Supplementary Information for "A primer for data assimilation with ecological models using Markov Chain Monte Carlo (MCMC)"

J.M. Zobitz*
Department of Mathematics
Augsburg College
2211 Riverside Avenue
Minneapolis, MN 55454
zobitz@augsburg.edu
+1-612-330-1065
fax: +1-612-330-1393

A.R. Desai
Department of Atmospheric & Oceanic Sciences
University of Wisconsin-Madison
1225 W Dayton St
Madison, WI 53706 USA
+1-608-265-9201
desai@aos.wisc.edu

D. J. P. Moore
Department of Geography
King's College London
Strand, London, WC2R 2LS, UK.
+44 (0)20 7848 2571
dave.moore@kcl.ac.uk

Current address:
School of Natural Resources and Environment
University of Arizona
1955 E. Sixth Street, Suite 205, Tucson, AZ 85721
davidjpmoore@email.arizona.edu

M. A. Chadwick
Department of Geography
King's College London
Strand, London, WC2R 2LS, UK.
+44 (0)20-7848-2641
michael.chadwick@kcl.ac.uk

* Denotes Corresponding Author

The data and programs used in the outlined examples are available at:
http://www.augsburg.edu/home/math/faculty/zobitz/dataAssimilation.html
or by contacting the corresponding author.

I. Six helpful rules of thumb for data assimilation

Through the examples outlined in our paper we have illustrated the applicability of data assimilation across a range of ecological areas and measurement techniques. Example 1 focused on techniques of model selection and interpreting model results and prediction uncertainties. Example 2 focused on using data assimilation to test model structure and generating scientific hypotheses through model manipulations. From these examples, we have found six helpful rules of thumb to remember when undertaking data assimilation. These rules are applicable to any study even though the model and data assimilation algorithm may vary.

Rule 1: Be a scientist

The first rule is a helpful reminder attempting any data assimilation routine. Data assimilation is not magic; it is a tool. As with any scientific study, the goal of any data assimilation scheme depends on the scientific hypothesis to be tested and questions asked. Data assimilation quantifies uncertainty estimates for the study results, validating or rejecting the study hypotheses. In cases where the study hypotheses were not confirmed, then either the model or measurements were non-informative, or did not help reduce any *a priori* uncertainty on the process to be studied.

Rule 2: Know your data & model

Interdisciplinary collaboration in the sciences is increasing (Porter and Rafols, 2009). One application of data assimilation has been the synthesis of different datasets (Doney and Ducklow, 2006; Friend et al., 2007), where interdisciplinary collaboration is crucial. With the implementation of programs such as NCEAS, NEON, and FLUXNET, public access and distribution of data is likely to become even more common. We are of the opinion that the proliferation of data to a broader audience will broaden the impact of the data. However caution must be exercised when performing data assimilation. Publicly shared data should follow best practices for identifying any gaps in the data (Zimmerman, 2008). When doing data assimilation, pay attention to data flags. From our experiences incorrect conclusions can be made by not understanding both the data and the model. Generally speaking, gap-filled data should not be assimilated - these data are already a secondary product. While we recognize that it may be necessary to include gap-filled data to ensure model continuity, make sure these data are flagged and not used in parameter or state estimation.

Data assimilation practitioners should also be aware of systematic observation biases. As shown in the mayfly example, different results were obtained when temperature data were perturbed. The potential for uncorrected biases in certain types of observations (e.g. eddy covariance data, see Richardson and Hollinger (2005)) will always affect the results. Data assimilation practitioners should be aware of these biases and interpret results accordingly. Collaboration with the researchers who collected the data prior is essential.

Dynamic models need an initial condition in order to run the model. Models may exhibit sensitive dependence on initial conditions, or that quantitatively different results are obtained even when the initial condition is only slightly varied (see Lorenz (1963) for a classical example of sensitive dependence on initial conditions). Because most ecological models are non-linear in structure, they are more susceptible to this sensitive dependence on initial conditions. The MCMC method may converge to a local, rather than global, optimum when the initial condition is an estimated parameter. We recommend careful consideration of model initialization before applying the data assimilation technique.

Rule 3: **Know your prior**

Fixing some parameters with ‘known’ values is critical for complex models with many parameters. When optimizing parameters often multiple parameter value combinations may equally explain the data. Robust priors prevent the production of “unreasonable” parameter sets. However, priors with narrow bounds may prevent you from examining possible deviations from literature. Too many wide, non-informative priors should be avoided where possible. The best bet is to use priors conditioned on literature and field data, and then inflate those priors by an arbitrary amount to allow for solutions outside the range of observed variability but within the range of reality.

Rule 4: Respect uncertainty

Respect the uncertainty on both the data and the model. Any data collected should have an associated uncertainty. Bayesian methods already specify a prior uncertainty on estimated parameters. For a linear model, the uncertainty on estimated parameters will only be reduced by the inclusion of extra data (Tarantola, 2005). The level of uncertainty on both the data and the model will influence the results. If raw data are aggregated together (e.g. aggregation of half-hourly *NEE* data to twice daily), then the overall uncertainty may be reduced. For our first example, we assumed that the errors were Gaussian and proportional to the variance of the model-data residual. Determining and evaluating different assumptions on error distributions may also be necessary to understand the impact of uncertainty on results (Braswell et al., 2005; Richardson and Hollinger, 2005). As mentioned in Rule 2, systematic biases in data should also be accounted for when dealing with data uncertainty.

Model uncertainty should be considered as well; we recognize assigning uncertainty to a model may be conceptually challenging. For a linear model, the uncertainties for data and model can be additive under certain assumptions (Tarantola, 2005). However different model structures can be investigated (Example 1) and the results can be investigated to qualitatively determine the model uncertainty. Information criteria such as the *BIC* provide a way to test the most parsimonious model for the data. We caution that the *BIC* is a relative measure and cannot be compared between two different model sets (e.g. between Example 1 and Example 2).

Rule 5: Trust but verify

Data assimilation is analogous to running a controlled experiment on the computer. Data assimilation techniques are well established, however a small error in coding of the model or the data may lead to incorrect conclusions from the data assimilation. From our experience we have found three ways that create greater confidence in our results. First, because data assimilation can be computationally intensive, if you are preparing data to be assimilated and want to test the routine, use simpler test data before doing a full-scale data assimilation. Second, verify results by challenging the data assimilation technique to replicate known results. Generate a dummy data set from known parameters. Randomly assign an error to these data and then assimilate the data. If your data assimilation technique is working correctly then you should retrieve reasonable estimate on the initial parameters (Braswell et al., 2005). Third, verify results through cross-validation. Zobitz et al. (2008) estimated parameters on a subset (termed “validation data”) of an eight-year record of *NEE* data, and then compared model predictions on a subset of “corroboration data” that weren’t used in the parameter estimation routine.

