Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER Applied Mathematics and Computation 163 (2005) 783–800
APPLIED MATHEMATICS AND COMPUTATION
www.elsevier.com/locate/amc

Carbon sequestration: inversion of FACE data and prediction

L.W. White [a,*], Y. Luo [b], T. Xu [a]

[a] Department of Mathematics, University of Oklahoma, Norman, OK 73019, USA
[b] Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA

Abstract

FACE data is used with a compartmental carbon forest model to obtain information in the form of cumulative distributions and estimates of carbon transfer parameters in the face of uncertainties in initial conditions, flux inputs, and environmental observations. An assessment of the data is conducted to determine its effectiveness in inversion. Predictions are made of future carbon pools sizes based on the joint probability density function obtain from the data and trends of environmental functions. Uncertainty is carried forward into future predictions of pool sizes.

1. Introduction and objectives

In this paper we use a carbon sequestration model and relevant data sets to estimate transfer coefficients between carbon pools and to make predictions of the size of future carbon pools. A system of initial value problems based on the linear compartmental carbon sequestration concept [2] serves as the underlying set of admissible models. The solution (state) of the resulting initial value

* Corresponding author.
E-mail address: lwhite@ou.edu (L.W. White).

0096-3003/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.amc.2004.04.021

problem is a vector the ith component of which expresses the quantity of carbon residing in the ith carbon pool at a given time. A sample (or parameter) space is defined with respect to the set of parameters in admissible models that reflect those quantities of interest (the transfer coefficients) and uncertainties the initial conditions and flux terms as indicated by cumulative distribution functions. In addition the model also contains terms such as environmental conditions that are available as time series measurements. These time series give rise to possible uncertainties due to errors in measurements and noise. Our objective is to use the data from observations to obtain information on the parameters of interest, to assess the value of information from the data, and to predict future carbon pool sizes in the face of uncertainties.

To carryout the above program, various data are compared with model outputs of the corresponding attributes of the system that are formulated in terms of model parameters and states. The data are thus associated with mappings of the parameter and state vectors. These mapping are regarded as part of the model as well. Because initial value problems are not exact and the data measurements themselves contain errors, the variance terms between forward model outputs and observed data are included in the collection of parameters for which there is uncertainty. A sample space is defined that consists of vectors whose components are transfer coefficients, initial conditions, flux terms, and variance weights.

A joint probability density function (pdf) is constructed on the sample space. This pdf conceptually represents our knowledge (or uncertainty) of the system given the data and the model. The joint pdf is used to determine marginal cumulative distribution functions (cdfs) for transfer coefficients. Hence, the uncertainty in the initial conditions and flux terms is marginalized. The effectiveness of the data can also be assessed by comparing the ratio between likelihood intervals in which the length of a 95% likelihood parameter interval obtained from the pdf constrained by observations and that of the a priori pdf in which only a priori parameter bounds are imposed. The joint pdf is also used to calculate functions of the parameters. For example, we make predictions of future pool sizes based on the model and data. In this work we use data from 5 years of observations to make predictions of likely pools sizes 5 years after observation period. We also compare the predictions with those be obtained based on only a priori information unconstrained by data as a further indicator of the information added from observations.

In Section 2 the mathematical model and its approximation are discussed. Also, the time series of the temperature and moisture environmental effects as well as the carbon flux are introduced. The trend functions used in future predictions and the cdfs of the deviations from those trends are indicated. After introducing the data and the observational operators in Section 3 the functional form of the joint probability density function is obtained. The use of the joint pdf to invert the observed data to obtain distribution and likelihood

intervals of parameters and random variables is discussed. We also assess the effectiveness of data in the estimation of various parameters. In Section 4 we discuss sampling and the dependence of results on the number of simulations used in evaluating the joint pdf as well as reporting results based on the data sets we use.

2. Underlying system and approximation

In this section we present the underlying model and the finite difference approximations used for its numerical approximation. In the subsequent application, the numerical model will be used to generate synthetic state vectors. The state, $\mathbf{x}(t)$, is a vector-valued function of time each component of which gives the quantity of carbon in a particular pool. In this work the compartmental model yields a state vector, $\mathbf{x}(t)$, consisting of seven components corresponding to the carbon pools: nonwoody biomass, woody biomass, metabolic litter, structural litter, microbes, slow organic matter (SOM), and passive SOM. Synthetic data are obtained from the state by means of observation operators that enable us to compare the synthetic data with measurement from real observations.

