# ESTIMATING SOIL CARBON LEVELS USING AN ENSEMBLE KALMAN FILTER

J. W. Jones, W. D. Graham, D. Wallach, W. M. Bostick, J. Koo

**ABSTRACT.** *Soils have been proposed as carbon storage sinks to help reduce atmospheric carbon dioxide levels and global warming. Benefits could accrue to farmers, due to beneficial effects of soil organic carbon on productivity, and to society by managing land to increase soil carbon. Measurements are needed to determine if practices aimed at increasing carbon levels are effective and to quantify amounts of carbon stored for verification purposes. However, measurements are expensive and have high errors relative to annual changes in soil carbon. In this article, we develop an Ensemble Kalman filter (EnKF) approach that combines measurements with predictions from a simple model, taking into account errors in measurements, model parameters, and model predictions. The EnKF was used to estimate soil carbon at annual time steps and to estimate an uncertain soil carbon decomposition rate parameter. A sensitivity analysis was conducted to evaluate the effects on EnKF estimates of uncertainties in measurements, model predictions, and the decomposition rate parameter. The EnKF estimates of soil carbon were compared with true values that were generated using a Monte Carlo method. Results showed that EnKF estimates of soil carbon levels and annual changes of soil carbon were more accurate than measurements alone for all combinations of conditions studied. The root mean square error of estimation was reduced from around 700 kg/ha based on measurements alone to about 225 kg/ha using the EnKF procedure. The unknown soil carbon decomposition parameter converged to its true value after about seven years. This EnKF method can be modified to incorporate more comprehensive models of cropping systems and soil carbon, to incorporate spatial variability, and to assimilate remote sensing inputs. It is simple to implement and has considerable promise for practical use in soil carbon sequestration projects.*

**Keywords.** *Data assimilation, Soil carbon, Stochastic model, Uncertainty.*

Increasing atmospheric carbon dioxide ($CO_2$) and other trace gases are causing increases in air temperature and possibly affecting regional patterns of precipitation and hydrological processes (IPCC, 1996; Rosenzweig and Hillel, 1998). The concentration of $CO_2$ in the atmosphere at the end of the 20th century exceeded 360 volumetric parts per million (ppmv), higher than that projected by Waggoner in 1969 (Allen, 1994), and continuing increases of about 1% per year are projected (IPCC, 1995, pp. 196-197). Research has shown that global average temperature has increased by about 0.5°C since the early 1960s (Parker et al., 1994). A number of studies conducted during the last 15 years have projected likely negative impacts of climate change on agriculture as well as other sectors (Adams et al., 1990; Rosenzweig et al., 1995; Kaiser and Drennen, 1993; IPCC, 1996). Methods for reducing this buildup of $CO_2$ in the Earth's atmosphere have been suggested, and include reduction in emission of $CO_2$ as well as increasing storage of carbon (C) in forests, soils, and the oceans (Rosenberg, 2000; IPCC, 1996). Research is being conducted to determine the potential for storing C in soils, taking into consideration biophysical, land management, and economic factors (Antle et al., 2001; Antle and McCarl, 2001; Yost et al., 2002; Jones et al., 2002). Incentives for this research derive from not only concern about global climate change, but also from the considerable benefits that higher soil organic matter levels will have on agricultural productivity, especially in developing countries where soil degradation has led to decreases in yield and food insecurity (Antle and Uehara, 2002).

If soil C sequestration is to become an accepted mechanism for reducing atmospheric $CO_2$ levels, a soil carbon accounting system needs to be developed (Antle and Uehara, 2002). Direct measurements of soil C can be made, but they are expensive, and yearly changes in soil C are small relative to the errors associated with sampling and measuring soil C. For example, consider a field in which the mass of organic C in the top 20 cm of a soil is 16,000 kg[C]/ha (about 0.6% C on a mass basis), and annual change in soil C is about 300 kg[C]/ha (about 0.01% on a mass basis). Yost et al. (2002) reported standard deviations of soil C measurements that ranged between 0.05% and 0.50%. Thus, standard errors of soil C measurement may be several times higher than the annual change in soil C. Although geostatistical methods may help reduce these errors (Yost et al., 2002), measurement errors will remain high. Biophysical models can be used to

Article was submitted for review in May 2003; approved for publication by the Information & Electrical Technologies Division of ASAE in November 2003.

Florida Agricultural Experiment Station Journal Series No. R-09486. This research was supported in part by the Soil Management CRSP Project, funded by US AID, and by the "Carbon from Communities: A Satellite View" project funded by NASA.

The authors are **James W. Jones, ASAE Fellow Member,** Distinguished Professor, **Wendy D. Graham, ASAE Member Engineer,** Professor, **Welch M. Bostick, ASAE Student Member,** Research Associate, and **Jawoo Koo, ASAE Student Member,** Research Associate, Department of Agricultural and Biological Engineering, University of Florida, Gainesville, Florida; and **Daniel Wallach,** Department Head, Department of Agronomy, INRA, Toulouse, France. **Corresponding author:** James W. Jones, P.O. Box 110570, 289 Rogers Hall, University of Florida, Gainesville, FL 32611; phone: 352-392-1864, ext. 289; fax: 352-392-4092; e-mail: jjones@agen.ufl.edu.

estimate soil C and its changes under different weather, soil, and management practices (Parton et al., 1988, 1994; Jones et al., 2002). However, although these models may produce precise estimates, they are imperfect, and parameters for specific field situations are uncertain. Thus, predictions of changes in soil C by models are also uncertain.

Techniques exist to combine models and measurements to obtain estimates of system states, such as soil C, and model parameters, such as decomposition coefficients for soil organic C. One widely used technique is the Kalman filter (Maybeck, 1979; Welch and Bishop, 2002). This technique is based on dynamic models that describe rates of change of state variables and predict their values over time. The Kalman filter approach starts out with a model prediction to estimate the state of a system and then uses measurements to conditionally update the estimates in an optimal way. The uncertainties in model predictions and measurements are used along with measurements at a point in time to obtain a maximum likelihood estimate of the true state of the system and the uncertainty associated with this estimate. When the model is linear and errors are Gaussian and time independent, procedures are relatively straightforward. This approach has been applied to a number of problems in agriculture and forestry (Or and Groeneveld, 1994; Tani et al., 1992; Wendroth et al., 1999). Extensions to the Kalman filter have been developed to overcome difficulties associated with non-linear models (e.g., Albiol et al., 1993; Graham, 2002). The Extended Kalman filter linearizes a non-linear model at each discrete time step before applying the Kalman filter estimator. This approach is also useful when parameters of the model are to be estimated. It works well for simple models, but linearizing complex models, such as soil C models or crop models with uncertain parameters, is complicated. Another extension to Kalman filtering has been developed for such situations, the Ensemble Kalman filter (Burgers et al., 1998; Eknes and Evensen, 2002; Margulis et al., 2002).