Rule 6: Have fun! ... eventually

Data assimilation and Bayesian hypothesis testing are by their very nature tedious and not as “clean” as simple model fitting. Parameter and prediction results from these models require greater analysis (e.g., assessing posteriors relative to priors) and caution when drawing inferences or falsifying hypotheses (e.g., role of parameter correlations). However, in the long run, we hope that one finds Bayesian data assimilation a richer way to test ecological hypotheses that are consistent with theory (model) and experiment (observation) and potentially lead to improved ecological forecasting.

II. Glossary of terms used in Data Assimilation

Cost Function: A function used to compare measurements to model estimates.

Data Assimilation: Data assimilation is a general term for methods that systematically combine measurements with a model with the goal of improving model performance

Driver Variable: Quantities that cause responses in model components, e.g. precipitation, temperature, net radiation, day length.

Likelihood Function: A type of cost function that assumes measurement error follows an assumed probability density function.

Monte Carlo Markov Chain: A numerical technique where the current state or quantity depends only on the previous state or quantity.

Optimization: The generic name for any data assimilation procedure which improves the fit between a model and observations by modifying either model parameters or states.

Parameter: A quantity which controls the response of some portion of the model to another quantity, often a driver variable, e.g. the temperature sensitivity of respiration, the response of growth to temperature, the degree to which stomatal aperture closes in response to dryness.

Parameter or State Space: The range of allowed values for a given parameter or state variable.

Posterior distribution: The distribution of a parameter or state calculated after applying a data assimilation technique.

Prior distribution: The allowable range, kurtosis and skewness of a parameter or state inferred from ecological knowledge.

Probability Density Function: A function that describes the probability of a parameter or state variable within a given range of values, e.g. the probability density function of a mayfly being a certain length.

State Variable: A variable of a dynamic model that contains the current value of a ‘state’, e.g. the number of individuals in a population, the amount of leaf area per unit ground area.

III. Supplementary Figures and Tables

Table 1: Overview of studies incorporating data assimilation, sorted by method utilized.

Reference	Method utilized	Description
Review		
Canadell et al. (2004)		Review of future directions of carbon cycle research and role of data assimilation.
Doney and Ducklow (2006)		Review of model-data assimilation in ocean flux studies.
Friend et al. (2007)		Synthesis of modeling results derived from FLUXNET data.
Hodges (2010)		Cautionary note on selection of appropriate models and prior information.
Hurtt et al. (2010)		Review of applications of dynamic spatial ecosystem models and data.
Mathieu and O'Neill (2008)		Broad overview of data assimilation
Olden et al. (2008)		Review of machine learning methods for ecology research.
Rayner (2010)		Review current state of data assimilation of carbon cycle
Raupach et al. (2005)		Review paper of model-data assimilation.
Vargas et al. (2010)		Synthesis of data assimilation for soil respiration measurements.
Wang et al. (2009c)		Review of model-data assimilation for carbon fluxes.
Williams et al. (2009)		Review of data assimilation with special focus to FLUXNET data.
Atmospheric		
Merinik et al. (2008)	4D-Var	Estimation of regional methane emissions
Vukicevic & Bao (1998)	4D-Var	Numerical study of errors and assumptions of 4D-Var data assimilation method
Chevallier et al. (2006)	Bayesian parameter estimation	Determine appropriate probability distribution for coupled ecosystem-atmosphere models.
Franks et al. (1999)	Bayesian parameter estimation	Determination of sensible and latent heat fluxes
Matross et al. (2006)	Bayesian parameter estimation	Regional surface flux estimation from satellite and atmospheric data
Ogee et al. (2004)	Bayesian parameter estimation	Estimation of ecosystem-scale photosynthesis and respiration
Rayner et al. (2005)	Bayesian parameter estimation	Assimilation of remote-sensing and atmospheric CO2 data into a model
Scholze et al. (2007)	Bayesian parameter estimation	Uncertainty analysis of model-data assimilation of atmospheric CO2 data.

Zobitz et al. (2007)	Bayesian parameter estimation	Determine ecosystem carbon fluxes from atmospheric CO2 data.
Arellano et al. (2007)	EnKF	Determine spatial distribution of global atmospheric CO
Feng et al. (2009)	EnKF	Estimate regional CO2 fluxes from dry-air CO2 mole fractions.
Peters et al. (2005)	EnKF	Estimate surface CO2 fluxes from atmospheric CO2 measurements.
Peters et al. (2007)	EnKF	High spatial resolution of surface CO2 fluxes from atmospheric measurements.
Zupanski et al (2007a)	Maximum Likelihood Ensemble Filter	Estimate spatial variation in gross ecosystem production and whole-ecosystem respiration components
Stauch et al. (2008)	Monte Carlo	Comparison of measured NEE to stochastic generated NEE at multiple sites
Kawa et al. (2004)	NASA DAS (*)	Model global CO2 transport
Kaminsky et al. (2002)	Objective function	Model constraint through combining atmospheric and local measurements to constrain
Kaminsky et al. (2007)	Objective function	Feasibility study to determine process parameters from climatological data.
Wang et al. (2001)	Objective function	Model-data assimilation of carbon and water fluxes at six different sites.
Baker et al. (2006)	Variational data assimilation	Estimate surface CO2 fluxes on 4^o by 5^o spatial resolution.
Vukicevic et al. (2001)	Variational data assimilation	Determine influence of temperature on carbon exchange from global datasets.
Zupanski et al (2007b)	Various	Apply information theory to determine for model selection for weather forecasting.
Jimenez-Guerrero et al. (2008)		High resolution air quality modeling.
Young (2006)		Application of data-based mechanistic approaches to determine global average CO2 levels.
Zeng et al. (2005)		Determine global carbon fluxes and their influence on interannual CO2 variability
<u>Ecosystem and Ecological</u>		
Renzullo et al. (2008)	Bayesian parameter estimation	Model-data assimilation of soil moisture and temperature data into a biophysical model.
Tang & Zhuang (2008)	Bayesian parameter estimation	Uncertainty analysis of model parameters to predict regional carbon stocks.
Chen et al. (2008)	EnKF	Dual estimation of ecosystem state variables and model parameter values.
Mo et al. (2008)	EnKF	Model assimilation with seasonal and interannually changing parameters.

