Jul 1st, 12:00 AM

Upscaling Terrestrial Carbon Dynamics From Sites To Regions With Uncertainty Measures: The GEMS Experience

Shuguang Liu

Z. Li

Jinxun Liu

Thomas R. Loveland

M. Chen

See next page for additional authors

Presenter/Author Information
Shuguang Liu, Z. Li, Jinxun Liu, Thomas R. Loveland, M. Chen, and Larry L. Tieszen

Upscaling Terrestrial Carbon Dynamics From Sites To Regions With Uncertainty Measures: The GEMS Experience

S. Liu[a], Z. Li[a], J. Liu[a], T. Loveland[b], M. Chen, and L Tieszen[b]

[a] SAIC, U.S. Geological Survey (USGS) Center for Earth Resources Observations and Science (EROS), Sioux Falls, SD 57198

[b] USGS Center for EROS, Sioux Falls, SD 57198

Abstract: Upscaling the spatial and temporal changes of carbon stocks and fluxes from sites to regions is challenging owing to the spatial and temporal variances and covariance of driving variables and the uncertainties in both the model and the input data. Although various modeling approaches have been developed to facilitate the upscaling process, few deal with error transfer from model input to output, and error propagation in time and space. We develop the General Ensemble Biogeochemical Modeling System (GEMS) for upscaling carbon stocks and fluxes from sites to regions with measures of uncertainty. GEMS relies on site-scale biogeochemical models (e.g., the Erosion-Deposition-Carbon Model (EDCM) and CENTURY) to simulate the carbon dynamics at the site scale. The spatial deployment of the site-scale model in GEMS is based on the spatial and temporal joint frequency distribution of major driving variables (e.g., land cover and land use change, climate, soils, disturbances, and management). At the site scale, GEMS uses stochastic ensemble simulations to incorporate input uncertainty and to quantify uncertainty transfer from input to output. Using data assimilation techniques, GEMS simulations can be constrained by field and satellite observations or census data including estimates of net primary productivity (NPP) from the Moderate Resolution Imaging Spectroradiometer (MODIS), grain yield and cropping practices, and forest inventories.

The modeling philosophy embedded in GEMS makes it ideal for incorporating and assimilating information with various uncertainties at a range of spatio-temporal resolutions. The application of GEMS to quantify the contemporary terrestrial carbon dynamics in the United States is presented as an example of GEMS applications.

Keywords: Biogeochemical cycle; Modelling; Scaling; Ensemble simulation; Spatial and temporal variance

1. INTRODUCTION

The direct approach of incorporating variance and covariance of input variables in the simulation process can be expressed as:

$$E(p) = \int E[p(X)]f(X)dX \qquad (1)$$

where E is the operator of expectation, p is the nonlinear model, X is a vector of model variables, and f is the joint probability density function (pdf) of X. It is assumed in this equation that the spatial interactions (i.e., the interactions among plots) are insignificant and that the joint pdf of X is generated at the spatial scale of the model, namely, the specific spatial scale from which the model was developed and successfully tested.

Generally, it is impossible to analytically integrate equation (1) because of the complexity of models. In practice, equation (1) is often modified as a summation over a JFD table of the major input variables X. To do this, the major input variables can be partitioned in space into discrete, homogeneous regions using GIS techniques:

$$E(p) = \sum_{i=1}^{n} E[p(X_i)]F(X_i) \qquad (2)$$

where n is the number of strata or unique homogeneous regions as defined by the GIS overlay of the data layers of major input variables, and F is the frequency of cells or the total area of strata i as defined by the vector of Xi. We refer to the area defined by a unique combination of generalized map classes/values for all of the driving variables Xi as a simulation unit. Because

of the spatial heterogeneity of each data layer in Xi, the resulting simulation units in a region usually vary in size.