The underlying model is a system of seven differential equations with initial conditions given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x}(t) = \xi(t)AC\mathbf{x}(t) + \mathbf{b}u(t), \tag{1}$$

$$\mathbf{x}(0) = \mathbf{x}_0.$$

The matrix C is a 7×7 diagonal matrix whose entries consist of the components of the vector $\boldsymbol{c}$

$$C = \mathrm{diag}(\mathbf{c}).$$

The vector $\boldsymbol{c}$ represents fractions of C left in their own pools after each time step. The 7×7 matrix A gives interaction weights among the various pools with

$$A = \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0.712 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0.288 & 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0.45 & 0.275 & -1 & 0.42 & 0.45 \\ 0 & 0 & 0 & 0.275 & 0.296 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0.004 & 0.03 & -1 \end{pmatrix}.$$

The nonzero elements in the matrix A are coefficients that partition exit carbon into other pools (see [2] for a detailed explanation). The column vector $\boldsymbol{b}$ has partitioning coefficients of carbon influx function $u(t)$ into different plant pools. The scalar environmental function $\xi(t)$ takes into account dependence of carbon fluxes on temperature and moisture related functions $\tau(t)$ and $\mu(t)$ and is expressed as the product

$$\xi(t) = \tau(t) \quad \mu(t).$$

The solution vector $\boldsymbol{x}(t)$ is a column 7-vector giving carbon pool sizes as a function of time.

The system (1) is derived from ecological observation of ecosystem carbon processes see [2]. Carbon inputs into the system through photosynthesis as represented by $u(t)$, is partitioned into different plant parts as represented by the vector $\boldsymbol{b}$, and then transferred among pools, $\boldsymbol{x}(t)$, as described by AC. From the perspective of this work we consider the vectors $\boldsymbol{c}$, $\boldsymbol{b}$, and $\boldsymbol{x}_0$ to be parameters of the problem. Actually the vector $\boldsymbol{c}$ is of primary importance for our estimations here. Nevertheless the vectors $\boldsymbol{b}$ and $\boldsymbol{x}_0$ may not be well known, and we wish to include this uncertainty in our considerations. We thus define the vector $\boldsymbol{q} = (\boldsymbol{c},\boldsymbol{b},\boldsymbol{x}_0)$. When the parameter vector $\boldsymbol{q}$ is specified, the associated state function at time $\boldsymbol{x}(\boldsymbol{q})(t)$ may be determined.

The real-valued functions $u(t)$, $\tau(t)$, and $\mu(t)$ are functions obtained through observation of time series over a given time period (5 years). As such, there appears to be considerable noise in the functions. For inversion of the data we use these functions directly. However, for prediction we use approximating trend functions that may be extrapolated into the future (say an additional 5 years). To include the uncertainty of the measurements of C flux, temperature, and moisture, deviations from the trends during the observational interval are used to generate deviations from the trends in the future. In Figs. 1–3 are the graphs to show variations in temperature, moisture and flux for a 5-year periods. Also, are included trend curves that are smooth approximations of the respective time series. The deviations of the time series from the trends are expressed in terms of the cumulative density functions obtained from the histograms of the deviations over the 5-year period. These are presented in Figs. 4 and 5. Explicit descriptions of the trend functions will be given in the case study in Section 4.

To study the problem, a finite difference approximation of the initial value problem (1) is introduced in [6]. Thus, we consider the following system of difference equations. Setting

$$\mathbf{x}_j = \mathbf{x}(\mathbf{q})(t_j), \quad \xi_j = \xi(t_j) \quad \text{and} \quad u_j = u(t_j)$$

the difference approximation to (1) that we use is given by

$$(\mathbf{x}_{j+1} - \mathbf{x}_j)/\Delta t = AC(\xi_{j+1}\mathbf{x}_{j+1} + \xi_j\mathbf{x}_j)/2 + \mathbf{b}(u_{j+1} + u_j)/2.$$

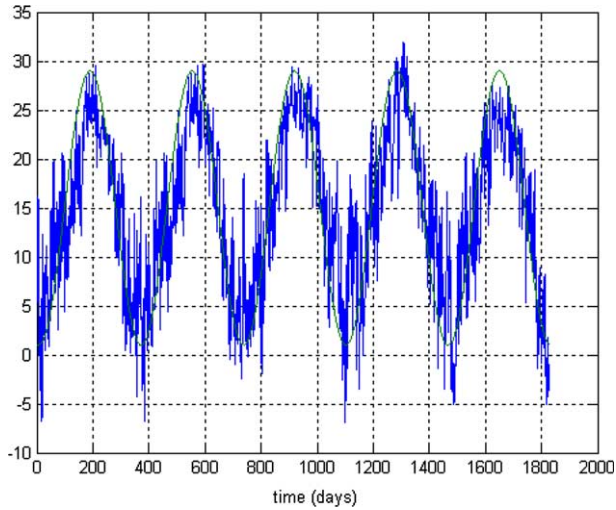

Fig. 1. Comparison of temperature time series with trend in degrees Centigrade.

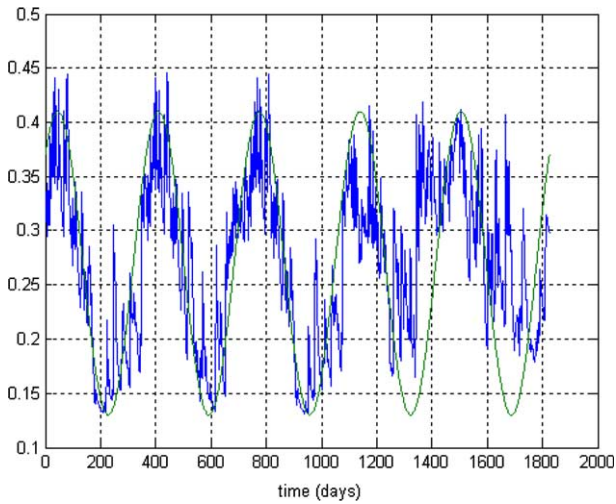

Fig. 2. Comparison of moisture time series with trend in volumetric moisture content.