Quaife et al. (2008)	EnKF	Assimilating canopy reflectance data into an ecosystem model.
Rastetter et al. (2010)	EnKF	Model-data assimilation of carbon flux data for arctic ecosystems.
Stockli et al. (2008)	EnKF	Data assimilation of satellite data to determine ecosystem phenology.
Williams et al. (2005)	EnKF	Model-data assimilation to determine forest carbon fluxes and stocks.
Lokupitiya et al. (2008)	Maximum Likelihood Ensemble Filter	Global determination of CO2 fluxes
Braswell et al. (2005)	MCMC	Estimate gross ecosystem production and whole-ecosystem respiration
Higgins et al. (2010)	MCMC	Model-data assimilation of African Savannah dynamics.
Kattge et al. (2009)	MCMC	Determine photosynthetic model parameters on a global scale.
Prihodoko et al. (2008)	MCMC	Model-data assimilation and analysis of carbon fluxes from a tall-tower site.
Ricciuto et al. (2008a)	MCMC	Forecasting of future carbon stocks from past measurements.
Ricciuto et al. (2008b)	MCMC	Model-data assimilation of carbon flux data into a biophysical model.
Sacks et al. (2006)	MCMC	Model-data assimilation of carbon flux data and coupling of carbon cycle to climate.
Wang (2009)	MCMC	Estimation of time-varying parameters for ecological population models.
Wu et al. (2009)	MCMC	Model-data assimilation of a terrestrial ecosystem model.
Zobitz et al. (2008)	MCMC	Model-data assimilation of ecosystem flux data to constrain soil respiration models.
Mitchell et al. (2011)	GLUE	Multi-site model-data assimilation investigating the sensitivity of carbon uptake to soil moisture.
Carvalhais et al (2008)	Objective function	Investigate model NEP variability under steady state assumptions.
Liu et al. (2008)	Objective function	Seasonal determination of LAI from satellite reflectance data.
Pinty et al. (2007)	Objective function	Model-data assimilation of satellite leaf area indices.
Turner et al. (2006)	Objective Function	Regional estimation of NEP from satellite data.
Turner et al. (2009)	Objective function	Integration of satellite data and flux tower data to constrain surface carbon fluxes.
Vohland & Jarmer (2007)	Objective function	Model-data assimilation for a radiative transfer model.
Jarlan et al. (2008)	Variational data assimilation	Assimilation of satellite LAI data to determine latent heat and carbon fluxes.
Wang et al. (2007)		Two-tier simulation of carbon and nitrogen stocks for a biophysical model.
Zhang et al. (2008)		Modeling of regional carbon fluxes in Northeastern China.
Wang et al. (2009b)		Determination of biophysical model parameters using a hierarchical approach.

Hydrological		
Durand et al. (2009)	Kalman Filter	Validation of snowpack depth from radiance measurements for land-surface models.
Moore et al. (2008)	MCMC	Estimate transpiration and carbon uptake fluxes for a subalpine forest.
See & Abrahart (2001)	Neural networks	Data assimilation to forecast river levels.
Abrahart & See (2002)	Various	Comparison of data assimilation techniques to forecast river flow.
Jacobs et al. (2008)		Comparison of three models to constrain regional soil moisture observations.
Milesi et al. (2005)		Modeled the contribution of turf grasses for carbon sequestration.
Teuling et al. (2009)		Parameter sensitivity analysis of soil moisture models.
Wood et al. (2002)		Hydrological forecasting using climate forecasts coupled to a model.
Methods		
Clason & Hepperger (2009)	Variational data assimilation	Numerical study of variational data assimilation method.
Han & Li (2008)	Various	Mathematical implementation of Bayesian filters
Bailey et al. (2010)		Investigate parameter redundancy for a multistate mark-recapture model.
Other		
Matear and Holloway (1995)	Adjoint method	Modeling ocean phosphorous
Tjiputra et al. (2007)	Adjoint method	Model-data assimilation and parameter sensitivity analysis of ocean ecosystem model.
Burgers et al.(1998)	EnKF	Numerical analysis and implementation of EnKF
Ridgwell et al. (2007)	EnKF	Model-data assimilation of ocean biogeochemical model.
Spitz el al. (2001)	Objective function	Model-data assimilation of an ocean ecosystem model.
Schneider et al. (2008)		Comparison of model-predicted marine productivity to satellite data.
Wang et al. (2009a)		Model sea-ice circulation patterns in the Northern Hemisphere.

Supplementary Information for Example 1: Determining the temperature sensitivity of ecosystem carbon respiration

Table 2: Results from data assimilation of ecosystem respiration. The column "Prior values" shows the starting value and the range of allowed values used in the optimization. The column "Model" represents the combination of Equations Error! Reference source not found.-Error! Reference source not found. used to generate for a particular model. The column "Posterior Values" shows the results of the MCMC algorithm, with the parameter value that maximizes the likelihood, followed by the range of accepted values from the data assimilation algorithm.

Parameter	Prior values	Model	Posterior values
B_R	1 [0.25, 4]	1	2.77 [2.54, 2.96]
		2	3.09 [1.58, 3.99]
		3	2.65 [2.44, 2.86]
		4	0.90 [0.86, 1.64]
Q_{10}	2 [1.25, 3.5]	1	1.89 [1.64, 2.15]
		2	1.87 [1.62, 2.21]
		3	2.85 [2.42, 3.26]
		4	3.25 [2.56, 3.50]
R	50 [10,100]	1	
		2	57.7 [48.0, 63.2]
		3	
		4	48.1 [47.5, 51.8]
A_{max}	20 [5,50]	1	15.4 [14.3, 16.6]
		2	15.2 [14.2, 16.1]
		3	16.7 [15.7, 17.6]
		4	16.5 [15.4, 17.4]
LUE	0.001 [0.001, 0.1]	1	0.025 [0.023, 0.028]
		2	0.025 [0.023, 0.028]
		3	0.032 [0.026, 0.037]
		4	0.032 [0.027, 0.036]
T_{max}	25 [-10, 30]	1	
		2	
		3	14.0 [9.6, 16.1]
		4	13.5 [9.5, 15.6]
T_{min}	0 [-20, 20]	1	
		2	
		3	-4.3 [-6.0, -3.0]
		4	-3.7 [-5.5, -2.5]

Supplementary Information for Example 2: Predicting mayfly emergence with an age structured model

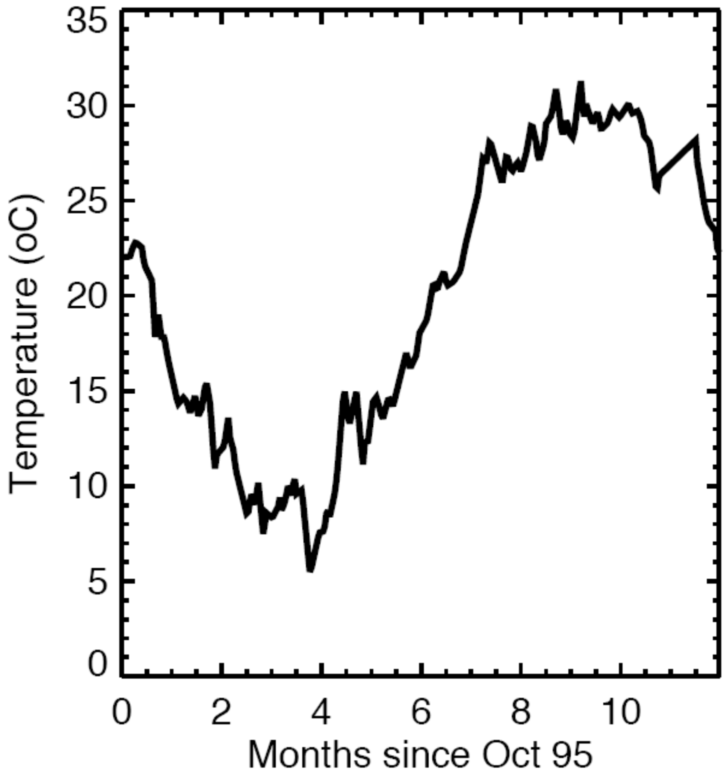

Figure 1: Mean daily temperature in the Lower Mobile River, Alabama October 1995 to September 1996 digitized from data presented in Chadwick and Feminella (2001).