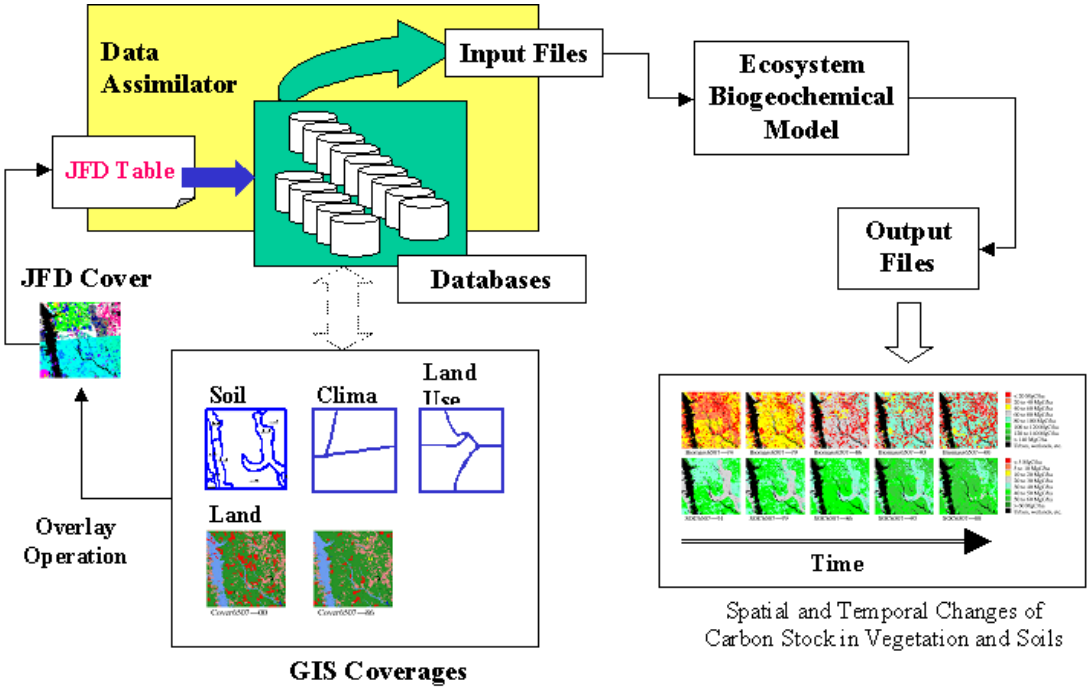

Figure 1. Diagram of the General Ensemble-based biogeochemical Model System (GEMS) and major steps in the spatially deployment of the encapsulated plot-level biogeochemical model over large areas.

Any difference between the model scale and the spatial resolution of the data is subject to introducing biases because of model nonlinearity. An ensemble approach can be used to assimilate the fine-scale heterogeneities in the databases in order to reduce the potential biases. The mean conditions (e.g., C stocks and fluxes) of simulation unit i can be estimated using multiple stochastic model simulations:

$$E[p(X_i)] = \frac{1}{m}\sum_{j=1}^{m} p(X_{ij}) \qquad (3)$$

where m is the number of stochastic fine-scale model runs for simulation unit i, and Xij is the vector of model input values at the fine scale generated using a Monte Carlo approach within the space defined by Xi. Therefore, input values for each stochastic model run are sampled from their corresponding potential value domains (i.e., Xi) that are usually described by their statistical information, such as moments and distribution types. The variance of the model simulations for simulation unit i can be quantified as follows:

$$\sigma^2[p(X_i)] = \sum_{j=1}^{m}[p(X_{ij}) - E[p(X_i)]]^2 \quad (4)$$

Other descriptive statistics (e.g., skewness) can also be calculated from the ensemble simulations. These moments not only characterize the spatial and temporal trends and patterns of simulated variables but also show their uncertainties in both space and time. Uncertainty transfer from input to output can be assessed in a spatial and temporal context.

2. GEMS AND DATA ASSIMILATION

2.1 GEMS Structure

GEMS is a modeling system that was developed for a better integration of well-established ecosystem biogeochemical models with various spatial databases for the simulations of the biogeochemical cycles over large areas (Figure 1) [*Liu et al.*, 2003b*; Liu et al.*, 2004]. It uses a Monte-Carlo-based ensemble approach to incorporate the variability (as measured by variances and covariance) of state and driving variables of the underlying biogeochemical models into simulations. Consequently, GEMS can not only simulate the spatial and temporal trends of C and N dynamics such as CO_2 exchange between the terrestrial biosphere and the atmosphere, but can also provide uncertainty estimates of the predicted variables in time and space.

GEMS consists of three major components: one or multiple encapsulated ecosystem biogeochemical models, an automated model parameterization

system (AMPS), and an input/output processor (IOP). Plot-scales models such as CENTURY (Parton et al., 1987) and EDCM (Liu et al., 2003) can be the encapsulated ecosystem biogeochemical model in GEMS (Liu et al., 2004a, 2004b, 2006; Tan et al., 2005).