Combining terms we find that

$$[I - \Delta t\xi_{j+1}AC/2]\mathbf{x}_{j+1} = [I + \Delta t\xi_j AC/2]\mathbf{x}_j + \Delta t\mathbf{b}(u_{j+1} + u_j)/2$$

for $j=0,1,\ldots,NT-1$. Set

$$B_j = I - \Delta t\xi_j AC/2,$$

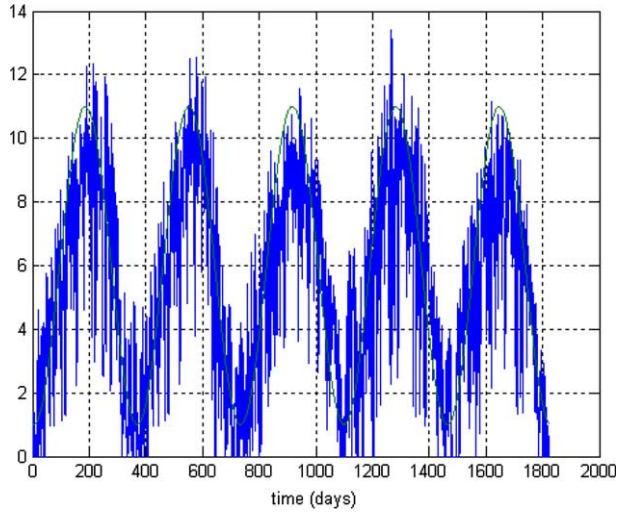

Fig. 3. Comparison of CO_2 influx function with trend in gC/(m^2 day).

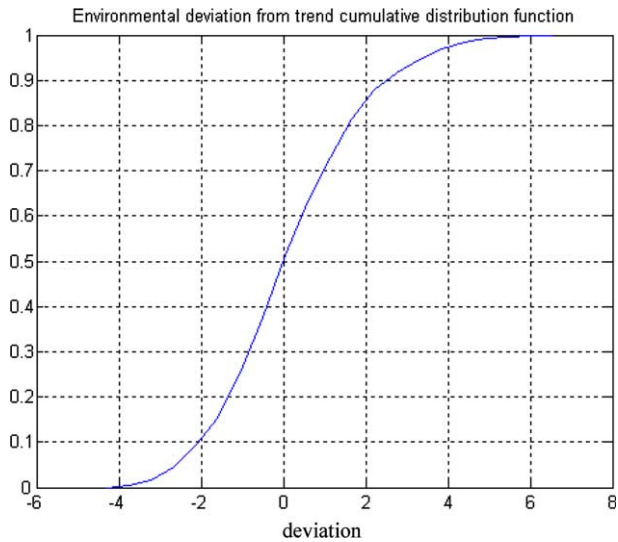

Fig. 4. CDF of the deviation from of the environmental time-series from trend over a 5-year record.

$$B'_j = I + \Delta t \xi_j AC/2$$

and

$$\mathbf{f}_j = \Delta t \mathbf{b}(u_{j+1} + u_j)/2.$$

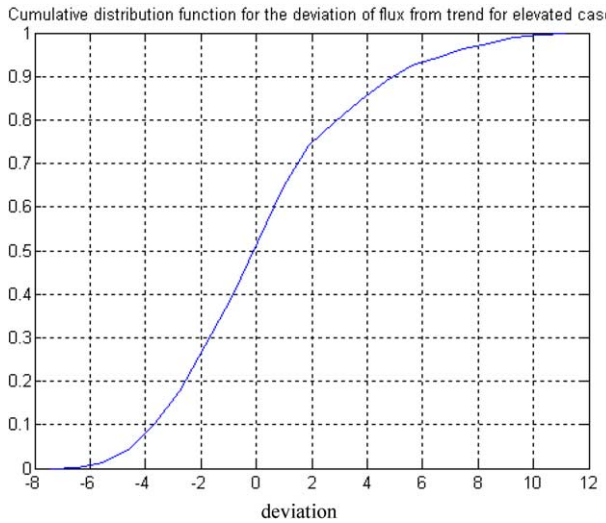

Fig. 5. CDF of the deviation of the flux time series from trend over a 5-year record.