Table 3: Results from data assimilation of the mayfly population model. The column "Prior values" shows the starting value and the range of allowed values used in the optimization. The column "Posterior Values" shows the results of the MCMC algorithm, with the parameter value that maximizes the likelihood, followed by the range of accepted values from the data assimilation algorithm. The column "Mortality" denotes the four parameter model where δ is an estimated parameter ranging from 0 to 1, and "No mortality" is when δ is fixed at 0.

Parameter	**Prior value**	**Posterior values**	
		Mortality	No mortality
α	0.4 [0-1]	0.28 [0.27-0.31]	0.97 [0.10-1.00]
δ	0.2 [0-1]	0.27 [0.27-0.27]	
T_1	10 [-10-20]	9.8 [9.3-10.2]	15.0 [7.6-15.2]
T_2	25 [10-40]	12.8 [11.9-13.4]	15.9 [14.4-18.0]

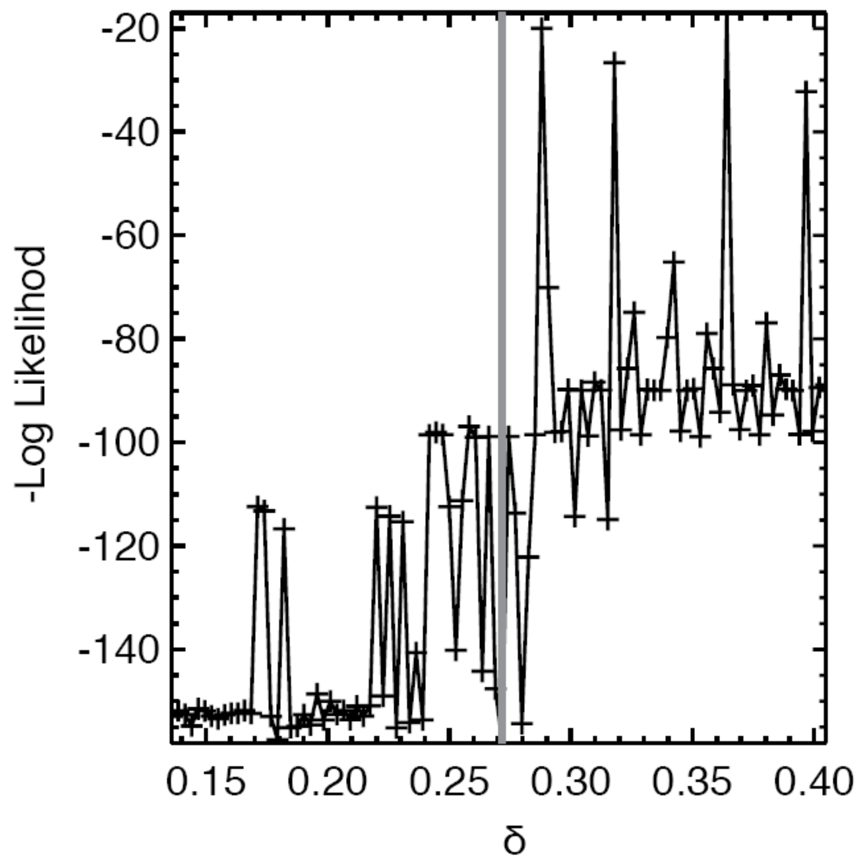

Figure 2. The negative log likelihood of the mortality model exhibits strong variations when the mortality parameter is perturbed 50% of its accepted value, denoted with the vertical line in the figure.

A lesson from this mayfly model is that mortality and growth rates are relatively similar, highlighting an interesting nature of mayfly population dynamics. Using the optimized model parameters from our MCMC method as a baseline, we can explore the sensitivity of the model to parameter variations. This sensitivity analysis enhances a quantitative understanding of how parameters affect model results. If a parameter is not well-constrained by a model, the log-likelihood function will not indicate a dependency on parameter variation. Figure 2 shows variation in the log likelihood function when δ is varied +/- 50% from its optimized value of 0.27 while the other three parameters are fixed. Essentially what we see here is that the model exhibits a strong non-linear sensitivity to this parameter. From a modeler's perspective this implies that either (a) MCMC will require a larger number of iterations to sample the parameter space or (b) a tighter prior is needed to ensure that all of these options are adequately sampled.