The purpose of this article is to evaluate the use of Ensemble Kalman filter (EnKF) methodology for estimating soil C and its changes over time. To do this, we introduce a simple model of soil C dynamics with one state variable and one unknown parameter. The state estimate and state covariance values are propagated between measurement times using Monte Carlo methods, and they are updated periodically when measurements are available. Using a synthetic example, measurements are simulated using the model to estimate the true states at the measurement time, and then a random variation around the perfect state is added based on an assumed measurement error variance. A sensitivity analysis is performed to demonstrate the effects of uncertainties in measurements, model predictions, and model parameters on uncertainty in soil C estimates. This article focuses on estimation of soil C over time (years) at a point in space; it does not address spatial variability or aggregation of estimates over large areas.

## Materials and Methods

### Soil Carbon Model

A discrete time model is used to simulate soil organic carbon ($X_t$) as it changes over time, using a time step of one year. We assume that there is one pool of C in the soil and that fresh organic matter carbon ($U_t$) may increase this pool, while during the same annual time step, microbial activity decomposes both existing soil C and $U_t$. The model also has one parameter (the rate constant for decomposition, R) that is constant over time, but it is not known with certainty. Conceptually, the field in which soil C estimates are being made belongs to a population of fields with decomposition rate parameters that are not known with certainty. The R value for the field in question thus belongs to a distribution of true values that are normally distributed with mean and variance. Here, we assume that the mean R is $R_0$, and this value is used as an initial estimate of R in the EnKF procedure. The resulting model thus has one state variable ($X_t$) and one parameter (R), which are to be estimated using the EnKF. We assume that there are uncertainties in model predictions of soil C and in the decomposition rate parameter. State equations for the non-linear model are:

$$X_t = X_{t-1} - R \cdot X_{t-1} + b \cdot U_{t-1} + \varepsilon_t$$
$$R = R_0 + \eta \tag{1}$$

where

$X_t$ = soil organic carbon in year $t$ (kg[C]/ha)
R = rate of decomposition of existing soil C (1/yr)
$R_0$ = initial estimate of soil C decomposition rate (1/yr)
$b$ = fraction of fresh organic C that is added to the soil in year $t$ that remains after one year
$U_t$ = amount of C in crop residue that is added to the soil in year $t$
$\varepsilon_t$ = model error for soil C (kg[C]/ha)
$\eta$ = error in initial estimate of decomposition rate R (1/yr).

Model error ($\varepsilon_t$) includes uncertainties in U and $b$ as well as uncertainties due to the fact that the model is a simplification of reality. We assume that model errors and the parameter estimator error are normally distributed and are not correlated. Thus:

$$\varepsilon_t \sim N\left(0, \sigma_\varepsilon^2\right)$$
$$\eta \sim N\left(0, \sigma_\eta^2\right) \tag{2}$$

where

$\sigma_\varepsilon^2$ = variance of model error for soil C
$\sigma_\eta^2$ = variance of error for estimate of soil C decomposition rate R.

The model error ($\varepsilon_t$) is a random process that changes over time but is uncorrelated with time (i.e., white noise), whereas the decomposition rate parameter error ($\eta$) is a random variable that does not change with time.

### Measurements

In order to evaluate the use of EnKF to estimate soil C, a time series of measurements was necessary. Soil C measurements ($Z_t$) may be made each year or less frequently, but measurements of R are not possible. Thus, the model has two variables that are to be estimated, but only one is observable. Furthermore, it was assumed that the soil C measurement error is normally distributed, independent in time and independent from X and R. In this article, a time series of measurements was generated using equation 1 (to generate true soil C values) and an assumed measurement error. In reality, true values of soil C are seldom, if ever, known. Measurement values were generated using two steps. First, one time series realization of soil C ($X_t$) was computed using equation 1 with the true value of the parameter R. Then, measurements were generated by randomly sampling from the distribution of Z ($\varepsilon_Z$) and adding this random error to soil C values at each discrete time step. Thus, $Z_t$ was generated by:

$$Z_t = X_t + \varepsilon_{z,t} \tag{3}$$

where

$Z_t$ = measurement of soil C in year $t$ (kg[C]/ha)

$\varepsilon_{z,t}$ = error in measurement, $\varepsilon_{z,t} \sim N(0,\sigma_Z^2)$

This procedure was repeated for different values of $\varepsilon_Z$ and $R_0$ for sensitivity analyses, as described below. In real applications, actual measurements would be used.

### The Ensemble Kalman Filter

The Kalman filter is a set of mathematical equations that are used to obtain optimal estimates of the state of a system. There are two types of equations in a Kalman filter: (1) time update equations, and (2) measurement update equations (Welch and Bishop, 2002). The time update equations project forward in time the current predictions of the system state and covariance. The measurement equations provide feedback by incorporating a new measurement to obtain an improved estimate of system state and covariance. In a discrete-time Kalman filter, a linear stochastic model is used to project the state and covariance estimates forward to the next time step $k$. At measurement times, the model-projected state and covariance values are updated by using the measurement and its covariance characteristics. A Kalman gain matrix is computed to update estimates of system state and covariance. This process is repeated over time in a recursive fashion, projecting values for each discrete time step and updating those estimates for time steps when measurements are available.

The Ensemble Kalman filter (EnKF) follows this same general approach for non-linear models but relies on Monte Carlo methods to project state and covariance values between measurement times (Burgers et al., 1998; Margulis et al., 2002). The soil C model (eq. 1) is non-linear due to multiplication of R and X, both "states" of the system to be estimated. For our model, the EnKF is used to estimate states of the system for each time step (denoted $\hat{X}_t$ and $\hat{R}_t$). The EnKF used in this article was adapted from Margulis et al. (2002). Measurements are combined with model predictions to obtain the best estimates of soil C and the decomposition rate parameter, given all of the measurements that have been made up to the current time.

A "truth" scenario is first generated using equation 1 for $X_t$, with one particular sequence of $\varepsilon_t$ and value of R to represent reality. The EnKF uses Monte Carlo simulation to propagate an ensemble of equally likely $X_t$, each with a unique sequence of $\varepsilon_t$ and value of R. This is the full range of possible outcomes $\{X_t\}$. Adding a random measurement error to the "truth" scenario described above produces "measurements" of reality. Each equally likely outcome in the ensemble of realizations is then updated, using the Kalman gain, to take advantage of the measurement of soil C at time $t$. This step will produce a new ensemble of estimates, which will have reduced variance relative to model-projected estimates. The mean of this updated ensemble is the optimal estimate of soil C ($\hat{X}_t$) and the decomposition rate parameter ($\hat{R}_t$).