It follows that the system of difference equations is

$$\begin{aligned} B_1\mathbf{x}_1 &= B_0'\mathbf{x}_0 + \mathbf{f}_0, \\ B_2\mathbf{x}_2 &= B_1'\mathbf{x}_1 + \mathbf{f}_1, \\ &\cdots \\ B_{\mathrm{NT}}\mathbf{x}_{\mathrm{NT}} &= B_{\mathrm{NT}-1}'\mathbf{x}_{\mathrm{NT}-1} + \mathbf{f}_{\mathrm{NT}-1}. \end{aligned} \tag{2}$$

Define the column vectors of length 7(NT)

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{x}_{\mathrm{NT}} \end{pmatrix},$$

$$\mathbf{F} = \begin{pmatrix} \mathbf{f}_0 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{f}_{\mathrm{NT}-1} \end{pmatrix}$$

and

$$\mathbf{F_0} = \begin{pmatrix} B_0\mathbf{x}_1 \\ \mathbf{0} \\ . \\ . \\ . \\ \mathbf{0} \end{pmatrix},$$

and the 7(NT)×7(NT) matrix

$$\underline{B}(\mathbf{c}) = \begin{pmatrix} B_1 & 0 & \cdots & & 0 \\ -B_1' & B_2 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & -B_{\mathrm{NT}-1}' & B_{\mathrm{NT}} \end{pmatrix}.$$

The approximating system is given by

$$\underline{B}(\mathbf{c})\mathbf{X}(\mathbf{q}) = \mathbf{F}(\mathbf{b}) + \mathbf{F}_0(\mathbf{x}_0). \tag{3}$$

Thus, given a transfer coefficient $\mathbf{c}$, the matrix $\underline{B}(\boldsymbol{x}(t))$ is defined. Assuming invertibility of $\underline{B}(\boldsymbol{c})$, we may solve the above equation for $\boldsymbol{X}(\boldsymbol{q})$. The mapping $\mathbf{q} \to \mathbf{X}(\mathbf{q})$ is defined from the prescribed set of admissible transfer coefficients Q_{ad} to the state vector $\mathbf{X}(\mathbf{q})$.

The actual solution of the difference equations (2) is, of course, carried out iteratively [4]. The closed form (3) is useful in analysis of the system, for example in stability studies in which the derivative of the state with respect to the parameters ($\mathbf{c}$, $\mathbf{b}$, $\boldsymbol{x}_0$) is needed, but will be studied in subsequent work.

3. Data observation operators and probability density functions

In this section we describe the measurement models. These operators map the model state obtained in the previous section to observable attributes to be compared with observed data. We assume that a parameter vector $\mathbf{q}$ is specified and that the column 7-vector $\boldsymbol{x}_j(\mathbf{q})$ gives the state associated with $\mathbf{q}$ at time t_j. Observation operators are generally of the form

$$\varphi(\mathbf{q})(t) = \Phi(t)\mathbf{c} + \phi(t),$$

where for each t, $\Phi(t)$ is a 7×7 matrix and $\phi(t)$ is a column 7-vector. A synthetic measurement at t_j takes the form

$$z_j(\mathbf{q}) = \varphi(\mathbf{q})(t_j)^{\mathrm{T}}\mathbf{x}_j(\mathbf{q}).$$

We consider the following observation operators to be compared with data.

Soil respiration

$$\Phi = \xi(t)\mathrm{diag}([0.25, 0.25, 0.55, 0.45, 0.7, 0.55, 0.55])$$

and

$$\phi = \mathbf{0}$$

with the observation given by

$$z_j(\mathbf{q}) = \mathbf{c}^{\mathrm{T}}\Phi(t_j)\mathbf{x}_j(\mathbf{q}).$$

Woody biomass

$$\Phi = 0$$

and

$$\phi = [0, 1, 0, 0, 0, 0, 0]^{\mathrm{T}}$$

with the observation given by

$$z_j(\mathbf{q}) = \phi^{\mathrm{T}}\mathbf{x}_j(\mathbf{q}).$$

Litterfall

$$\Phi(t) = \xi(t)\mathrm{diag}([0.75, 0.75, 0, 0, 0, 0, 0])$$

and

$$\phi = \mathbf{0}$$

with the observation given by

$$z_j(\mathbf{q}) = \mathbf{c}^{\mathrm{T}}\Phi(t_j)\mathbf{x}_j(\mathbf{q}).$$

Foliage biomass

$$\Phi = 0$$

and

$$\phi = [0.75, 0, 0, 0, 0, 0, 0]^{\mathrm{T}}$$

with the observation given by

$$z_j(\mathbf{c}) = \phi^{\mathrm{T}}\mathbf{x}_j(\mathbf{q}).$$

Mineral carbon

$$\Phi = 0$$

and

$$\phi = [0, 0, 0, 0, 1, 1, 1]^{\mathrm{T}}$$

with the observation given by

$$z_j(\mathbf{c}) = \phi^{\mathrm{T}}\mathbf{x}_j(\mathbf{q}).$$

Forest floor carbon

$$\Phi = 0$$

and

$$\phi = [0, 0, 0.75, 0.75, 0, 0, 0]^{\mathrm{T}}$$

with the observation given by

$$z_j(\mathbf{c}) = \phi^{\mathrm{T}}\mathbf{x}_j(\mathbf{q}).$$

To compare data we define a fit-to-data functional as follows. For the kth data set above, let

$$J_k(\mathbf{q}) = \varepsilon_k \sum_{i=1}^{\mathrm{NT}k} [\varphi(\mathbf{c})^{\mathrm{T}}\mathbf{x}_j(\mathbf{q}) - -z_i^k]^2,$$