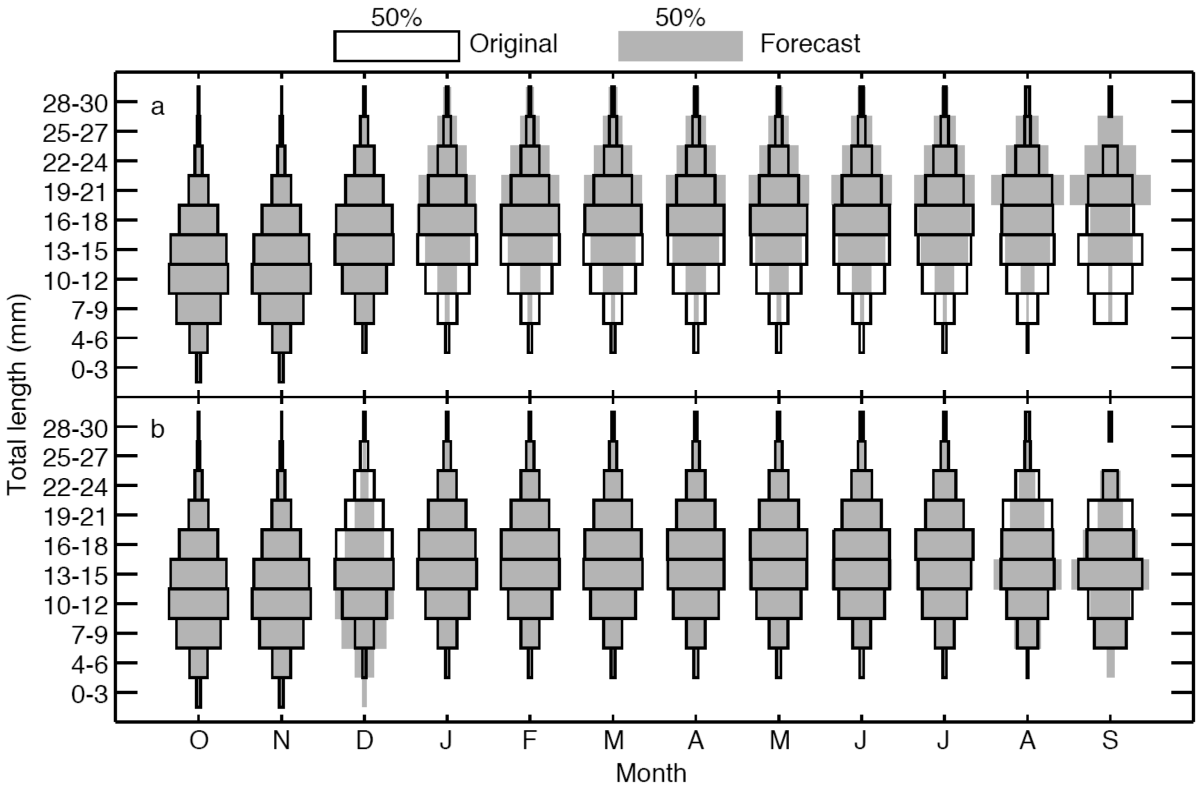

Figure 3. The mayfly mortality model parameter set propagated into the model with temperatures perturbed by a) +2^{o} C and b) +4^{o} C (gray bars) compared to the model with the original temperature observations (black bars).

In Figure 3 the best parameter set is run with temperature observations perturbed by +2 ^{o}C (Figure 3a) and +4 ^{o}C (Figure 3b). The model does show an enhancement of growth of large mayflies and a decrease in relative abundance of small mayflies with a small temperature change, but a small decline in large mayflies and no loss of small mayflies with a larger temperature change. One might conclude that a large temperature change may result in less mayfly emergence as the individuals spend more days above their temperature threshold. Given the simplistic nature of the model, it would be unwise to assume that these are accurate predictions.

IV. Supplementary Information References

Abrahart, R., and L. See (2002), Multi-model data fusion for river flow forecasting: an evaluation of six alternative methods based on two contrasting catchments, *Hydrology And Earth System Sciences*, *6*(4), 655-670.

Arellano, A., K. Raeder, J. Anderson, P. Hess, L. Emmons, D. Edwards, G. Pfister, T. Campos, and G. Sachse (2007), Evaluating model performance of an ensemble-based chemical data assimilation system during INTEX-B field mission, *Atmospheric Chemistry and Physics*, *7*(21), 5695-5710.

Bailey, L. L., S. J. Converse, and W. L. Kendall (2010), Bias, precision, and parameter redundancy in complex multistate models with unobservable states, *Ecology*, *91*(6), 1598-1604, doi:10.1890/09-1633.1.

Baker, D., S. Doney, and D. Schimel (2006), Variational data assimilation for atmospheric CO2, *Tellus B*, *58*(5), 359-365, doi:10.1111/j.1600-0889.2006.00218.x.

Braswell, B. H., W. J. Sacks, E. Linder, and D. S. Schimel (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*(2), 335-355, doi:10.1111/j.1365-2486.2005.00897.x.

Burgers, G., P. Jan van Leeuwen, and G. Evensen (1998), Analysis Scheme in the Ensemble Kalman Filter, *Monthly Weather Review*, *126*(6), 1719-1724.

Canadell, J., P. Ciais, P. Cox, and M. Heimann (2004), Quantifying, Understanding and Managing the Carbon Cycle in the Next Decades, *Climatic Change*, *67*(2), 147-160, doi:10.1007/s10584-004-3765-y.

Carvalhais, N. et al. (2008), Implications of the carbon cycle steady state assumption for biogeochemical modeling performance and inverse parameter retrieval, *Global Biogeochemical Cycles*, *22*(2), doi:10.1029/2007GB003033.

Chadwick, M. a, and J. W. Feminella (2001), Influence of salinity and temperature on the growth and production of a freshwater mayfly in the Lower Mobile River, Alabama, *Limnology and Oceanography*, *46*(3), 532-542.

Chen, M., S. Liu, L. Tieszen, and D. Hollinger (2008), An improved state-parameter analysis of ecosystem models using data assimilation, *Ecological Modelling*, *219*(3-4), 317-326, doi:10.1016/j.ecolmodel.2008.07.013.

Chevallier, F., N. Viovy, M. Reichstein, and P. Ciais (2006), On the assignment of prior errors in Bayesian inversions of CO2 surface fluxes, *Geophysical Research Letters*, *33*(13), doi:10.1029/2006GL026496.

Clason, C., and P. Hepperger (2009), A Forward Approach to Numerical Data Assimilation, *SIAM J. Sci. Comput.*, *31*(4), 3090-3115, doi:10.1137/090746240.

Doney, S., and H. Ducklow (2006), A decade of synthesis and modeling in the US Joint Global Ocean Flux Study, *Deep Sea Research Part II: Topical Studies in Oceanography*, *53*(5-7), 451-458, doi:10.1016/j.dsr2.2006.01.019.

Durand, M., E. Kim, and S. Margulis (2009), Radiance assimilation shows promise for snowpack characterization, *Geophysical Research Letters*, *36*, doi:10.1029/2008GL035214.

Feng, L., P. Palmer, H. Bosch, and S. Dance (2009), Estimating surface CO2 fluxes from space-borne CO2 dry air mole fraction observations using an ensemble Kalman Filter, *Atmospheric Chemistry and Physics*, *9*(8), 2619-2633, doi:10.5194/acp-9-2619-2009.

Franks, S. W., K. J. Beven, and J. H. C. Gash (1999), Multi-objective conditioning of a simple SVAT model, *Hydrol. Earth Syst. Sci.*, *3*(4), 477-488.

Friend, A. et al. (2007), FLUXNET and modelling the global carbon cycle, *Global Change Biology*, *13*(3), 610-633, doi:10.1111/j.1365-2486.2006.01223.x.

Han, X., and X. Li (2008), An evaluation of the nonlinear/non-Gaussian filters for the sequential data assimilation, *Remote Sensing of Environment*, *112*(4), 1434–1449.

Higgins, S. I., S. Scheiter, and M. Sankaran (2010), The stability of African savannas: insights from the indirect estimation of the parameters of a dynamic model, *Ecology*, *91*(6), 1682-1692, doi:10.1890/08-1368.1.

Hodges, J. S. (2010), Are exercises like this a good use of anybody's time?, *Ecology*, *91*(12), 3496-3500.

