Ecological Modelling 221 (2010) 1880–1888

Contents lists available at ScienceDirect

Ecological Modelling

journal homepage: www.elsevier.com/locate/ecolmodel

Uncertainty analyses for calibrating a soil carbon balance model to agricultural field trial data in Sweden and Kenya

John Juston[a,*], Olof Andrén[b], Thomas Kätterer[b], Per-Erik Jansson[a]

[a] Department of Land and Water Resources Engineering, Teknikringen 76, KTH Royal Institute of Technology, Stockholm 10044, Sweden
[b] Department of Soil and Environment, SLU, Uppsala, Sweden

ARTICLE INFO

Article history:
Received 13 January 2010
Received in revised form 27 April 2010
Accepted 29 April 2010
Available online 11 June 2010

Keywords:
Soil organic carbon
Soil carbon
Carbon budgets
Model
Modeling
Agriculture
Uncertainty analysis
GLUE
ICBM

ABSTRACT

How do additional data of the same and/or different type contribute to reducing model parameter and predictive uncertainties? Most modeling applications of soil organic carbon (SOC) time series in agricultural field trial datasets have been conducted without accounting for model parameter uncertainty. There have been recent advances with Monte Carlo-based uncertainty analyses in the field of hydrological modeling that are applicable, relevant and potentially valuable in modeling the dynamics of SOC. Here we employed a Monte Carlo method with threshold screening known as Generalized Likelihood Uncertainty Estimation (GLUE) to calibrate the Introductory Carbon Balance Model (ICBM) to long-term field trail data from Ultuna, Sweden and Machang'a, Kenya. Calibration results are presented in terms of parameter distributions and credibility bands on time series simulations for a number of case studies. Using these methods, we demonstrate that widely uncertain model parameters, as well as strong covariance between inert pool size and rate constant parameters, exist when root mean square simulation errors were within uncertainties in input estimations and data observations. We show that even rough estimates of the inert pool (perhaps from chemical analysis) can be quite valuable to reduce uncertainties in model parameters. In fact, such estimates were more effective at reducing parameter and predictive uncertainty than an additional 16 years time series data at Ultuna. We also demonstrate an effective method to jointly, simultaneously and in principle more robustly calibrate model parameters to multiple datasets across different climatic regions within an uncertainty framework. These methods and approaches should have benefits for use with other SOC models and datasets as well.

1. Introduction

Long-term agricultural field trials are necessary tools for assessing how agricultural soil stocks will respond to various management strategies, such as tillage practices as well as fertilizer, crop residue, and/or manure amendments (Nandwa, 2001; Debreczeni and Körschens, 2003; Richter et al., 2007). These experiments take on an added dimension of significance in famine-prone regions, such as much of Africa, where maintenance of soil organic matter is seen as a critical link in maintaining soil fertility and crop productivity (e.g. Nandwa, 2001; Vanlauwe and Giller, 2006; Bationo et al., 2007; Zingore et al., 2008).

Modeling is necessary for extracting information from field trial datasets and applying it to other areas, such as regional scale simulations (e.g. Falloon and Smith, 2002; Guo et al., 2007; Kamoni et al., 2007a). Questions remain, however, regarding model and parameter portability to different climatic regions. For example, several studies (Diels et al., 2004; Traore et al., 2008) have shown that simulations of agricultural field trial data in tropical regions using the RothC model were improved if rate constants were changed from the default originally calibrated in the U.K. (Coleman and Jenkinson, 1996).

It has also been recognized in recent years that we need to increase awareness and abilities to express uncertainties inherent in modeling soil carbon data (e.g., Ogle et al., 2003; O'Neill and Melnikov, 2008; Post et al., 2008). In fact, it has been argued that environmental modeling in general (including soil carbon models) is inherently and unavoidably uncertain due to uncertainties and errors in input and output data, parameter estimation, and model structures (Beven, 2009). Yet, the vast majority of model calibrations and predictions with SOC models are still conducted with optimization methods and deterministic perspectives (e.g. Smith et al., 1997; Andrén and Kätterer, 1997; Diels et al., 2004; Petersen et al., 2005; Lugato et al., 2007; Kamoni et al., 2007b; Guo et al., 2007; Traore et al., 2008).

There have been recent advances with Monte Carlo-based uncertainty analyses in the field of hydrological modeling (e.g. Beven and Freer, 2001; Kuczera and Parent, 1998) that are appli-

* Corresponding author. Tel.: +46 706 674855 (cell).
E-mail address: juston@kth.se (J. Juston).