### Ensemble Creation, Propagation, and Update

A number of pairs (X, R) are generated at time $t = 0$ as ensemble replicates or realizations of the initial values of soil C and the decomposition rate parameter, respectively. The basis for these pairs of variables is our prior knowledge of their probability distributions, namely $\varepsilon_0$ and $\eta$. We assume that initial values of X and R are not correlated. Then, each ensemble replicate ($X^j_{t|z,t}$ and $R^j_{t|z,t}$) is simulated, using the model shown in equation 1 and a Monte Carlo approach to produce random deviates at each time step. This provides an ensemble of equally likely realizations of X and R at time $t$ + 1, denoted $X^j_{t+1|z,t}$ and $R^j_{t+1|z,t}$, before using the new measurement to update the estimate at $t + 1$. Each ensemble replicate is treated the same. After the model is used to simulate the values of both X and R at time $t + 1$, the Kalman gain matrix is used along with the measurement to update the estimates for each ensemble member, denoted $X^j_{t+1|z,t+1}$ and $R^j_{t+1|z,t+1}$ (where $j$ represents the $j$th member of the ensemble), taking into account the measurement taken at $t + 1$.

Figure 1 shows this process of using the model to estimate values of soil C at time $t + 1$ for each replicate and the updated estimates of soil C, conditioned on measurement at time $t$ + 1. Although the figure only shows points depicting soil C, each ensemble replicate consists of both the soil C ($X^j_{t|z,t}$) and the rate parameter ($R^j_{t|z,t}$) estimates. After the state variable and parameter are updated at a time step, the resulting ensemble of values ($X^j_{t+1|z,t+1}$ and $R^j_{t+1|z,t+1}$) is used as initial values for another prediction-updating step at time $t = t + 2$. If there are no measurements at a time step, then the uncorrected values of the ensemble are used as initial values for the next step. This process continues sequentially, first by using equation 1 to propagate the variables over a time step, and then by updating the values at each time that a measurement is available.

The equations for the update step are given by:

$$\begin{aligned} X^j_{t+1|z,t+1} &= X^j_{t+1|z,t} + K_{X,t+1} \cdot \left(Z_{t+1} + \varepsilon^j_{z,t+1} - X^j_{t+1|z,t}\right) \\ R^j_{t+1|z,t+1} &= R^j_{t+1|z,t} + K_{R,t+1} \cdot \left(Z_{t+1} + \varepsilon^j_{z,t+1} - X^j_{t+1|z,t}\right) \end{aligned} \tag{4}$$

where $K_{X,t+1}$ and $K_{R,t+1}$ are Kalman gain values for updating X and R, respectively. In the equations above, the superscript $j$ represents an ensemble replicate. The use of $|z,t$ in the subscript indicates that estimates of the variables have been conditioned on measurements made up to time $t$ (or time $t$ + 1). The $\varepsilon^j_{z,t+1}$ is a random deviate based on the error associated with measurement of soil C for the $j$th replicate at time $t + 1$ (Margulis et al., 2002). This random sample is taken from the distribution of measurement error to adjust each replicate measurement value to account for variations among samples. Since an ensemble is used at each step, expected

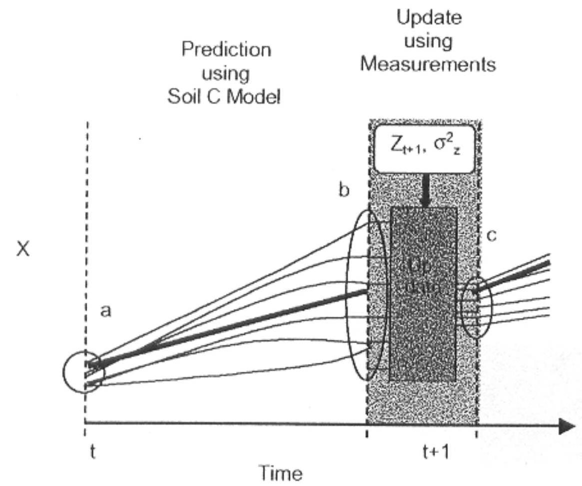


**Figure 1. Schematic of the Ensemble Kalman filter procedure. An ensemble of variables, created at time 0 (a), is simulated over time. The model is used to predict the state variables (for each ensemble replicate) for time $t+1$ (b) using values at time $t$. Then measurements are used to update the estimates of variables for time ($t+1$). These updated values (c) are used as initial values for the next prediction/update step. The heavy line in the figure shows the true value of soil C.**

values of X and R ($\hat{X}_t$ and $\hat{R}_t$) can be estimated at each time step, as well as their variances and covariance.

For our model, which has two equations and two variables, the Kalman gain matrix can be written as:

$$\begin{bmatrix} K_{X,t} \\ K_{R,t} \end{bmatrix} = \tag{5}$$

$$\begin{bmatrix} \sigma^2_{X,t} & \sigma_{XR,t} \\ \sigma_{XR,t} & \sigma^2_{R,t} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \left\{ \begin{bmatrix} 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \sigma^2_{X,t} & \sigma_{XR,t} \\ \sigma_{XR,t} & \sigma^2_{R,t} \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \sigma^2_Z \right\}^{-1}$$

where the terms of the covariance matrix are the variance of soil C predictions at time $t$ ($\sigma^2_{X,t}$), the variance of estimates of soil C decomposition rate at time $t$ ($\sigma^2_{R,t}$), and the covariance between X and R estimates at time $t$ ($\sigma_{XR,t}$). These variance and covariance values are estimated from the ensemble before state estimates are updated. The measurement matrix for this problem is the vector [1 0], since measurement is made only of soil C, the first state variable, and R is not measured. See Welch and Bishop (2002) for the general formulation of the Kalman gain matrix. The terms inside the brackets in equation 5 simplify to the scalar term $\left[\sigma^2_{X,t} + \sigma^2_Z\right]^{-1}$. Thus, after matrix multiplication, equation 5 can be simplified and written as two terms, as follows:

$$K_{X,t} = \frac{\sigma^2_{X,t}}{\sigma^2_{X,t} + \sigma^2_Z}$$

$$K_{R,t} = \frac{\sigma_{XR,t}}{\sigma^2_{X,t} + \sigma^2_Z} \tag{6}$$

Note that although R is not measured, the measurement of soil C provides information for refining the estimate of R via the covariance term. Note also that K varies with time; it is recalculated each time a measurement is made using covariance terms computed from the ensemble before the update step.