where

$$\varepsilon_k = 1/((\mathrm{NT}k - 1)\mathrm{var}(z^k)).$$

For each k the fit-to-data functional $J_k(\mathbf{q})$ gives an expression of the squared error between the data vector z^k and the corresponding output associated with a given parameter vector $\mathbf{q}$, $\{\boldsymbol{\varphi}^k(\mathbf{c})\ \mathrm{x}_i(\mathbf{q})^{\mathrm{NT}k}{}_{i=1}\}$. Observations occur at NTk times for a given data set. The fit-to-data functional is given by

$$J(q) = \sum_{k=1}^{6} \theta_k J_k(q).$$

Instead of considering a minimization problem, as done in [2], to find that coefficient vector minimizing the fit-to-data function over an admissible set, we take a probabilistic view. Thus, we consider the set of admissible parameters as elements in a sample space $\mathbf{Q}$ consisting of the vector $\boldsymbol{p} = (\boldsymbol{q}, \theta)$ where θ is the vector of coefficients θ_k, $k = 1, \ldots, 6$ in the fit-to-data functional above. From this perspective, we view the weight (or in fact the variance) as part of the model Burnham and Anderson [1]. Thus, the vector of parameters consist of the transfer coefficients, the flux vector, and the vector of initial conditions, $\boldsymbol{q} = (\boldsymbol{c}, \boldsymbol{b}, \boldsymbol{x}_0)$ as well as the vector of weights θ.

We introduce the joint probability density function (pdf)

$$f(\mathbf{p}) = K \exp(-J(\mathbf{p})/2),$$

where the constant K is a normalization constant such that the integral of $f(\mathbf{p})$ over the subset of admissible parameters Q_{ad} of the sample space $\mathbf{Q}$ is one. The

function $f(\mathbf{p})$ is a pdf that carries the information from measurements, the model equation, and a priori bounds in the parameters as well as uncertainties in the data. Without the data the joint pdf $f(\mathbf{p})$ may be regarded as the uniform distribution defined over $\mathbf{Q}_{\text{ad}}$. The data and the model define correlations among the parameters through the error expression in the fit-to-data and joint probability density functions. This constitutes a conjunction of information as discussed in [5] to obtain joint pdf $f(\mathbf{p})$.

From the joint probability density function $f(\mathbf{p})$ we obtain the following.

(1) Quantification of system information based on the model, prior information, and data.
(2) Marginal distributions for individual parameters,

$$f_i(\mathbf{p}_i) = \int_{Q'i} f(\mathbf{p})\,\mathrm{d}\mathbf{p}_i'$$

for the marginal pdf where integration is with respect to all parameters except $\mathbf{p}_i$ and Q'_i designates the parameter space excluding the $\mathbf{p}_i$ variable. The marginal cumulative distribution is given by

$$F_i(s) = \int_{P_i(s)} f(\mathbf{p})\,\mathrm{d}\mathbf{p},$$

where

$$P_i(s) = \{\mathbf{p} : p_i \leqslant s\}.$$

(3) Mean and maximum likelihood estimators are given by

$$p_{i\,\text{mean}} = \int_Q p_i f(\mathbf{p})\,\mathrm{d}\mathbf{p},$$

$$p_{i\,\max} = \max\{f_i(p_i) : p_i \in Q_i\}.$$

(4) Likelihood intervals and bounds are determined as follows. Let $\underline{p}_i$ and $\underline{p}_i$ be defined by

$$F_i(\overline{p}_i) = 0.975 \quad \text{and} \quad F_i(\underline{p}_i) = 0.025$$

as left and right endpoints of a 95% likelihood interval $[\underline{p}_i, p_i]$.

We can use the information contained in the joint pdf obtained above to make predictions of likely future C pool sizes. If γ: $\mathbf{Q}_{\text{ad}} \to \mathbf{R}$ is a reasonable (measurable) real-valued function of the parameters $\mathbf{p}$, the cumulative distribution function is defined by

$$F_\gamma(s) = \text{Prob}\,\mathbf{Q}_{\text{ad}}(s;\gamma),$$

where

$$\mathbf{Q}_{\text{ad}}(\mathbf{s};\theta) = \{p \in \mathbf{Q}_{\text{ad}} : \gamma(\mathbf{p}) \leqslant s\}.$$

Thus, $F_\theta(s)$ is calculated by

$$F_\gamma(s) = \int_{\mathbf{Q}_{\mathrm{ad}}(\mathbf{s};\gamma)} f(\mathbf{p})\,\mathrm{d}\mathbf{p}.$$

Setting γ above to $x_i(T,\boldsymbol{p})$, the size of the ith pool at a future time T, then the cumulative distribution of x_i is obtained by computation of the integral for each value of s

$$F_{xi}(s) = \int_{\mathbf{Q}_{\mathrm{ad}}(\mathbf{s};xi)} x_i(T,\mathbf{p}) f(\mathbf{p})\,\mathrm{d}\mathbf{p}.$$

4. A case study using FACE data

In this section we specify a case study using FACE data. The sample space consists of a bounded set in $\boldsymbol{R}^{21}$ in which the parameters are 21-tuples of the 7-vector of transfer coefficients **c**, the 7-vector of initial conditions $\mathbf{x}_0$, the 2-vector of fluxes **b**, and the weights θ on the variances of the error between the simulated model output and the observations. The vector of weights θ are all bounded between 0 and 2 and the terms multiply the inverse of the variances of the data.