Models developed for site-scale applications do not have apparent linkages with geo-referenced data. Most information in spatial databases is aggregated to the map-unit level, making the direct injection of such information into the modeling processes problematic and potentially biased [*Reiners et al.*, 2002]. Consequently, a AMPS is usually needed for incorporating field-scale spatial heterogeneities of state and driving variables into simulations. A AMPS generally consists of two major interdependent parts: (1) data search and retrieval algorithms, and (2) data processing mechanisms. The first part searches for and retrieves relevant information from various databases according to the keys provided by a joint frequency distribution (JFD) table [*Reiners et al.*, 2002*; Liu et al.*, 2004a]. The data processing mechanisms downscale the aggregated information at the map-unit level to the field scale, using a Monte Carlo approach. Once the data are assimilated, they are injected into the modeling processes through the IOP which updates the default input files with the assimilated data. Values of selected output variables are also written by the IOP to a set of output files after each model execution.

In the next section, we will describe the generation of land cover sequences and land use practices in GEMS. Automated stochastic parameterization of soil data can be found in Liu et al. (2004a).

2.2 Generation of Land Cover and Land Use Sequences in GEMS

In order for the encapsulated model EDCM to run for any given individual plot, a land cover and land use change file (i.e., the event schedule file in CENTURY) must be created first. This file specifies the time period for the simulation, the type and timing of any LCLUC change, as well as the types and timing of management practices such as cultivation and fertilization. In GEMS, the schedule file is created on-the-fly before each model simulation via the AMPS.

The following procedures are implemented in GEMS to assimilate the LCLUC data:

1. Land cover conversion, as indicated by different land cover classes at two consecutive points in time, is assumed to occur during the interval of the LCLUC observation. The exact year that the conversion happened was estimated randomly.
2. Even if no land cover change has been detected between two consecutive remote sensing (RS) observations, additional land use activities (e.g., clearcutting and selective cutting) might be prescribed to account for the activities that might have missed by Landsat observations. For example, because of the fast recovery of spectral reflectance after reforestation in part of the southern US, an interval of 8 years might be too long for detecting clearcutting activities that occurred during the early part of the time interval. How far back the remote sensing technique can go to detect clearcutting activities depends on the growth rates of forests. The faster they grow, the shorter the time period for which the remote sensing is effective. If the time period between two consecutive satellite observations is longer than the effective time frame, additional clearcutting events are scheduled by assuming that the detected annual clearcutting rate can be applied to the entire interval. In this study, selective cutting probability (selective cutting could not be picked up by Landsat) derived from the FIA database will be used to schedule selective cutting events on forestlands.
3. The general land cover class “agricultural land” used in LCLUC characterization is downscaled to crop species according to the crop composition information from the agricultural census. It is necessary to disaggregate the agricultural land into a combination of specific croplands for biogeochemical modeling because different crops with different biological characteristics and management practices likely have different impacts on carbon dynamics in vegetation and soils. Disaggregation of the agricultural land data will be done stochastically based on county level crop composition statistics. Cropping practices including shares of various crops and rotation probabilities have been derived from the National Resources Inventory (NRI) database, developed by USDA/Natural Resources Conservation Service (http://www.nrcs.usda. gov/technical/NRI/). The NRI database is a statistically based sample of land use and natural resources conditions and trends on U.S. nonfederal lands. The inventory, covering about 0.8 million sample points across the country, was done once every five years.
4. Another task in generating LULCC sequences is to fill the gaps between consecutive land cover maps. This will be accomplished with crop rotation probabilities on agricultural land calculated from the NRI databases. Management practices such as cultivation and fertilization will be incorporated in the generated LULCC sequences according to crop

or forest types assigned and the geographic region of the site.

5. Natural (e.g., fires, hurricanes, tornados) and anthropogenic (e.g., clearcutting) have been differentiated in the LCLUC mapping activities. There detailed information will be parameterized in the model accordingly because the biogeochemical consequences of these types of disturbances can be vastly different.

2.3 Data Assimilation

Inherent drawbacks exist on the measurement and modeling of ecosystem carbon dynamics at the plot to global scales. Measurements are usually patchy in space and discontinuous in time, while modeling is always built on some principles with assumptions and imperfectly-defined parameters (Chen *et al.* 2004, 2005, Ngodock *et al.* 2005). Advanced data assimilation techniques can overcome these drawbacks by combining a series of measurements with dynamic models based on statistics or optimal theory. Simply speaking, data assimilation is to find an optimal match between measurements and model estimates. A data assimilation process has three components: (1) dynamical models, (2) observation data, and (3) model–data synthesis or assimilation techniques (Raupach *et al.* 2005).