The Kalman gain variables are used to weight the updated estimate on the basis of error variances. Note, for example, that if measurement error variance ($\sigma^2_Z$) is very small relative to model prediction variance ($\sigma^2_{X,t}$), then $K_{X,t}$ approaches 1.0, and the updated estimate of $X^j_{t+1|z,t+1}$ (eq. 4) will be approximately the value that was measured. In contrast, if measurement error is large relative to prediction error, then $K_{X,t}$ will be closer to 0.0, and the updated estimate will be near the predicted value. Furthermore, if the covariance term used to compute $K_{R,t}$ is small, then the estimate of R ($R^j_{t+1|z,t+1}$) will remain near its estimate from the previous step. However, if the covariance term is large, then the differences between measured and predicted soil C will result in adjustments to R in the update step.

## Implementation of the Ensemble Kalman Filter

A computer program was written to implement the EnKF. Values for model inputs and parameters that represent soils in western Africa were used. First, a set of base case values is described. Then a sensitivity analysis is described in which various inputs and parameters are changed to study their impacts on EnKF estimates of soil C and the decomposition rate parameter.

### *Base Case*

The initial value of soil C was assumed to be 16,000 kg[C]/ha in the top 20 cm of soil, which is about 0.6% carbon on a mass basis. This value approximates the levels of soil carbon found by Yost et al. (2002) and J. B. Naab (personal communication) in agricultural fields in Mali and Ghana, respectively. Variance of this initial soil C estimate was assumed to be 20,000 (kg[C]/ha)$^2$ (standard deviation of 141 kg[C]/ha). We also assumed that the model error variance ($\sigma^2_\varepsilon$) was equal to 20,000 (kg[C]/ha)$^2$. The initial value of $R_0$ was assumed to be 0.020, which is in the range of values given by Pieri (1992) for soils in western Africa. The uncertainty in R was represented by a variance ($\sigma^2_\eta$) of 0.0001. The value of $U_t$ was set at a relatively high value of 2,000 kg[C]/ha, constant across all years. This value represents a crop with a vegetative biomass of 5,000 kg/ha, assuming 40% of the biomass is C. The value of $b$ was assumed to be 0.20, meaning that 20% of the vegetative biomass would remain on the field. Typically in western Africa, animals graze fields after grain is harvested. The variance of soil C measurements ($\sigma^2_Z$) was set at 500,000, which is a standard deviation of 707 kg[C]/ha or a standard error of measurement of 0.0253% C on a mass basis. This value is somewhat low relative to the range in standard deviations of measurement errors (0.058% to 0.21% C) reported by Yost et al. (2002); additional values will be used in the sensitivity analysis. The number of replicates was set to 1,000, and measurement frequency was set to one measurement each year. In preliminary runs, results were not affected when the number of replicates was varied between 100 and 4,000.

Equation 3 was used to generate measurements ($Z_t$) for $t$ = 1 through 50 years, starting with an initial value of soil C of 16,000 kg[C]/ha and a true value of R of 0.010 for the specific field. The EnKF estimates of $\hat{R}_t$ should converge to the true value for the specific field. Table 1 summarizes the values of parameters and initial conditions used to simulate measurements and to implement the EnKF for the initial set of values. The EnKF was used to estimate X and R and their variances for each of the 50 years for which measurements were generated.

Results of mean estimates of $\hat{X}_t$ and $\hat{R}_t$ and their variances ($\sigma^2_{X,t}$ and $\sigma^2_{R,t}$), computed from estimates made after the EnKF update procedure) were plotted versus time to demonstrate behavior of the EnKF procedure. The resulting estimates of soil C and its annual changes were compared with those estimated using measurements alone. Root mean square errors (RMSE) between true soil C in each year and measurements ($RMSE_Z$) were computed and compared with RMSE between true soil C and EnKF estimates ($RMSE_{KF}$) for the first 10 years. In addition, annual changes in soil C estimated from measurements ($Z_t - Z_{t-1}$ ) and from EnKF estimates (using expected values of replicates, $\hat{X}_t - \hat{X}_{t-1}$) were compared with true values that were generated, and RMSE values of annual soil C changes were computed using these errors over the first 10 years ($RMSE'_Z$ and $RMSE'_{KF}$, respectively).

### *Sensitivity Analysis*

A sensitivity analysis was conducted to characterize the effects of different variables on uncertainties in soil C estimation and on the stability of the decomposition rate parameter estimates. Table 2 summarizes the variables modified in this analysis and the values used for each run. The values shown in the first row of table 2 are those used in the base case. For each sensitivity analysis run, all variables were kept at their base case values except the variable being studied. For example, when studying the effects of the mean value of R, three runs were made changing $R_0$ for each run (to 0.005, 0.03, and 0.06) but keeping the other variables at the values listed in the top row of table 2. Results were analyzed by comparing values of expected values of ($\hat{X}_t$, $\hat{R}_t$) with measurements and true values. In addition, variances in estimates of both $\hat{X}_t$ and $\hat{R}_t$ were compared along with $RMSE_Z$, $RMSE_{KF}$, $RMSE'_Z$, and $RMSE'_{KF}$.

**Table 1. Values of parameters, initial conditions, and inputs for the base case implementation of the Ensemble Kalman filter.**

| Variable | Definition | Units | Value |
|---|---|---|---|
| $X_0$ | True value of soil C at time 0 | kg[C]/ha | 16,000 |
| R | True value of mineralization parameter | 1/yr | 0.010 |
| $\sigma^2_Z$ | Variance of measurement, constant over time | $(kg[C]/ha)^2$ | 500,000 |
| $\sigma^2_\varepsilon$ | Variance in model estimates of soil C, each year time step | $(kg[C]/ha)^2$ | 20,000 |
| $R_0$ | Initial value of soil C decomposition parameter | 1/yr | 0.015 |
| $\sigma^2_\eta$ | Variance of decomposition rate parameter | $(1/yr)^2$ | 0.0001 |
| $U_t$ | Input of C to the soil each year (assumed constant) | kg[C]/ha | 2,000 |
| $h$ | Proportion of annual soil C that remains after one year | - - | 0.20 |
| nreps | Number of ensemble replicates used | - - | 1000 |
| Zfreq | Measurement frequency | 1/yr | 1 |