0304-3800/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.ecolmodel.2010.04.019

cable, relevant and potentially valuable in modeling the dynamics of SOC. These methods are fundamentally different from simpler Monte Carlo error-propagation approaches (e.g. Ogle et al., 2003) as they seek to refine (condition) knowledge of parameter uncertainty based on the information content in the available data and criteria defined by the user of the model. Outputs from these methods provide conditioned (posterior) distributions of parameter values plus uncertainty bounds on simulated dynamics. There are very few examples in the current literature of these methods applied to soil carbon studies (e.g. Xenakis et al., 2008).

This paper addresses the questions:

- What types of insights can be gained by calibrating a soil carbon model with uncertainty analysis rather than deterministically?
- Can an uncertainty framework provide insights for model portability from study sites in cold temperate Sweden to tropical Kenya?

We employed the Introductory Carbon Balance Model (ICBM) for this investigation, and use the original dataset and model parameterization (Andrén and Kätterer, 1997) as a jumping off point for alternate analyses. New calibrations are performed to extended time series from the long-term field trial in Uppsala, Sweden (Persson and Kirchmann, 1994) using a Monte Carlo method with threshold screening known as Generalized Likelihood Uncertainty Estimation (GLUE) (Beven and Freer, 2001). We introduce an additional 16 years of data for model calibration compared to the original parameterization. Calibration results are presented in terms of parameter distributions and credibility bands on time series simulations. The same model and calibration approach were then applied to a second dataset from a long-term trial in Machang'a Kenya (Kihanda et al., 2005). A third calibration was then performed investigating joint and simultaneous calibration of model parameters to both datasets. This last case study builds on work by Petersen et al. (2005), wherein a simultaneous calibration to multiple datasets was conducted within an optimization framework.

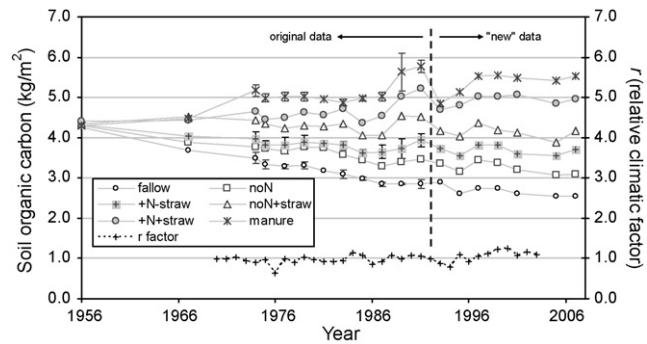

Fig. 1. Time series of SOC carbon storage for six treatments of the SLU-Ultuna field trial in Sweden. Each data point represents the mean of four replicates plots. Standard deviations of replicate data are shown for the fallow, +N-straw, and manure treatments. Soil depth was 20 cm and average bulk density approximately 1.35 g/cm^3. Calculated annual values for the time-variant climate factor, r, are along the bottom.

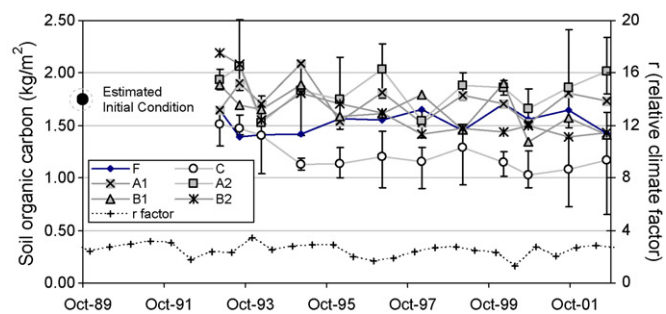

Fig. 2. Time histories of SOC storage for six treatments from the Machang'a soil fertility experiment in Kenya. Data represent the mean of three replicate plots for each treatment. The range of minimum and maximum observations amongst replicates is shown for the C (control) and A2 (10 t/year manure amendment) treatments. Soil depth was 20 cm and average bulk density approximately 1.28 g/cm^3. Calculated seasonal values for the time-variant climate factor, *r*, are along the bottom.