1. Models. Models can be expressed as:

$$X_{k+1} = f(X_k, U_k, P) + noise \qquad (5)$$

where $\boldsymbol{X_k}$ is a vector of state variables (such as carbon stocks, water and related entities); $\boldsymbol{f}$ is a vector of rates of change (net fluxes where components of $\boldsymbol{X_k}$ are stocks); $\boldsymbol{U_k}$ is a set of externally specified time-dependent forcing variables (such as meteorological variables and soil properties) and $\boldsymbol{P}$ is a set of time-independent model parameters (such as rate constants). The noise terms account for both imperfections in model formulation and stochastic variability in forcing variables ($\boldsymbol{U_k}$) or parameters ($\boldsymbol{P}$).

2. Data. The measured quantities ($\boldsymbol{Z}$) are related to the system state and external forcing variables by an observation model of the form

$$Z_k = H(X_k, U_k) + noise \qquad (6)$$

Where the operator H specifies the deterministic relationship between the measured quantities and the system state. The noise term accounts for both measurement error (instrumental and processing errors in the measurements) and representation error (errors in the model representation of $\boldsymbol{Z}$, introduced by shortcomings in the observation model $\boldsymbol{H}$).

3. Synthesis. The final component is a synthesis process including three basic choices:

a. Target variables. These are the model properties to be adjusted in the optimization process which may be model parameters ($\boldsymbol{P}$); forcing variables ($\boldsymbol{U_k}$), if there is substantial uncertainty on them; initial conditions on the state variables ($\boldsymbol{X_k}$); and time-dependent components of state variables ($\boldsymbol{X_k}$) (Aalto *et al.* 2004, Williams *et al.* 2005).

b. Cost function. That is the measure of distance between data and model and define as follows

$$J(Y) = \left(Z - H(Y)\right)^T R^{-1}\left(Z - H(Y)\right) + \left(Y - \hat{Y}\right)^T B^{-1}\left(Y - \hat{Y}\right) \qquad (7)$$

Where Y is the vector of target variables and a subset of all model variables ($\boldsymbol{X_k}$,U_k,P); $\hat{Y}$ denotes a priori estimate of Y; R and B are covariance matrices for Z and $\hat{Y}$.

c. Search strategy for finding the optimum values. Search strategies can be classified broadly into sequential and nonsequential methods. In the sequential methods the data arrive in time sequence and are incorporated into the model-data synthesis step by step, such as the well-known Kalman filter and ensemble Kalman filter (EnKF) (Evensen, 2003, Chen *et al.* 2005),. While in nonsequential methods the data are treated all at once, such as the least square fitting algorithms, the general, straightforward, and versatile Markov Chain Monte Carlo (MCMC) methods (Arsham 1998, Braswell *et al.* 2005).

3. APPLICATIONS

GEMS has been applied to simulate carbon dynamics in various regions including the United States, Africa, Asia, and Central America. Here we described its application to simulate carbon sources and sinks and its uncertainty in the Ridge and Valley ecoregion in the eastern United States. The Ridge and Valley (RV) is one of the Omernik level III ecoregions of North America. It is located in the eastern part of the United States covering parts of Pennsylvania, West Virginia, Virginia, Kentucky, North Carolina, Tennessee, Alabama, and Georgia. The RV ecoregion, with an area of 115,000 km^2, is an important crop production region that has about 56% forest cover. Almost all of the forest in these ecoregions is re-growth following cutting or following agricultural abandonment.

Land cover change information was generated by the U.S. Land Cover Trends Project (Loveland et

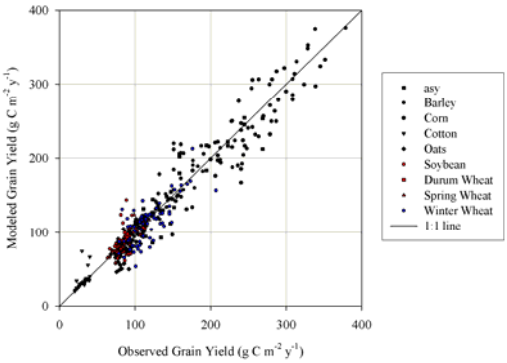

Figure 2. Comparison of grain yields between GEMS simulations and agricultural census data from 1973 to 2000 in selected states (e.g., SD, ND, WY, MT).

al., 2002). The land cover types were derived from five dates of the Landsat MSS, TM, and ETM+ data (nominally 1973, 1980, 1986, 1992, and 2000), which were analysed at a common pixel size of 60-m resolution. Ten land cover types were used including water, developed (urban), human disturbed, mining, natural barren, forest, grass and shrub, agricultural, wetland, and natural disturbed area.

