhou et al. 2009 JGR-Biogeosciences

Relationship between the departure of annual soil respiration and annual mean temperature

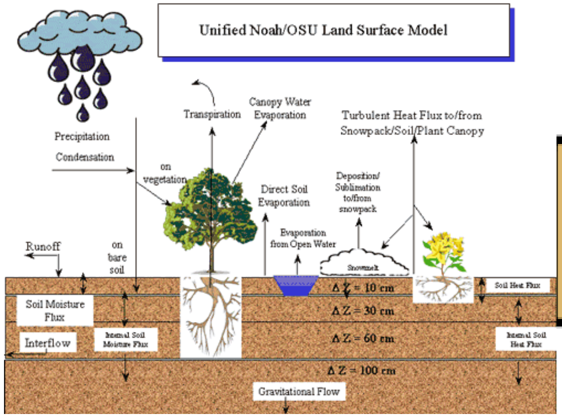


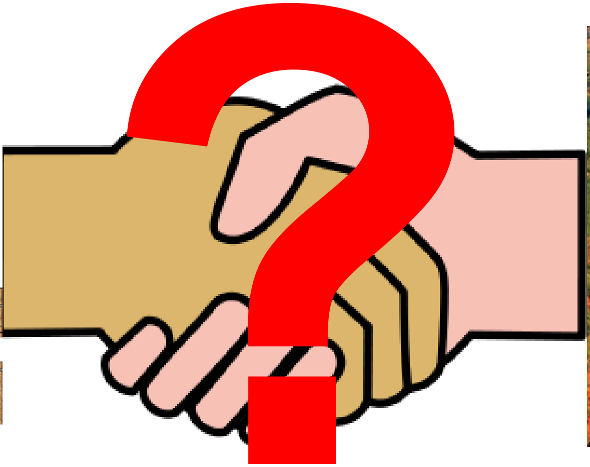
Zhou et al. 2009 JGR-Biogeosciences

Summary

- We have developed techniques to pool spatially distributed ecological data to constrain regional and global models
- Almost all of parameters are constrained to different degrees
- Ecosystem C residence time ranged from 16 in croplands to 86 years in evergreen forests with an average of 57 years. Large uncertainty appeared in the southern and eastern USA
- The estimated C sequestration was 0.20 Pg C yr^{-1} with the largest portion in the evergreen forests, grasslands, and croplands with large uncertainty in the central and eastern USA.
- Spatially distributed Q_{10} can help improve model predictions of carbon cycles