2. Methods

2.1. Study sites

The "Frame trial" is situated at the Swedish University of Agricultural Sciences (SLU-Ultuna) in Uppsala, Sweden. Plots are 4 m^2 in a randomized block design consisting of 15 treatments with four replicates each (Persson and Kirchmann, 1994). Soil was classified as Eutric Cambisol with 37% clay, 41% silt, 23% sand. Average annual temperature at the site is 5.4 °C and average annual precipitation is 520 mm/year. We selected six of the 15 experimental treatments from the SLU-Ultuna Frame Trial for modeling (Table 1). This selection was consistent with prior modeling by Andrén and Kätterer (1997). Soil carbon inputs varied among treatments and consisted of biennial amendments of straw or farmyard manure in addition to below-ground contributions from root production (Table 1). A rather unique aspect of the Ultuna trials is that carbon content in biennial amendments are precisely measured and controlled across replicates and treatments. Carbon inputs from root production are not directly measured but have been previously estimated with allometric functions (Andrén and Kätterer, 1997). Nonetheless, response variations across replicates are generally representative of stochasticity and uncertainty in responses and observations, as input variation is for the most part controlled for (i.e., minimal). Eighteen measurements of soil carbon content (%) in the upper 20 cm for the six selected treatments were available as cross-replicate means from 1956 to 2007. Time series of soil carbon storage (kg/m^2) in the upper 20 cm (Fig. 1) were estimated from carbon content using an interpolated time series of available bulk density data. The initial organic carbon content in the upper 20 cm was about 1.5% (Persson and Kirchmann, 1994), resulting in an initial condition of approximately 4.3 kg/m^2 total storage used in model simulations. Data on the variability (standard deviation) of soil carbon content observations between treatment replicates were available for 1971–1991 (Fig. 1). The mean standard deviation in replicate responses had no trend with the mean magnitude of measured SOC and was approximately 0.14 kg/m^2. Daily meteorological data were available from the site from 1970 to 2004. The SLU-Ultuna Frame Trial data has also been used for model validation in several other studies using RothC (Falloon and Smith, 2002), Century (Paustian et al., 1992; Falloon and Smith, 2002), Q-model (Hyvönen et al., 1996; Nilsson et al., 2005) and CN-SIM (Petersen et al., 2005) models.

The field trial in Kenya was located approximately 200 km northeast of Nairobi outside the town of Machang'a. Plots were 5 m^2 in a randomized block design consisting of nine treatments with three replicates each (Kihanda et al., 2005). Soils were classified as Chrombic Cambisol with 57% clay, 13% silt, and 31% sand. Average annual temperature at the site was 22.8 °C and average annual precipitation was 789 mm/year. We selected six of the nine treatments at Machang'a for modeling (Table 1, Fig. 2). There were two cropping seasons at Machang'a (typical to the region). Soil carbon amendments were made in select treatments (Table 1) once per year for manure (September) and twice per year from crop residues of the previous growing season (Kihanda et al., 2005). Biannual time series of dry-weight crop residues and grain yields

Table 1
Treatments from the Ultuna and Machang'a long-term agricultural field trails that were used for modeling. Treatment names are consistent with those used in source materials (Andrén and Kätterer, 1997; Kihanda et al., 2005). Carbon inputs used in modeling are summarized as long-term average annual values for crop residue, manure and below-ground sources. Treatments B1 and B2 at Machang'a received manure supplementation for the first 3 years, but none thereafter.

Site	Treatment name	Averaged carbon inputs to soil (kg/m^2/year)		
		Crop residue additions	Manure additions	Below-ground contributions
Ultuna	Fallow	0	0	0
	+N + straw	0.19	0	0.095
	noN + straw	0.19	0	0.058
	noN	0.0	0	0.057
	+N − straw	0.0	0	0.091
	Manure	0.0	0.19	0.092
Machang'a	A1	0.16	0.15	0.060
	A2	0.17	0.31	0.067
	B1	0.10	0.15/0.0	0.038
	B2	0.10	0.31/0.0	0.040
	C	0.052	0	0.017
	F	0.11	0	0.040

were available. Crop residue inputs at Machang'a were not standardized as they were at the SLU-Ultuna site and therefore varied from season-to-season and year-to-year. This represents a more typical practice for long-term trials. Carbon concentration in crop residues at Machang'a was in average 36% (a comparatively low value) measured during a 3-year interval (Kihanda et al., 2005). This value was applied to the full time series of dry-weight residues to estimate carbon inputs for modeling. Below-ground contributions from root biomass and exudates were estimated as 25% of carbon in total above ground storage (stover plus grain yields) based on an average value from Bolinder et al. (2007), and then multiplied by 0.7 as an approximation of the portion of total below-ground contributions to the upper 20 cm of soil. The net effect of accounting for below-ground contributions was approximate 36% additional carbon input compared to crop residues alone. There are uncertainties in this estimate of below-ground contributions, but these were not unaccounted for in our analysis. Input time series were then synchronized to the available soil carbon data, such that soil observations corresponded to end-of-season status. The experiment was initiated in 1989, however carbon content in the soil was not measured until 1993, after which it was sampled at 6–12 months intervals. Twelve measurements of soil carbon in the upper 20 cm were available for each replicate from 1993 to 2002 (Fig. 2). A plausible SOC initial condition at 1989 has been back extrapolated (0.68%) by Kihanda et al. (2005) and has been used as a starting value in several other modeling studies (Kamoni et al., 2007b; Micheni et al., 2004). Soil bulk density data at Machang'a were limited and were assumed constant at 1.28 g/cm^3 (Micheni et al., 2004), resulting in an initial soil carbon storage used for model simulations of 1.75 kg/m^2 for all treatments. Daily meteorological data were also available from the site for 1989–2002.