Soil respiration=1.401
Woody biomass=283,000
Litter fall=1150
Foliage biomass=2570
Forest floor carbon=8710
Mineral carbon=36,200

The bounds of admissible transfer coefficients, initial conditions, and flux terms are given in Tables 1–3.

Table 1
Bounds on admissible transfer coefficients

Transfer coefficient	Minimum bound (1/day)	Maximum bound (1/day)
c_1	8.82×10^{-5}	3.9×10^{-3}
c_2	2.74×10^{-5}	5.48×10^{-4}
c_3	2.7395×10^{-3}	5.468×10^{-2}
c_4	2.74×10^{-4}	5.48×10^{-3}
c_5	1.37×10^{-3}	1.7068×10^{-2}
c_6	2.74×10^{-5}	5.48×10^{-4}
c_7	6.85×10^{-7}	1.826×10^{-5}

Table 2
Bounds on admissible initial conditions

Initial condition	Minimum bound (g/m^2)	Maximum bound (g/m^2)
x_1	4.22×10^2	5.159×10^2
x_2	3.69×10^3	4.51×10^3
x_3	5.76×10^1	7.04×10^1
x_4	6.246×10^2	7.5634×10^2
x_5	1.107×10^2	1.353×10^2
x_6	1.2465×10^3	1.235×10^3
x_7	8.307×10^2	1.0153×10^3

Table 3
Bounds on admissible flux terms

Flux	Minimum bound	Maximum bound
b_{o1}	0.225	0.275
b_{o2}	0.27	0.33

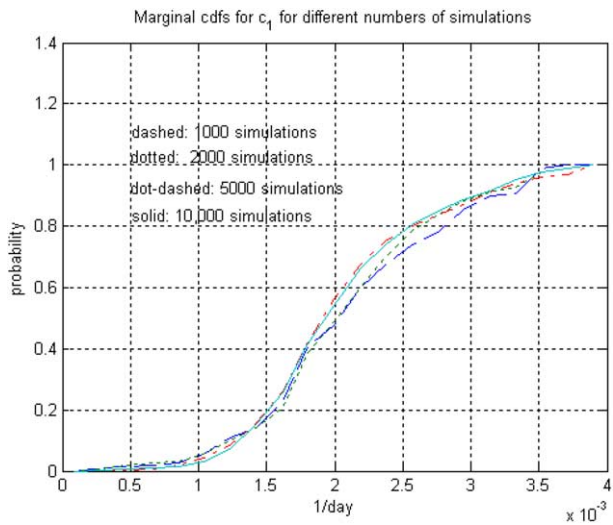

Fig. 6. Comparison of cdfs for c_1 for different numbers of simulations.

To calculate the integrals indicated above requires multiple simulations see [3]. In Fig. 6 is portrayed the curves for different numbers of simulations for the marginal cumulative distributions for the transfer coefficient c_1. Based on the computational experience indicated in Fig. 6 we use 10,000 simulations for in the present study. Moreover, for our computations, we use directly the time