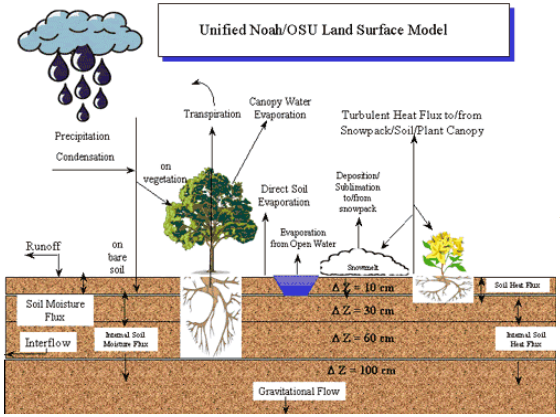
Future states of climate and ecosystem





Data from experiments and observations

Future states of climate and ecosystem



Data from experiments and observations

23rd New Phytologist Symposium

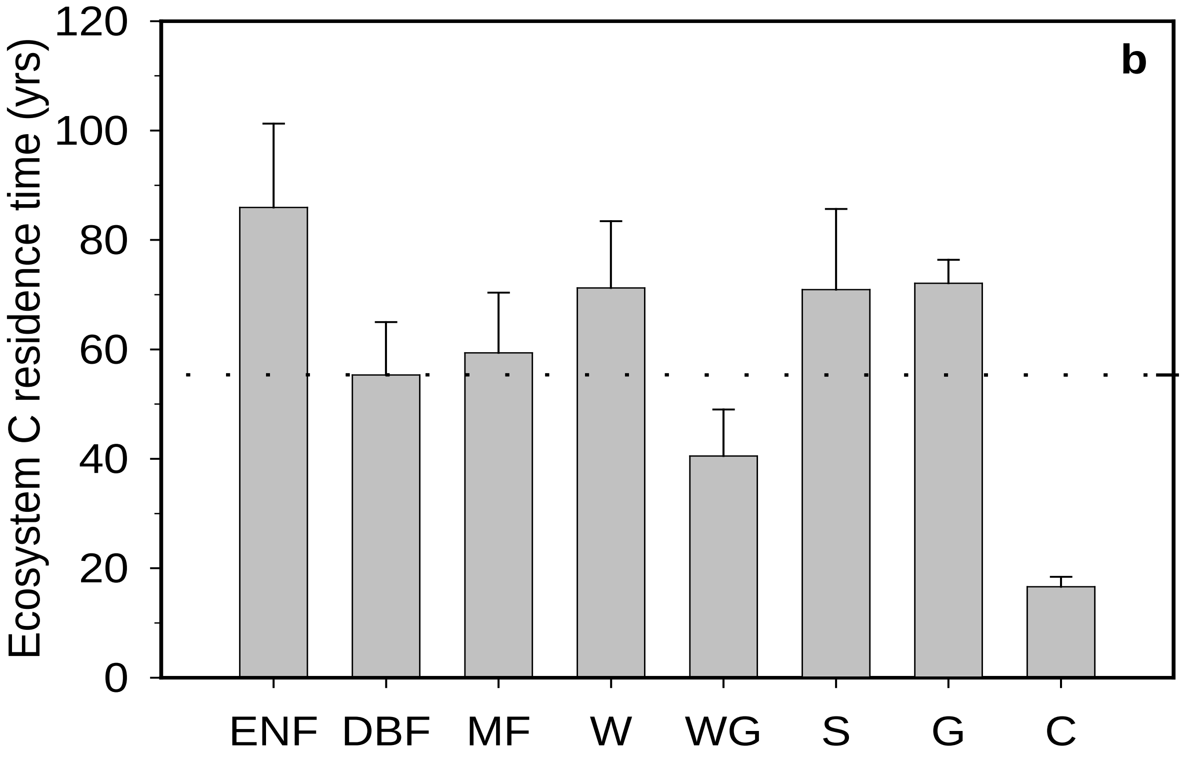
Carbon cycling in tropical ecosystems

Yanling Hotel, Guangzhou, China
17–20 November 2009

Organisation

Ian Alexander *University of Aberdeen, UK*
Yiqi Luo *University of Oklahoma, USA*
Juxiu Liu *South China Botanical Garden, China*
Rich Norby *Oak Ridge National Laboratory, USA*
Xuli Tang *South China Botanical Garden, China*
Yan Yu *South China Botanical Garden, China*
Guoyi Zhou *South China Botanical Garden, China*