**Table 2. Values used for each variable in sensitivity analysis. Values in first row (with *) were used as the base case. A total of 19 runs were made. Variables are defined in table 1.**

| $R_0$ | $\sigma^2_\eta$ | $\sigma^2_\varepsilon$ | $\sigma^2_Z$ | $U_t$ | Zfreq |
|---|---|---|---|---|---|
| 0.020* | 0.0001* | 20,000* | 500,000* | 2,000* | 1* |
| 0.005 | 0.00001 | 1,000 | 10,000 | 0 | 1/2 |
| 0.030 | 0.00005 | 40,000 | 2,000,000 | 1,000 | 1/3 |
| 0.060 | 0.00015 | 80,000 | 8,000,000 | 4,000 | 1/5 |

## Results

### Base Case

Soil C increased by 8,500 kg[C]/ha in the base case simulations over the 50-year time period. Figure 2 shows measurements, EnKF estimates, and true values of soil C over time for the base case (second line from top of graph). It is clear from this figure that the EnKF estimates were closer to the true values (heavier line) most of the time, and that those estimates did not fluctuate as much from year to year as measurements of soil C. The standard error of EnKF estimates of soil C was 397 kg[C]/ha in the tenth year (table 3). This error converged to a minimum value of about 300 kg[C]/ha after 30 years, which is less than half of the measurement error assumed for this base case (standard deviation of measurement = 707 kg[C]/ha). The $RMSE_{KF}$ value was 226 over the 50 years.

Annual increases in true soil C values varied from year to year due to the random variability denoted in equation 1. Figure 3 shows that annual changes in soil C varied from about -100 to about 460 kg[C]/ha per year for the first 30 years of the simulations. Estimates of annual changes in soil C based on measurements, however, varied between -1200 and +1800 kg[C]/ha per year. The EnKF estimates of annual changes were much more stable (fig. 3). $RMSE'_{KF}$ was 241 kg[C]/ha per year, considerably less than $RMSE'_Z$ value of 869 (table 3).

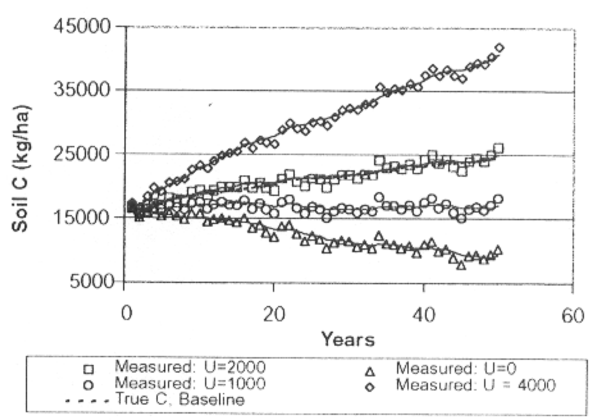


**Figure 2. Effect of different levels of fresh organic C input on soil C estimates from measurements and from the EnKF. Starting with the lowest curve, C inputs were 0, 1000, 2000, and 4000 kg[C]/ha each year, respectively. Solid lines are EnKF estimates, points are measurements, and the dashed line is the time course of true soil C values for the base case.**

Table 3. Effect of variance of measuring soil C on errors associated with soil C and decomposition rate parameter (R) estimates. Values in top row are for the base case. RMSE are errors in estimating true soil C from measurements (subscript z) and from the EnKF procedure. RMSE' are errors in estimating annual changes in soil C from measurements and from the EnKF procedure.

| Measurement Variance, $\sigma_Z^2$ (kg/ha)$^2$ | Standard Error of C Estimate, $t$ = 10 (kg/ha) | Standard Error of R Estimate, $t$ = 10 (1/yr) | $RMSE_Z$ (kg/ha) | $RMSE_{KF}$ (kg/ha) | $RMSE'_Z$ (kg/ha) | $RMSE'_{KF}$ (kg/ha) |
|---|---|---|---|---|---|---|
| 500,000 | 397 | 0.0036 | 577 | 226 | 869 | 241 |
| 10,000 | 86 | 0.0027 | 82 | 76 | 123 | 110 |
| 2,000,000 | 712 | 0.0050 | 1154 | 377 | 1738 | 288 |
| 8,000,000 | 1136 | 0.0073 | 2308 | 654 | 3477 | 312 |

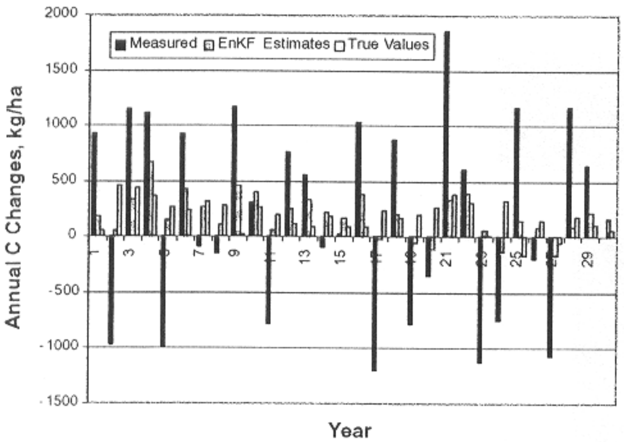


Figure 3. Annual estimates of soil C changes, comparing those made from measurements and from the EnKF with true values.

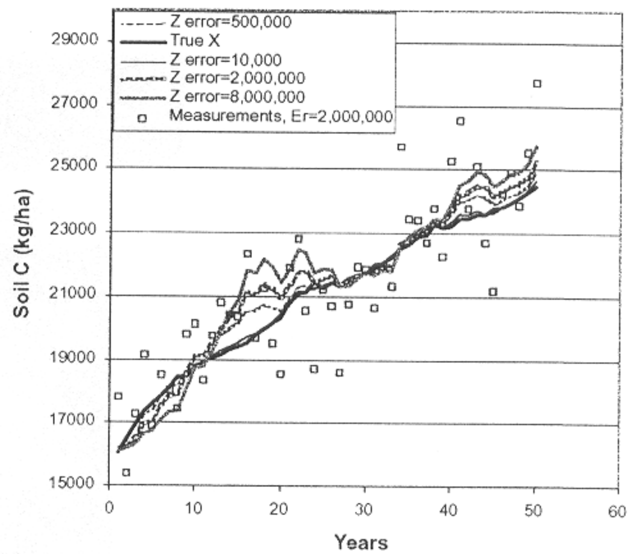


Figure 4. Ensemble Kalman filter estimates of soil C over time for different magnitudes of measurement error. The observations (symbols) were generated with an assumed measurement error variance of 2,000,000 (kg/ha)$^2$.

The decomposition rate parameter converged from its initial value of 0.020 to a value approximating the true field value of 0.010 after about seven years. The standard error of estimate for $\hat{R}_t$ after ten years was 0.0036 (1/yr) (table 3). This value continued to decrease over time as more measurements were used to update the estimates, converging to about 0.001 after 50 years.