Hurtt, G. C., J. Fisk, R. Q. Thomas, R. Dubayah, P. R. Moorcroft, and H. H. Shugart (2010), Linking models and data on vegetation structure, *J. Geophys. Res.*, *115*, 11 PP., doi:201010.1029/2009JG000937.

Jacobs, C. M. J. et al. (2008), Evaluation of European Land Data Assimilation System (ELDAS) products using in situ observations, *Tellus A*, *60*(5), 1023-1037, doi:10.1111/j.1600-0870.2008.00351.x.

Jarlan, L., G. Balsamo, S. Lafont, A. Beljaars, J. Calvet, and E. Mougin (2008), Analysis of leaf area index in the ECMWF land surface model and impact on latent heat and carbon fluxes: Application to West Africa, *Journal of Geophysical Research - Atmospheres*, *113*, doi:10.1029/2007JD009370.

Jimenez-Guerrero, P., O. Jorba, J. Baidasanoa, and S. Gasso (2008), The use of a modelling system as a tool for air quality management: Annual high-resolution simulations and evaluation, *Science of the Total Environment*, *390*(2-3), 323-340, doi:10.1016/j.scitotenv.2007.10.025.

Kaminski, T., W. Knorr, P. Rayner, and M. Heimann (2002), Assimilating atmospheric data into a terrestrial biosphere model: A case study of the seasonal cycle, *Global Biogeochemical Cycles*, *16*(4), doi:10.1029/2001GB001463.

Kaminski, T., S. Blessing, R. Giering, M. Scholze, and M. Vossbeck (2007), Testing the use of adjoints for parameter estimation in a simple GCM on climate time-scales, *Meteorologische Zeitschrift*, *16*, 643-652, doi:10.1127/0941-2948/2007/0259.

Kattge, J., W. Knorr, T. Raddatz, and C. Wirth (2009), Quantifying photosynthetic capacity and its relationship to leaf nitrogen content for global-scale terrestrial biosphere models, *Global Change Biology*, *15*(4), 976-991, doi:10.1111/j.1365-2486.2008.01744.x.

Kawa, S., D. Erickson, S. Pawson, and Z. Zhu (2004), Global CO2 transport simulations using meteorological data from the NASA data assimilation system, *Journal of Geophysical Research - Atmospheres*, *109*(D18), doi:10.1029/2004JD004554.

Liu, Q., L. Gu, R. Dickinson, Y. Tian, L. Zhou, and W. Post (2008), Assimilation of satellite reflectance data into a dynamical leaf model to infer seasonally varying leaf areas for climate and carbon models, *Journal of Geophysical Research - Atmospheres*, *113*(D19), doi:10.1029/2007JD009645.

Lokupitiya, R., D. Zupanski, A. Denning, S. Kawa, K. Gurney, and M. Zupanski (2008), Estimation of global CO2 fluxes at regional scale using the maximum likelihood ensemble filter, *Journal of Geophysical Research - Atmospheres*, *113*(D20), doi:10.1029/2007JD009679.

Lorenz, E. (1963), Deterministic Nonperiodic Flow, *Journal of the Atmospheric Sciences*, *20*, 130-141.

Matear, R. J., and G. Holloway (1995), Modeling The Inorganic Phosphorous Cycle Of The North Pacific Using An Adjoint Data Assimilation Model to Assess the Role of Dissolved Organic Phosphorus, *Global Biogeochemical Cycles*, *9*(1), 101-119, doi:10.1029/94GB03104.

Mathieu, P., and A. O'Neill (2008), Data assimilation: From photon counts to Earth System forecasts, *Remote Sensing of Environment*, *112*(4), 1258-1267, doi:10.1016/j.rse.2007.02.040.

Matross, D. M. et al. (2006), Estimating regional carbon exchange in New England and Quebec by combining atmospheric, ground-based and satellite data, *Tellus B*, *58*(5), 344-358, doi:10.1111/j.1600-0889.2006.00206.x.

Meirink, J., P. Bergamaschi, and M. Krol (2008), Four-dimensional variational data assimilation for inverse modelling of atmospheric methane emissions: method and comparison with synthesis inversion, *Atmospheric Chemistry and Physics*, *8*(21), 6341-6353.

Milesi, C., S. Running, C. Elvidge, J. Dietz, B. Tuttle, and R. Nemani (2005), Mapping and Modeling the Biogeochemical Cycling of Turf Grasses in the United States, *Environmental Management*, *36*(3), 426-438, doi:10.1007/s00267-004-0316-2.

Mitchell, S., K. Beven, J. Freer, and B. Law (2011), Processes influencing model-data mismatch in drought-stressed, fire-disturbed eddy flux sites, *J. Geophys. Res.*, *116*, 15 PP., doi:201110.1029/2009JG001146.

Mo, X., J. Chen, W. Ju, and T. Black (2008), Optimization of ecosystem model parameters through assimilating eddy covariance flux data with an ensemble Kalman filter, *Ecological Modelling*, *217*(1-2), 157-173, doi:10.1016/j.ecolmodel.2008.06.021.

Moore, D. J. P., J. Hu, W. J. Sacks, D. S. Schimel, and R. K. Monson (2008), Estimating transpiration and the sensitivity of carbon uptake to water availability in a subalpine forest using a simple ecosystem process model informed by measured net CO2 and H2O fluxes, *Agricultural and Forest Meteorology*, *148*(10), 1467-1477, doi:10.1016/j.agrformet.2008.04.013.

Ogee, J., P. Peylin, M. Cuntz, T. Bariac, Y. Brunet, P. Berbigier, P. Richard, and P. Ciais (2004), Partitioning net ecosystem carbon exchange into net assimilation and respiration with canopy-scale isotopic measurements: An error propagation analysis with (CO2)-C-13 and (COO)-O-18 data, *Global Biogeochemical Cycles*, *18*(2), doi:10.1029/2003GB002166.

Olden, J. D., J. J. Lawler, and N. L. Poff (2008), Machine Learning Methods Without Tears: A Primer For Ecologists, *Quarterly Review of Biology*, *83*(2), 171-193, doi:Article.

Peters, W. et al. (2007), An atmospheric perspective on North American carbon dioxide exchange: CarbonTracker, *Proceedings of the National Academy of Science*, *104*(48), 18925-18930, doi:10.1073/pnas.0708986104.

Peters, W., J. Miller, J. Whitaker, A. Denning, A. Hirsch, M. Krol, D. Zupanski, L. Bruhwiler, and P. Tans (2005), An ensemble data assimilation system to estimate CO2 surface fluxes from atmospheric trace gas observations, *Journal of Geophysical Research - Atmospheres*, *110*(D24), doi:10.1029/2005JD006157.