2.2. Modeling approach

We extended the model structure of the two-storage Introductory Carbon Balance Model (ICBM) by adding an inert pool (Fig. 3). Here, we define "inert" as essentially stationary over the duration of the simulation period. We treat this as a model parameter that is determined with calibration, and not specified a priori. Additionally, the model had four other calibration parameters that determined rate constants for model fluxes (Table 2). These five model parameters were estimated for each calibration in this study using methods for uncertainty analyses described below.

The scale factor, r, in model flux equations (Fig. 3) is a relative index of climate-dependent decomposer activity. It is intended to provide model portability to differing climatic regions. We calculated r off-line using daily-step soil water balance and temperature estimations, coupled to equations relating these soil climate parameters to net biological activity as described in Andrén et al. (2007). Daily time series results for r were averaged over annual or bi-annual intervals, depending on the simulation dataset, such that each simulation time interval had a characteristic activity factor based on the available climatic data. Following the approach of Andrén et al. (2007), r was treated as a relative measure of decomposer activity where the SLU-Ultuna site in Sweden was taken as the reference condition. By definition then, the long-term average r value for SLU-Ultuna was normalized to 1.0, but there was inter-annual variability ($\sigma = 0.12$, Fig. 1). Calculations for Machang'a were also normalized to the Swedish site and had a 13-year average value of $r = 2.5$, also with both annual and seasonal variability ($\sigma = 0.54$ and 0.39, respectively; Fig. 2). In general, the higher value at Machang'a suggested approximately 2.5 times higher rate of decomposition of SOC at Machang'a than at Ultuna.

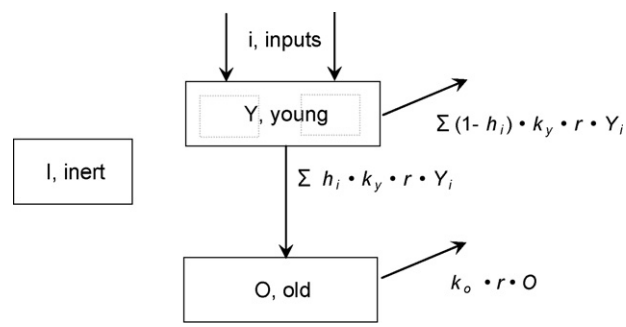

Fig. 3. The ICBM model structure, as used here. Carbon inputs were categorized as either crop residues or manure. The net young pool, Y, contained separate storages for the two input types, each with a humification factor, h_i. A fraction of the initial soil organic carbon content was assumed effectively inert over the simulation period. Fluxes out of Y and O pools (diagonally upward arrows) are to the atmosphere.

The governing differential equations for ICBM (see Appendix A) were coded as finite difference equations. Simulations were conducted at time steps appropriate to the available data at each site. The modeling time step was annual for the Ultuna data, and seasonal (two-per-year: March–August and September–February) for the Machang'a data. Initial conditions for total soil carbon were treated as fixed constants in all simulations with values as discussed above for each study site. These initial total carbon stocks were apportioned to initialize O and Y pools in the ICBM structure (Fig. 3) using steady-state storage ratios from closed-form solutions of the ICBM equations (Appendix A) after the inert fraction in each simulation (the model parameter I) was subtracted.

Five calibration case studies were investigated (Table 3). All were conducted with uncertainty analyses as described in the next

Table 2
Summary of model parameters for calibration and their feasible ranges.