### Effects of Measurement Error

Varying measurement error variance ($\sigma_Z^2$) had a large influence on standard errors of soil C estimate. At ten years, standard errors varied from 86 to 1136 kg[C]/ha as measurement error variance was increased from 10,000 to 8,000,000 (table 3). However, EnKF estimates of soil C and annual changes in soil C over 50 years varied much less; $RMSE_{KF}$ varied from 76 to 654, and $RMSE'_{KF}$ varied from 110 to 312 kg[C]/ha (table 3). Figure 4 shows the time courses of EnKF estimates for 50 years as well as the true values of soil C (heavy line). The measurements in figure 4 were generated with an error variance of 2,000,000. EnKF estimates of soil C based on different error variances follow the same pattern, but estimates derived from larger measurement errors deviate more from the true values, as expected. Estimates for the lowest value of measurement error, 10,000 (kg[C]/ha)$^2$, are almost indistinguishable from the true values. Although measurement error was less than model error for this case, the EnKF process still provided estimates of soil C with less error than the measurements (table 3), although not by much. Uncertainties in estimation of $\hat{R}_t$ varied with measurement error (table 3), but its expected values were not greatly affected by the large variations in measurement error used in this analysis (0.0108 for the base case at $t$ = 50 years vs. 0.0098 for the case with measurement error of 8,000,000).

### Effects of Model Error

Varying model error ($\sigma_\varepsilon^2$) from 1,000 to 80,000 only marginally affected uncertainties in EnKF estimates of soil C and its annual changes (table 4); at ten years, estimation errors ranged from 373 to 445 kg[C]/ha. $RMSE_{KF}$ was lowest (166) for the lowest model error used in the sensitivity analysis. This error increased to only 308 for model error of 80,000. Uncertainties in estimates of $\hat{R}_t$ ranged from 0.0024 to 0.0054 (1/yr) as model error varied from its lowest to highest value.

### Effects of Initial Value of Decomposition Rate Parameter

Changes in initial value of R ($R_0$) from less than half of the true value (0.005) to six times the true value (0.060) had little effect on errors of estimation of soil C after 10 years or on EnKF estimates of annual changes in soil C over 50 years ($RMSE'_{KF}$, table 5). However, errors in estimation of soil C over 50 years were degraded when $R_0$ was six times higher than the true value. After about 10 years, $\hat{R}_t$ values were close to the true value (fig. 5), regardless of its mean value.

Table 4. Effect of varying uncertainty in model predictions of soil C ($\sigma_\varepsilon^2$) on errors associated with soil C and decomposition rate parameter (R) estimates. Values in top row are for the base case. See table 3 for explanation of the different RMSE columns.

| Model Error $\sigma_\varepsilon^2$ (kg/ha)$^2$ | Standard Error of C Estimate, $t$ = 10 (kg/ha) | Standard Error of R Estimate, $t$ = 10 (1/yr) | $RMSE_Z$ (kg/ha) | $RMSE_{KF}$ (kg/ha) | $RMSE'_Z$ (kg/ha) | $RMSE'_{KF}$ (kg/ha) |
|---|---|---|---|---|---|---|
| 20,000 | 397 | 0.0036 | 577 | 226 | 869 | 241 |
| 1,000 | 373 | 0.0024 | 577 | 166 | 869 | 173 |
| 40,000 | 417 | 0.0044 | 577 | 260 | 869 | 292 |
| 80,000 | 445 | 0.0054 | 577 | 308 | 869 | 369 |

Table 5. Effect of initial estimate of R on errors associated with soil C and decomposition rate parameter (R) estimates. Values in top row are for the base case. See table 3 for explanation of the different RMSE columns.

| Initial Estimate of R ($R_0$) (1/yr) | Standard Error of C Estimate, $t$ = 10 (kg/ha) | Standard Error of R Estimate, $t$ = 10 (1/yr) | $RMSE_Z$ (kg/ha) | $RMSE_{KF}$ (kg/ha) | $RMSE'_Z$ (kg/ha) | $RMSE'_{KF}$ (kg/ha) |
|---|---|---|---|---|---|---|
| 0.020 | 397 | 0.0036 | 577 | 226 | 869 | 241 |
| 0.005 | 398 | 0.0036 | 577 | 261 | 869 | 238 |
| 0.030 | 397 | 0.0036 | 577 | 345 | 869 | 259 |
| 0.060 | 396 | 0.0038 | 577 | 860 | 869 | 366 |

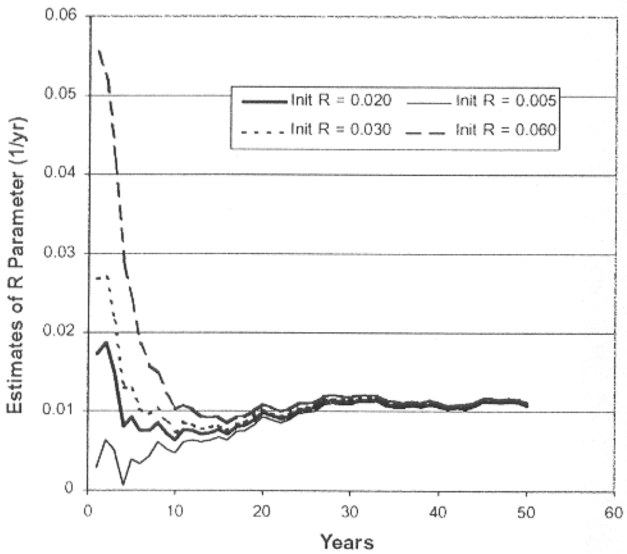


Figure 5. Effect of mean value of soil decomposition rate on estimates of R from the EnKF vs. time. Initial values of R were assumed to be the mean values in each case, and the true value was R = 0.01 for each case.

## Effects of Decomposition Rate Error

Table 6 shows uncertainties in EnKF estimates of soil C and decomposition rate for different assumed uncertainties in R. In most cases, these estimation errors were not affected much at all, surprisingly. However, it is interesting to note that the $RMSE_{KF}$ value was almost twice as high when the uncertainty in R was assumed to be very low (0.00001). Estimates of R improved very slowly from year to year in the EnKF procedure for this low value of uncertainty. This occurred because the actual error was high (a difference of 0.01 between initial and field value of R) compared with the assumed uncertainty in R. Estimates of $\hat{R}_t$ never dropped below 0.012 for this case, and only converged to this value after 15 years. This result demonstrates that the initial estimate of R must fall within the range characterized by the assumed error; otherwise the EnKF estimates may not converge.