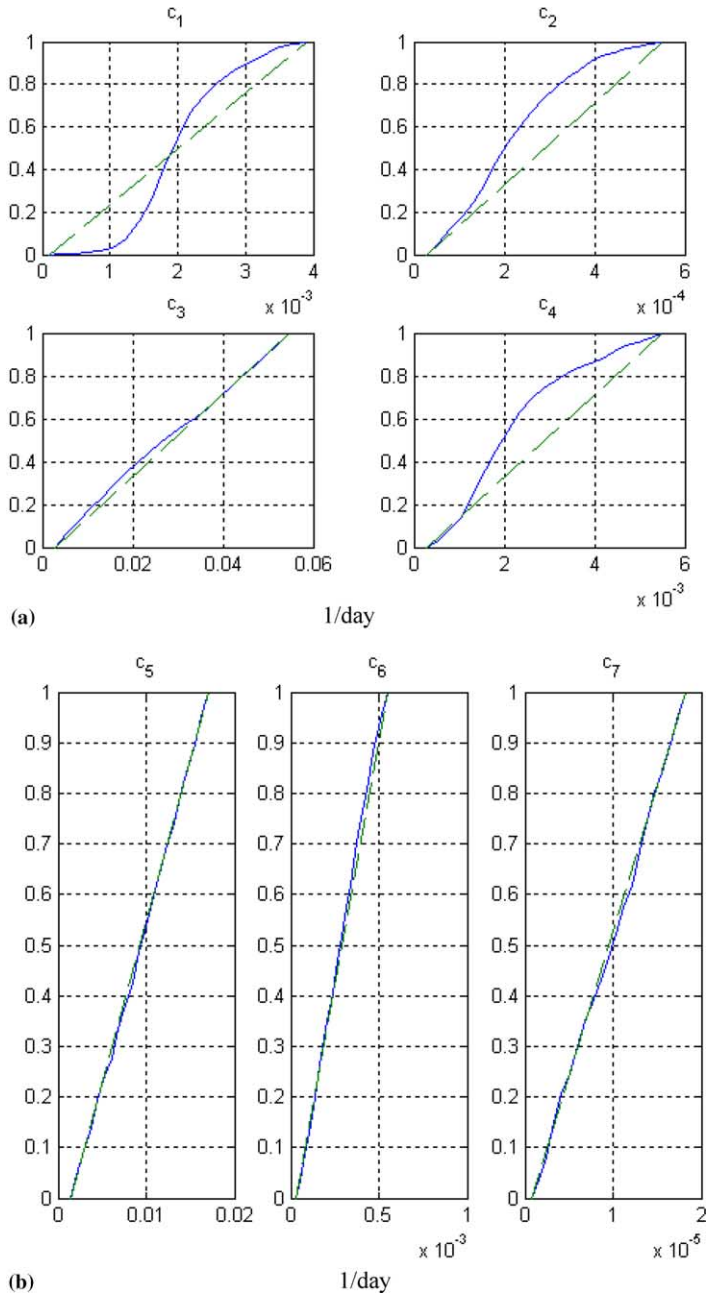

Fig. 7. (a) Marginal cdfs for c_1–c_4 with data (solid) compared with uniform cdfs (dashed). (b) Marginal cdfs for c_5–c_7 with data (solid) compared with uniform cdfs (dashed).

Table 4
Ratios of (a posteriori)/(a priori) 95% likelihood intervals for transfer coefficients with 5 years of data and predicted pools sizes after 10 years

Pool number	c (1/day)	x (g/m^2)
1	0.704	0.228
2	0.914	0.974
3	1.000	1.000
4	0.955	0.694
5	0.997	0.908
6	0.974	0.826
7	0.988	0.991

Table 5
Comparison of data types that reduce likelihood ratio to less than 0.8 for observed data and numerical data

Data type	Transfer coefficient with data with ratio < 0.8	Transfer coefficient with numerical data ratio < 0.8
Soil carbon	c_2, c_4	c_2, c_4, c_6[a]
Foliage biomass	c_1	c_1
Litter fall	–	–
Woody biomass	c_2	c_2
Soil respiration	–	–

[a] c_6 numerical examples used longer observational time periods.

series for temperature, moisture and flux as indicated in Figs. 1–3. Also, we use time step sizes of d$t = 1$ day and use data over 5-year-period.

In Fig. 7a and b are plotted the a posteriori marginal cdfs based on data along with the a priori cdfs obtained from uniform distributions over the parameter interval. As a measure of the information added to the inversion based on the data, we calculated the ratio of the length of the 95% likelihood interval using the a posteriori marginal distributions constrained by data divided by the length of the 95% likelihood interval using the a priori uniform distributions based on the parameter bounds indicated in Tables 1–3. This is presented in the second column of Table 4. We note that the 95% likelihood interval is substantially reduced for the c_1 transfer coefficient. The likelihood interval is also reduced for c_2 and c_4, although not as dramatically as c_1. For the transfer coefficients c_3, c_5, and c_7 there is very little information supplied from the data beyond that of the a priori uniform distribution.

The results portrayed in Fig. 7a and b as well as those indicated in Table 4 are obtained using all data sets. It is an interesting issue to determine the effects of different data sets on the different transfer coefficients. In Table 5 we portray in the second column which coefficients are impacted by which types of data. In the third column are results from [7] from numerical test. These coincide quite