Pinty, B., T. Lavergne, M. Vossbeck, T. Kaminski, O. Aussedat, R. Giering, N. Gobron, M. Taberner, M. M. Verstraete, and J. L. Widlowski (2007), Retrieving surface parameters for climate models from Moderate Resolution Imaging Spectroradiometer (MODIS)-Multiangle Imaging Spectroradiometer (MISR) albedo products, *J. Geophys. Res*, *112*, doi:10.1029/2006JD008105.

Porter, A., and I. Rafols (2009), Is science becoming more interdisciplinary? Measuring and mapping six research fields over time, *Scientometrics*, *81*(3), 719-745, doi:10.1007/s11192-008-2197-2.

Prihodko, L., A. S. Denning, N. P. Hanan, I. Baker, and K. Davis (2008), Sensitivity, uncertainty and time dependence of parameters in a complex land surface model, *Agricultural and Forest Meteorology*, *148*(2), 268-287, doi:10.1016/j.agrformet.2007.08.006.

Quaife, T., P. Lewis, M. De Kauwe, M. Williams, B. E. Law, M. Disney, and P. Bowyer (2008), Assimilating canopy reflectance data into an ecosystem model with an ensemble Kalman filter, *Remote Sensing of Environment*, *112*(4), 1347–1364, doi:10.1016/j.rse.2007.05.020.

Rastetter, E. B. et al. (2010), Processing arctic eddy-flux data using a simple carbon-exchange model embedded in the ensemble Kalman filter, *Ecological Applications*, *20*(5), 1285-1301, doi:10.1890/09-0876.1.

Raupach, M. R., P. J. Rayner, D. J. Barrett, R. S. DeFries, M. Heimann, D. S. Ojima, S. Quegan, and C. C. Schmullius (2005), Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications, *Global Change Biology*, *11*(3), 378-397, doi:10.1111/j.1365-2486.2005.00917.x.

Rayner, P., M. Scholze, W. Knorr, T. Kaminski, R. Giering, and H. Widmann (2005), Two decades of terrestrial carbon fluxes from a carbon cycle data assimilation system (CCDAS), *Global Biogeochemical Cycles*, *19*(2), doi:10.1029/2004GB002254.

Rayner, P. J. (2010), The current state of carbon-cycle data assimilation, *Current Opinion in Environmental Sustainability*, *2*(4), 289-296, doi:10.1016/j.cosust.2010.05.005.

Renzullo, L. J., D. J. Barrett, A. S. Marks, M. J. Hill, J. P. Guerschman, Q. Mu, and S. W. Running (2008), Multi-sensor model-data fusion for estimation of hydrologic and energy flux parameters, *Remote Sensing of Environment*, *112*(4), 1306–1319.

Ricciuto, D. M., K. J. Davis, and K. Keller (2008a), A Bayesian calibration of a simple carbon cycle model: The role of observations in estimating and reducing uncertainty, *Global Biogeochem. Cycles*, *22*, doi:10.1029/2006GB002908.

Ricciuto, D. M., M. P. Butler, K. J. Davis, B. D. Cook, P. S. Bakwin, A. Andrews, and R. M. Teclaw (2008b), Causes of interannual variability in ecosystem-atmosphere CO2 exchange in a northern Wisconsin forest using a Bayesian model calibration, *Agricultural and Forest Meteorology*, *148*(2), 309-327, doi:10.1016/j.agrformet.2007.08.007.

Richardson, A. D., and D. Y. Hollinger (2005), Statistical modeling of ecosystem respiration using eddy covariance data: Maximum likelihood parameter estimation, and Monte Carlo simulation of model and parameter uncertainty, applied to three simple models, *Agricultural and Forest Meteorology*, *131*, 191-208.

Ridgwell, A., J. Hargreaves, N. Edwards, J. Annan, T. Lenton, R. Marsh, A. Yool, and A. Watson (2007), Marine geochemical data assimilation in an efficient Earth System Model of global biogeochemical cycling, *Biogeosciences*, *4*(1), 87-104, doi:10.5194/bg-4-87-2007.

Sacks, W. J., D. S. Schimel, R. K. Monson, and B. H. Braswell (2006), Model-data synthesis of diurnal and seasonal CO_2 fluxes at Niwot Ridge, Colorado, *Global Change Biology*, *12*, 240-259, doi:10.1111/j.1365-2486.2005.01059.x.

Schneider, B., L. Bopp, M. Gehlen, J. Segschneider, T. Frolicher, P. Cadule, P. Friedlingstein, S. Doney, M. Behrenfeld, and F. Joos (2008), Climate-induced interannual variability of marine primary and export production in three global coupled climate carbon cycle models, *Biogeosciences*, (2), 597-614, doi:10.5194/bg-5-597-2008.

Scholze, M., P. Rayner, W. Knorr, and R. Giering (2007), Propagating uncertainty through prognostic carbon cycle data assimilation system simulations, *Journal of Geophysical Research - Atmospheres*, *112*(D17), doi:10.1029/2007JD008642.

See, L., and R. Abrahart (2001), Multi-model data fusion for hydrological forecasting, *Computers & Geosciences*, *27*(8), 987-994, doi:10.1016/S0098-3004(00)00136-9.

Spitz, Y. H., J. R. Moisan, and M. R. Abbott (2001), Configuring an ecosystem model using data from the Bermuda Atlantic Time Series (BATS), *Deep Sea Research Part II: Topical Studies in Oceanography*, *48*(8-9), 1733-1768, doi:10.1016/S0967-0645(00)00159-4.

Stauch, V. J., A. J. Jarvis, and K. Schulz (2008), Estimation of net carbon exchange using eddy covariance CO2 flux observations and a stochastic model, *J. Geophys. Res.*, *113*, doi:200810.1029/2007JD008603.

Stockli, R., T. Rutishauser, D. Dragoni, J. O'Keefe, P. Thornton, M. Jolly, L. Lu, and A. Denning (2008), Remote sensing data assimilation for a prognostic phenology model, *Journal of Geophysical Research - Biogeosciences*, *113*(G4), doi:10.1029/2008JG000781.

Tang, J., and Q. Zhuang (2008), Equifinality in parameterization of process-based biogeochemistry models: A significant uncertainty source to the estimation of regional carbon dynamics, *J. Geophys. Res.*, *113*, 13 PP., doi:200810.1029/2008JG000757.

Tarantola, A. (2005), *Inverse Problem Theory and Model Parameter Estimation*, SIAM Books, Philadelphia, PA.

Teuling, A., R. Uijlenhoet, B. van den Hurk, and S. Seneviratne (2009), Parameter Sensitivity in LSMs: An Analysis Using Stochastic Soil Moisture Models and ELDAS Soil Parameters, *Journal of Hydrometeorology*, *10*(3), 751-765, doi:10.1175/2008JHM1033.1.