## Effects of Annual Inputs of Carbon

Annual inputs of carbon ($U_t$) between 0 and 4,000 kg[C]/ha per year resulted in considerable differences in soil carbon levels over time. Without any C input, soil C decreased by 7,300 kg[C]/ha over 50 years, and it increased by 24,300 kg[C]/ha when annual input was 4,000 C per year (fig. 2). However, errors in estimating soil C and the decomposition rate parameter remained nearly constant over this entire range of $U_t$ (table 7).

## Effects of Measurement Frequency

Varying frequency of measurements had a significant influence on estimation errors. During years when measurements were not made, errors in soil C estimation increased (fig. 6). When measurements were available for updating the estimate, errors decreased. Figure 6 demonstrates the fact that prediction errors grow considerably when predictions are made without conditionally updating estimates based on measurements. Annual estimates of soil C changes were not made, since measurement frequency varied in this case. Instead, estimates were compared of changes in soil C over 5-year intervals. Errors in estimates of 5-year soil C changes

Table 6. Effect of varying uncertainty in decomposition rate parameter ($\sigma_\eta^2$) on errors associated with soil C and decomposition rate parameter (R) estimates. Values in top row are for the base case. See table 3 for explanation of the different RMSE columns.

| R Rate Error $\sigma_\eta^2$ (1/yr)$^2$ | Standard Error of C Estimate, $t$ = 10 (kg/ha) | Standard Error of R Estimate, $t$ = 10 (1/yr) | $RMSE_Z$ (kg/ha) | $RMSE_{KF}$ (kg/ha) | $RMSE'_Z$ (kg/ha) | $RMSE'_{KF}$ (kg/ha) |
|---|---|---|---|---|---|---|
| 0.0001 | 397 | 0.0036 | 577 | 226 | 869 | 241 |
| 0.00001 | 344 | 0.0025 | 577 | 487 | 869 | 210 |
| 0.00005 | 388 | 0.0034 | 577 | 274 | 869 | 227 |
| 0.00015 | 400 | 0.0037 | 577 | 219 | 869 | 251 |

Table 7. Effect of varying annual inputs of fresh organic C on errors associated with soil C and decomposition rate parameter (R) estimates. Values in top row are for the base case. See table 3 for explanation of the different RMSE columns.

| Annual C Input from Crops, $U_t$ (kg/ha) | Standard Error of C Estimate, $t$ = 10 (kg/ha) | Standard Error of R Estimate, $t$ = 10 (1/yr) | $RMSE_Z$ (kg/ha) | $RMSE_{KF}$ (kg/ha) | $RMSE'_Z$ (kg/ha) | $RMSE'_{KF}$ (kg/ha) |
|---|---|---|---|---|---|---|
| 2,000 | 397 | 0.0036 | 577 | 226 | 869 | 241 |
| 0 | 387 | 0.0039 | 577 | 222 | 869 | 235 |
| 1,000 | 392 | 0.0038 | 577 | 224 | 869 | 238 |
| 4,000 | 405 | 0.0033 | 577 | 229 | 869 | 246 |

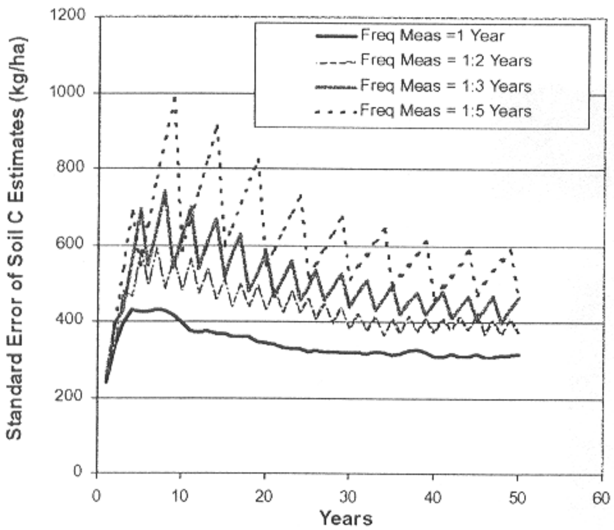


Figure 6. Variance of EnKF estimates of soil C vs. time for different frequencies of measurement (ranging from one per year to one every five years).

were 1,363 kg[C]/ha when measurements alone were used, 315 when the EnKF was used with annual measurements, and 581 when EnKF was used with measurements made at 5-year intervals.

## Discussion

Results of this analysis show that estimation of soil C can be improved by using the EnKF procedure. The use of the EnKF was superior to measurements alone in all cases compared. However, little information was available on which to base values of model error and decomposition rate error. Sensitivity analysis indicated that accurate estimate of error in R may not be critical, as long as it is not too low. However, soil C model error is critical. We used variances ranging from 1,000 to 80,000, corresponding to standard deviations of 32 to 383 kg[C]/ha. There are physical limits on how much C can be added from plant material each year, and limits on decomposition of existing soil C. In the cases analyzed, maximum soil C change per year was about 800 kg[C]/ha when $U_t$ was highest. Maximum annual decomposition was about -500 kg[C]/ha when $U_t$ was 0. Thus, model prediction was constrained between -500 and +800 kg[C]/ha per year, a range of 1,300. If we assume that maximum error is about 1/4th of this range, and is approximated by two standard deviations, then standard deviation of error would be about 162 (a variance of about 26,000), which compares well with the base case value of 20,000. Model error should probably depend on the amount of C added. For example, model error could be separated into that associated with bare soil plus that associated with addition of plant C.

Constant annual inputs of C from crops ($U_t$) were assumed, although we know that this is not likely to occur in nature. Measurements of annual crop residue added to the soil could be made, perhaps through the use of remote sensing, to implement this procedure under real conditions. Uncertainties in these measurements could also be included, as discussed above. Furthermore, a very simple model was used to evaluate the potential use of the EnKF approach. More complex models could be used to possibly reduce errors of estimation. For example, the simple model could be expanded to include more soil C pools, climate variable inputs, and a simple model to simulate crop biomass. It is also possible to incorporate existing soil carbon models, such as CENTURY (Parton et al., 1988) or DSSAT-CENTURY (Gijsman et al., 2002) into the EnKF procedure.

Finally, the procedure evaluated in this article estimated soil C at a point where measurements are made annually or less frequently. The ultimate value of this approach will be realized when it is used over large areas in which measurements are made only for a subset of fields. Additional work is needed to extend this approach for spatial and temporal estimation of soil C.