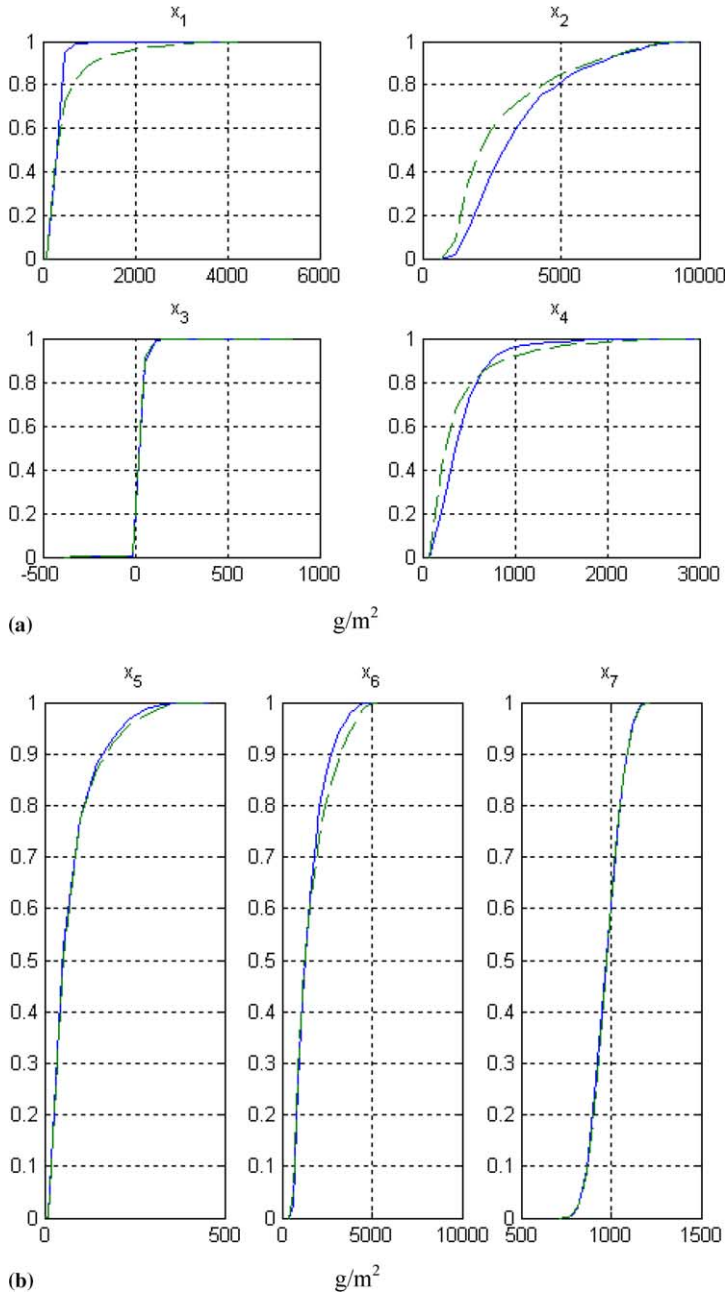

Fig. 8. (a) Cdf of predicted carbon pools x_1–x_4 after 10 years using data (solid) and uniform (dashed) in probability. (b) Cdf of predicted carbon pools x_5–x_7 after 10 years using data (solid) and uniform (dashed) in probability.

well with the results from real data used in this paper. We see that soil carbon, foliage biomass, and wood biomass are very important for estimation of c_1, c_2 and c_4. It appears that litter fall and soil respiration data are of little value.

Another way of measuring the value of data on the inversion process is to observe the spread of the predicted carbon pool sizes under a joint pdf that is obtained under the constraint of data with that obtained with only the a priori parameter bounds. Thus, for predictions to 5 years past the data period, we use the joint a posteriori pdf obtained based on the 5-year data above. It is necessary to extrapolate the temperature, moisture, and flux to the future in order to solve the initial value problem over the extended period. We used trends of the temperature, moisture, and flux time series obtained as a least squares fit to the time series given in Figs. 1–3. We determine

$$\text{Temperature:} \quad \tau(t) = 14.7651 + 14\sin(2\pi(t + 266)/365),$$

$$\text{Moisture:} \quad \mu(t) = 0.27 + 0.14\sin(2\pi(t + 46/365)),$$

and

$$\text{Flux:} \quad u(t) = 6 + 5\sin(2\pi(t + 269)/365).$$

To model uncertainty, we accumulate the deviation of the time series in Figs. 1–3 from these trends. Based on the resulting deviations we construct cdfs of the environmental function, i.e. the product of temperature and moisture and the flux, see Figs. 4 and 5. As the trend values are extrapolated to obtain future values of environmental and flux values, deviations are randomly generated to perturb the trends in an effort to capture future uncertainty. The value of the state vector at the time 10 years is obtained by solving the initial value problem over the time period of 10 years. Time steps of length 30 days are used in order to expedite the solution. The pool size vector obtained at $T = 10$ years is considered to be a random vector defined over the sample space of the parameters.

In Fig. 8a and b are portrayed the cdfs of the predicted pool sizes of different carbon pools. These are compared with the predicted pool sizes using the a priori uniform distributions for parameters before constraining with data. In Table 4 is presented in column 3 the reduction in the length of the 95% likelihood interval of the predicted pool sizes with respect to that obtained with only the a priori uniform distribution on the parameter sample space. We note that prediction ranges are narrowed for pools 1, 4, 5, and 6.

Acknowledgement

This study was financially supported by the Office of Science (BER), US Department of Energy Grant No. DE-FG 03-99ER62800.

References

[1] K.P. Burnham, D.R. Anderson, Model Selection and Multimodel Inference, second ed., Springer, 2002.
[2] Y.L. Luo, L. White, J. Canadell, E. DeLucia, D. Ellsworth, A. Finzi, J. Lichter, W. Schlesinger, Sustainability of terrestrial carbon sequestration: a case study in Duke Forest with an inversion approach. Global Biogeochem. Cycles 17(1) 1021 (2003) 21–34.
[3] H. Niederreiter, Random Number Generation and Quasi-Monte Carlo Methods, SIAM, Philadelphia, PA, 1992.
[4] J. Stoer, R. Bulirsch, Introduction to Numerical Analysis, second ed., Springer, New York, 1991.
[5] A. Tarantola, Inverse Problem Theory, Elsevier, New York, 1987.
[6] L. White, Y. Luo, Estimation of carbon transfer coefficients using Duke Forest free-air CO_2 enrichment data, Appl. Math. Comp. 130 (2002) 101–120.
[7] L. White, Y. Luo, Model-based data assessment for terrestrial carbon processes: implications for sampling strategy in FACE experiments, Appl. Math. Comp., in press.