Tjiputra, J., D. Polzin, and A. Winguth (2007), Assimilation of seasonal chlorophyll and nutrient data into an adjoint three-dimensional ocean carbon cycle model: Sensitivity analysis and ecosystem parameter optimization, *Global Biogeochemical Cycles*, *21*(1), doi:10.1029/2006GB002745.

Turner, D. P. et al. (2006), Evaluation of MODIS NPP and GPP products across multiple biomes, *Remote Sensing of Environment*, *102*(3-4), 282-292, doi:10.1016/j.rse.2006.02.017.

Turner, D. P., W. D. Ritts, S. Wharton, C. Thomas, R. Monson, T. A. Black, and M. Falk (2009), Assessing FPAR source and parameter optimization scheme in application of a diagnostic carbon flux model, *Remote Sensing of Environment*, *113*(7), 1529-1539, doi:10.1016/j.rse.2009.03.003.

Vargas, R., M. S. Carbone, M. Reichstein, and D. D. Baldocchi (2010), Frontiers and challenges in soil respiration research: from measurements to model-data integration, *Biogeochemistry*, *102*, 1-13, doi:10.1007/s10533-010-9462-1.

Vohland, M., and T. Jarmer (2007), Estimating structural and biochemical parameters for grassland from spectroradiometer data by radiative transfer modelling (PROSPECT plus SAIL), *International Journal of Remote Sensing*, *29*(1), 191-209, doi:10.1080/01431160701268947.

Vukicevic, T., B. Braswell, and D. Schimel (2001), A diagnostic study of temperature controls on global terrestrial carbon exchange, *Tellus Series B*, *53*(2), 150-170, doi:10.1034/j.1600-0889.2001.d01-13.x.

Vukićević, T., and J.-W. Bao (1998), The Effect of Linearization Errors on 4DVAR Data Assimilation, *Monthly Weather Review*, *126*(6), 1695-1706, doi:10.1175/1520-0493(1998)126<1695:TEOLEO>2.0.CO;2.

Wang, G. (2009), Signal extraction from long-term ecological data using Bayesian and non-Bayesian state-space models, *Ecological Informatics*, *4*(2), 69-75, doi:10.1016/j.ecoinf.2009.01.005.

Wang, J., H. Hu, K. Mizobata, and S. Saitoh (2009a), Seasonal variations of sea ice and ocean circulation in the Bering Sea: A model-data fusion study, *Journal of Geophysical Research - Oceans*, *114*, doi:10.1029/2008JC004727.

Wang, W., K. Ichii, H. Hashimoto, A. Michaelis, P. Thornton, B. Law, and R. Nemani (2009b), A hierarchical analysis of terrestrial ecosystem model Biome-BGC: Equilibrium analysis and model calibration, *Ecological Modelling*, *220*(17), 2009-2023, doi:10.1016/j.ecolmodel.2009.04.051.

Wang, Y.-P., B. Z. Houlton, and C. B. Field (2007), A model of biogeochemical cycles of carbon, nitrogen, and phosphorus including symbiotic nitrogen fixation and phosphatase production, *Global Biogeochem. Cycles*, *21*(1), doi:10.1029/2006GB002797.

Wang, Y.-P., R. Leuning, H. A. Cleugh, and P. A. Coppin (2001), Parameter estimation in surface exchange models using nonlinear inversion: how many parameters can we estimate and which measurements are most useful?, *Global Change Biology*, *7*(5), 495-510, doi:10.1046/j.1365-2486.2001.00434.x.

Wang, Y.-P., C. M. Trudinger, and I. G. Enting (2009c), A review of applications of model-data fusion to studies of terrestrial carbon fluxes at different scales, *Agricultural and Forest Meteorology*, *149*(11), 1829-1842, doi:10.1016/j.agrformet.2009.07.009.

Williams, M. et al. (2009), Improving land surface models with FLUXNET data, *Biogeosciences*, *6*(7), 1341-1359, doi:10.5194/bg-6-1341-2009.

Williams, M., P. A. Schwarz, B. E. Law, J. Irvine, and M. R. Kurpius (2005), An improved analysis of forest carbon dynamics using data assimilation, *Global Change Biology*, *11*(1), 89-105, doi:10.1111/j.1365-2486.2004.00891.x.

Wood, A. W. (2002), Long-range experimental hydrologic forecasting for the eastern United States, *J. Geophys. Res.*, *107*(D20), doi:10.1029/2001JD000659.

Wu, X., Y. Luo, E. Weng, L. White, Y. Ma, and X. Zhou (2009), Conditional inversion to estimate parameters from eddy-flux observations, *J Plant Ecol*, *2*(2), 55-68, doi:10.1093/jpe/rtp005.

Young, P. (2006), The data-based mechanistic approach to the modelling, forecasting and control of environmental systems, *Annual Reviews in Control*, *30*(2), 169-182, doi:10.1016/j.arcontrol.2006.05.002.

Zeng, N., A. Mariotti, and P. Wetzel (2005), Terrestrial mechanisms of interannual CO2 variability, *Global Biogeochemical Cycles*, *19*(1), doi:10.1029/2004GB0022763.

Zhang, Y., Q. Yu, J. Jiang, and Y. Tang (2008), Calibration of Terra/MODIS gross primary production over an irrigated cropland on the North China Plain and an alpine meadow on the Tibetan Plateau, *Global Change Biol*, *14*(4), 757-767, doi:10.1111/j.1365-2486.2008.01538.x.

Zimmerman, A. S. (2008), New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data, *Science Technology Human Values*, *33*(5), 631-652, doi:10.1177/0162243907306704.

Zobitz, J., S. Burns, J. Ogee, M. Reichstein, and R. Bowling (2007), Partitioning net ecosystem exchange of CO2: A comparison of a Bayesian/isotope approach to environmental regression methods, *Journal of Geophysical Research - Biogeosciences*, *112*(G3), doi:10.1029/2006JG000282.

Zobitz, J., D. Moore, W. Sacks, R. Monson, D. Bowling, and D. Schimel (2008), Integration of process-based soil respiration models with whole-ecosystem CO2 measurements, *Ecosystems*, *11*(2), 250-269, doi:10.1007/s10021-007-9120-1.

Zupanski, D., A. Denning, M. Uliasz, M. Zupanski, A. Schuh, P. Rayner, W. Peters, and K. Corbin (2007a), Carbon flux bias estimation employing maximum likelihood ensemble filter (MLEF), *Journal of Geophysical Research - Atmospheres*, *112*(D17), doi:10.1029/2006JD008371.

Zupanski, D., A. Y. Hou, S. Q. Zhang, M. Zupanski, C. D. Kummerow, and S. H. Cheung (2007b), Applications of information theory in ensemble data assimilation, *Quarterly Journal of the Royal Meteorological Society*, *133*(627), 1533-1545, doi:10.1002/qj.123.