Historical grain yield from agricultural census data, and forest inventory data (e.g., biomass stock and growth rates) were used during data assimilation to derive optimal model parameters (e.g., gross primary productivity). Figure 2 shows the comparison between the simulated and observed grain yields for the major crops from 1973 to 2000 in four states.

Figure 3 shows the land cover change, carbon stock change, standard deviation and coefficient of variation of the carbon stock simulated by GEMS in a 10-km by 10-km block in the Ridge and Valley ecoregion. The block consisted of 64% forest and 18% crop land. Model simulations were performed at 60-m length-scale resolution. The figure clearly shows that the spatial patterns of C stocks and uncertainties were (1) relatively stable over time (i.e., the pictures look similar across frames in a row), (2) strongly dependent on land cover classes with highest C stocks in forests, and (3) variable within land cover classes.

4. CONCLUSIONS

We develop the General Ensemble Biogeochemical Modeling System (GEMS) for upscaling carbon stocks and fluxes from sites to regions with measures of uncertainty. At the site

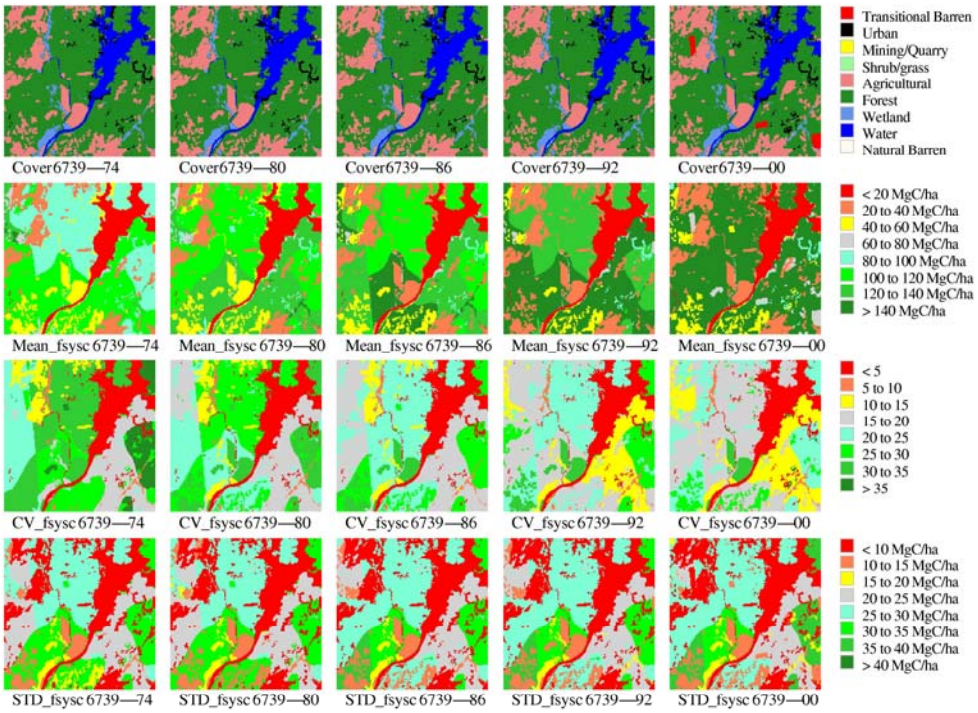

Figure 3 Spatiotemporal changes of land cover, patterns of the mean and uncertainty (i.e., CV -- coefficient of variation, and STD -- standard deviation) of ecosystem carbon stocks (fsysc) simulated by GEMS in a 10-km by 10-km sample block (#39) in the Ridge and Valley ecoregion

scale, GEMS uses stochastic ensemble simulations to incorporate input uncertainty and to quantify uncertainty transfer from input to output. GEMS relies on site-scale biogeochemical models (e.g., the Erosion-Deposition-Carbon Model (EDCM) and CENTURY) to simulate the carbon dynamics at the fine scale. The modeling philosophy embedded in GEMS makes it ideal for incorporating and assimilating information with various uncertainties at a range of spatio-temporal resolutions.