## References


Adams, R. M., C. Rosenzweig, R. M. Peart, J. T. Ritchie, B. M. McCarl, J. D. Glyer, R. B. Curry, J. W. Jones, K. J. Boote, and L. H. Allen, Jr. 1990. Global climate change and U.S. agriculture: An interdisciplinary assessment. *Nature* 345: 219-224.

Albiol, J., J. Robuste, C. Casas, and M. Poch. 1993. Biomass estimation in plant cell cultures using an extended Kalman filter. *Biotech. Progress* 9(2): 174-178.

Allen, L. H., Jr. 1994. Carbon dioxide increase: Direct impacts on crops and indirect effects mediated through anticipated climate change. In *Physiology and Determination of Crop Yield*, 425-459. K. J. Boote, J. M. Bennett, T. R. Sinclair, and G. M. Paulsen, eds. Madison, Wisc.: ASA.

Antle, J. M., and B. A. McCarl. 2001. The economics of carbon sequestration in agricultural soils. In *International Yearbook of Environmental and Resource Economics*, VI: 278-310. T. Tietenberg and H. Folmer, eds. Cheltenham U.K.: Edward Elgar Publishers.

Antle, J. M., and G. Uehara. 2002. Creating incentives for sustainable agriculture: Defining, estimating potential, and verifying compliance with carbon contracts for soil carbon projects in developing countries. In *A Soil Carbon Accounting System for Emissions Trading*, 1-12. Special Publication SM CRSP 2002-4. Honolulu, Hawaii: University of Hawaii.

Antle, J. M., S. M. Capalbo, and C. C. Crissman. 2001. Economic analysis of agricultural soil carbon sequestration: An integrated

assessment approach. *J. Agric. and Resource Economics* 26(2): 344-367.
Burgers, G., P. J. van Leeuwen, and G. Evensen. 1998. Analysis scheme in the Ensemble Kalman filter. *Monthly Weather Review* 126(6): 1719-1724.
Eknes, M., and G. Evensen. 2002. An Ensemble Kalman filter with a 1-D marine ecosystem model. *J. Marine Systems* 36(1-2): 75-100.
Gijsman, A. J., G. Hoogenboom, W. J. Parton, and P. C. Kerridge. 2002. Modifying DSSAT crop models for low-input agricultural systems using a soil organic matter-residue module from CENTURY. *Agronomy J.* 94: 462-474.
Graham, W. D. 2002. Estimation and prediction of hydrogeochemical parameters using Extended Kalman filtering. In *Stochastic Methods in Subsurface Contaminant Hydrology*, 327-363. R. S. Govindaraju, ed. Reston, Va.: ASCE Publications.
IPCC. 1995. *Climate Change 1994: Radiative Forcing of Climate Change and an Evaluation of the IPCC IS92 Emission Scenarios.* Intergovernmental Panel on Climate Change. New York, N.Y.: Cambridge University Press.
IPCC. 1996. *Climate Change 1995: The Science of Climate Change.* J. Houghton, L. B. Meira Filho, B. A. Callander, N. Harris, A. Kattenberg, and K. Maskell, eds. Intergovernmental Panel on Climate Change. New York, N.Y.: Cambridge University Press.
Jones, J. W., A. J. Gijsman, W. J. Parton, K. J. Boote, and P. Doraiswamy. 2002. Predicting soil carbon accretion: The role of biophysical models in monitoring and verifying soil carbon. In *A Soil Carbon Accounting System for Emissions Trading*, 41-68. Special Publication SM CRSP 2002-4. Honolulu, Hawaii: University of Hawaii.
Kaiser, H. M., and T. E. Drennen. 1993. *Agricultural Dimensions of Global Climate Change.* Delray Beach, Fla.: St. Lucie Press.
Margulis, S. A., D. McLaughlin, D. Entekhabi, and S. Dunne. 2002. Land data assimilation and soil moisture estimation using measurements from the Southern Great Plains 1997 field experiment. *Water Resources Research* 38(12): 1299-.
Maybeck, P. S. 1979. *Stochastic Models: Estimation, and Control*, vol. 1. Burlington, Mass.: Academic Press.
Or, D., and D. P. Groeneveld. 1994. Stochastic estimation of plant-available soil water under fluctuating water table depths. *J. Hydrology* 163(1-2): 43-64.
Parker, E. E., P. E. Jones, C. K. Folland, and A. J. Bevan. 1994. Interdecadal changes of surface temperature since the late nineteenth century. *J. Geophysical Research* 99(D7): 14373-14399.
Parton, W. J., J. W. B. Stewart, and C. V. Cole. 1988. Dynamics of C, N, P, and S in grassland soils: A model. *Biogeochemistry* 5(1): 109-131.
Parton, W. J., D. S. Ojima, C. V. Cole, and D. S. Schimel. 1994. A general model for soil organic matter dynamics: Sensitivity to litter chemistry, texture, and management. In *Quantitative Modeling of Soil Forming Processes*, 147-167. R. B. Bryant and R. W. Arnold, eds. Special Publication 39. Madison, Wisc.: SSSA.
Pieri, C. J. M. G. 1992. *Fertility of Soils: A Future for Farming in the West African Savannah*. Berlin, Germany: Springer-Verlag.
Rosenberg, N. J. 2000. Storing carbon in agricultural soils to help mitigate global warming. Issue Paper No. 14. Ames, Iowa: Council for Agricultural Science and Technology.
Rosenzweig, C., and D. Hillel. 1998. *Climate Change and the Global Harvest.* New York, N.Y.: Oxford University Press.
Rosenzweig, C., L. H. Allen, Jr., J. W. Jones, G. Y. Tsuji, and P. Hildebrand, eds. 1995. *Climate Change and Agriculture: Analysis of Potential International Impacts*. Special Publication 59. Madison, Wisc.: ASA.
Tani, A., H. Murase, M. Kiyota, and N. Honami. 1992. Growth simulation of alfalfa cuttings in vitro by Kalman filter neural network. *Acta Horticulturae* 319: 671-676.
Welch, G., and G. Bishop. 2002. An introduction to the Kalman filter. Report TR-95-041. Chapel Hill, N.C.: University of North Carolina, Department of Computer Science.
Wendroth, O., H. Rogasik, S. Koszinski, C. J. Ritsema, L. W. Dekker, and D. R. Nielsen. 1999. State-space prediction of field-scale soil water content time series in a sandy loam. *Soil and Tillage Research* 50(1): 85-93.
Yost, R. S., P. Doraiswamy, and M. Doumbia. 2002. Defining the contract area: Using spatial variation in land, cropping systems, and soil organic carbon. In *A Soil Carbon Accounting System for Emissions Trading*, 13-40. Special Publication SM CRSP 2002-4. Honolulu, Hawaii: University of Hawaii.