New Phytologist **www.newphytologist.org**



ENF - Evergreen needleleaf forest

DBF - Deciduous broadleaf forest

MF - Mixed forest

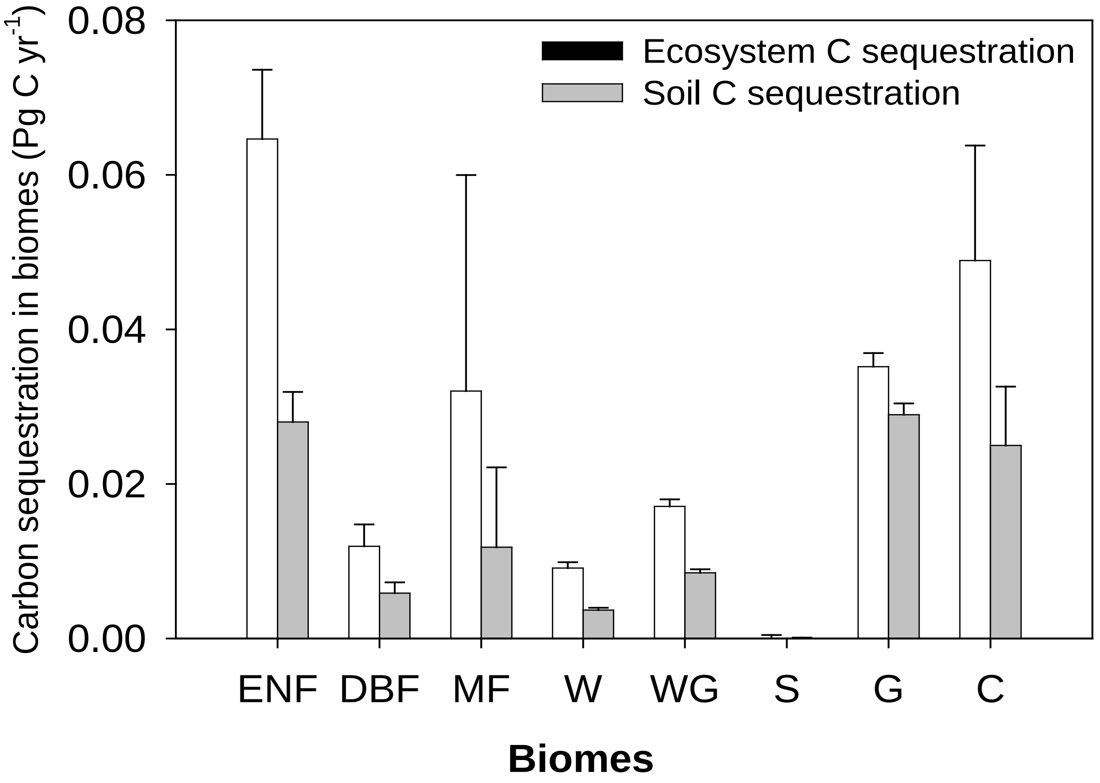
W – Woodland

WG - Wooded grassland

S – Shrubland

G – Grassland

C - Cropland



ENF - Evergreen needleleaf forest

DBF - Deciduous broadleaf forest

MF - Mixed forest

W – Woodland

WG - Wooded grassland

S – Shrubland

G – Grassland

C - Cropland

Area-weighted mean values of Q10, mean annual temperature (MAT), and mean annual precipitation (MAP)

NO.	Land cover type	Number of grids	Area (km^2)	Q_{10}	MAT (℃)	MAP (mm y^{-1})
1	Broadleaf evergreen forest	1093	1.34E+07	1.50	25.0	2201
2	Broadleaf deciduous forest and woodland	318	3.28E+06	1.75	16.6	961
3	Mixed forest and woodland	763	6.55E+06	1.61	9.2	934
4	Coniferous forest and woodland	2000	1.29E+07	1.69	-2.1	547
5	High latitude deciduous forest and woodland	952	5.75E+06	1.61	-5.7	442
6	Wooded C_4 grassland	1456	1.71E+07	1.59	23.1	1324
7	C_4 grassland	779	8.93E+06	2.02	23.8	580
9	Shrubs	1034	1.10E+07	1.82	17.6	266
10	Tundra	1507	7.00E+06	2.03	-10.4	335
11	Desert, bare ground	1598	1.68E+07	1.43	20.6	96
12	Cultivation	1368	1.33E+07	2.01	14.7	832
14	C_3 wooded grassland	440	4.46E+06	1.66	14.6	1145
15	C_3 grassland	1235	1.14E+07	1.96	7.3	423
All	Global average	14543	1.32E+08	1.72	13.7	789

Means and uncertainty (SD) of estimated parameters