5. ACKNOWLEDGEMENTS

Work performed by S Liu, Z Li, J Liu and M Chen is under SAIC contract 03CRCN0001. The research was funded by USGS's Earth Surface Dynamics Program and the Geographic Analysis and Monitoring Program.

6. REFERENCES

Liu, J., Liu, S. and Loveland, T.R. 2006. Temporal Evolution of Carbon Budgets of the Appalachian Forests in the U.S. from 1972 to 2000. Forest Ecology and Management, 222, 191-201.

Liu, S., Bliss, N., Sundquist, E., and Huntington, T.G., 2003, Modeling carbon dynamics in vegetation and soil under the impact of soil erosion and deposition: Global Biogeochemical Cycles, v. 17, no. 2, p. 1074–1097.

Liu, S., Loveland, T.R., and Kurtz, R.M., 2004a, Contemporary carbon dynamics in terrestrial ecosystems in the Southeastern Plains of the United States: Environmental Management, v. 33, p. S442–S456.

Liu, S., Kaire, M., Wood, E., Diallo, O., and Tieszen, L.L., 2004b. Impacts of land use and climate change on carbon dynamics in south-central Senegal. Journal of Arid Environment 59 (3): 583-604.

Loveland, T.R., Sohl, T.L., Stehman, S.V., Gallant, A.L., Sayler, K.L. and Napton, D.E., 2002. A strategy for estimating the rates of recent United States land-cover Changes. Photogrammetric Engineering & Remote Sensing 68(10): 1091-1099.

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 Science Society of America Journal, p. 1173-1179.

Tan, Z., S. Liu, C. A. Johnston, T. R. Loveland, L. L. Tieszen, J. Liu, and R. Kurtz (2005), Soil organic carbon dynamics as related to land use history in the northwestern Great Plains, Global Biogeochem. Cycles, 19, GB3011, doi:10.1029/2005GB002536.

Xiao, X., Zhang, Q., Braswell, B., Urbanski, S., Boles, S., Wofsy, S., Moore, B., Ojima, D., Modeling gross primary production of temperate deciduous broadleaf forest using satellite images and climate data, *Remote Sensing of Environment,* 91, 256-270 2004.

Zhao, M., Heinsch, F., Nemani, R., Running, S., Improvements of the MODIS terrestrial gross and net primary production global data set, *Remote Sensing of Environment,* 95, 164-176, 2005.

Arsham H. (1998), Techniques for Monte Carlo optimizing, Monte Carlo Methods and Appl. Vol. 4, No.3: 181-229

Braswell B., et al (2005), Estimating diurnal to annual ecosystem parameters by synthesis of a carbon flux model with eddy covariance net ecosystem exchange observations, Global Change Biology 11: 89-105

Chen M., et al (2004), Dynamic data driven wildfire modeling, F. Darema (ed.) Dynamic Data Driven Applications Systems. Kluwer Acadmec Publishers, the Netherlands.

Chen M., et al (2005), Towards a Dynamic Data Driven Application System for Wildfire Simulation. pp. 632-639. In V.S. Sunderam et al. (eds.) Computational Science – Proceedings ICCS'2005, Lecture Notes in Computer Science. 3515.

Evensen G. (2003), The ensemble kalman Filter: theoretical formulation and practical implementation. Ocean Dynamics 53: 343-367.

Ngodock, H., M. Chen and G. Jacobs (2005), The representer method, the Ensemble Kalman filter and the Ensemble Kalman smoother: a comparison study using a nonlinear reduced gravity ocean model. Ocean Modeling (in press)

Raupach, M.R., Rayner P. J., Barrett D. J., DeFries R. S., Heimann M., Ojima D. S., Quegan S., and Schmullius C. C, 2005. Model-data synthesis in terrestrial carbon observation: methods, data requirements and data uncertainty specifications, Global Change Biology, 11(3), 378-397(20).

Reiners, W.A., S. Liu, K.G. Gerow, M. Keller, and D.S. Schimel. 2002. Historical and future land use effects on trace gas emissions using an ensemble modeling approach: Costa Rica's Caribbean Lowlands as an example. Global Biogeochemical Cycles 16[4]:1068, doi: 10.1029/2001GB001437.