Symbol	Name	Units	Feasible range
k_Y	Rate constant for young carbon	year^{-1}	0.4–1.0
k_O	Rate constant for old carbon	year^{-1}	0.001–0.1
h_C	Humification coeff. for crop residues		0.0–1.0
h_M	Humification coeff. for manure	–	0.0–1.0
I	Inert SOC storage	kg/m^2	0.0–4.5 (Ultuna) 0.0–1.7 (Machang'a)

Table 3
Summary of calibration case studies, ranges of tested GLUE acceptance criterion, and minimum identified RMSE for each case study from all Monte Carlo realizations.

Dataset	Case	Data interval	Calibration parameters	RMSE criterion (g/m^2)	Min RMSE (g/m^2)
SLU Frame	1	1956–1991	5	0.18–0.28	0.167
	2	1956–2007	5	0.19–0.30	0.184
	3	1956–2007	4 (I fixed)	0.19	0.184
Machang'a	4	1989–2002	5	0.21–0.28	0.202
Combined	5	1956–2007 (SLU) 1989–2002 (Mach.)	7	0.21	0.193

section. The first case study revisited the original ICBM calibration to the SLU Frame Trial data (Andrén and Kätterer, 1997). The second case study examined the information value (in the context of parameter identification for ICBM) of an additional 16-year data from the Frame Trial. The third case study examined the impact of prior knowledge, here treated as uncertain knowledge, in the size of the inert pool, I, on ICBM parameter identification to the extended Frame Trial data. The fourth case study examined parametric uncertainty for model calibration with the Machang'a field trial data. The last case study examined how information in both datasets could be used to simultaneously constrain model parameters. Here, we allowed for seven degrees of freedom for calibration: k_Y, k_O, h_C were treated as joint parameters between datasets, while h_M and I were treated as site-specific parameters. It is reasonable to expect that inert storage could be quite different amongst regions (Falloon et al., 1998). The humification factor, h_M, was also treated as site-specific based on evidence that the two manure treatments at Machang'a did not produce significantly different response in SOC (Kihanda et al., 2005), which we judged as an unusual and difficult to explain response.

2.3. Calibration and uncertainty analysis

The uncertainty analysis was the Generalized Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992; Beven and Freer, 2001; Beven, 2009). Generally speaking, GLUE is a non-statistical Monte Carlo procedure that utilizes threshold screening. Plausible prior boundaries were defined for ICBM model parameters (Table 2). These prior ranges allowed the Monte Carlo procedure to search turnover times of 1–2.5 years for the Y pool and 10–1000 years for the O pool; the full fractional range of 0–1 for h_C and h_M parameters; and a range of inert carbon storage from zero to just below the initial condition at each site. Multiple random realizations of parameter sets were drawn within these boundaries. ICBM was run for each realization and time series responses in all treatments were simulated based on initial conditions and the parameter set. A performance measure was calculated to compare simulated SOC time series to data observations. Here, we utilized the root mean square error (RMSE) of simulated and observed values. GLUE is flexible in the choice of performance measure and it is important to recognize that the use of other performance indices, such as a model efficiency measure (Nash and Sutcliffe, 1970) or limits of acceptability approach (Liu et al., 2009), could yield different insights and even slightly different results than presented here. A user-defined threshold was applied based on RMSE scores to retain simulations and parameter sets for post-processing that had RMSE scores below the threshold criteria. Simulations with RMSE scores above the threshold were discarded. Post-processing included calculating distribution functions for the accepted parameter sets and calculating upper and lower boundaries on sets of time series simulations.

The parameter space of each case study (Table 3) was explored with a series of GLUE calibrations, each conducted with threshold criteria starting just above the lowest achieved RMSE for each case study and then incrementing in steps of 0.01 kg/m^2. Note that this increment (0.01 kg/m^2) is between 2 and 15% of annual C loads (crop residues plus manure) applied at either experiment (Table 1), and is less than 1% of SOC store in the upper 20 cm of either of these experiments (Figs. 1 and 2). The resulting sequence of parameter distributions and time series boundaries provided insights into the shapes and characteristics of the ICBM parameter space as conditioned by these data. The number of Monte Carlo runs for each calibration was limited to either 50 million total, or $N = 250$ in the set of accepted model runs.

3. Results

3.1. Frame trial calibrations (Cases 1 and 2)

Uncertainty bounds for time series simulations of the Frame Trial data using all data (Case 2) were different depending of threshold used and treatment considered (Fig. 4). When a smaller part of the data set was used (Case 1) the results were in general within or quite close to the bounds of the Case 2 RMSE < 0.19 kg/m^2 simulations, with the exception of the FYM treatment (Fig. 4). For this treatment the 16-year additional time series data seemed to provide additional constraint on simulations and model parameters (discussed more below).

Zigzag patterns in some treatment responses resulted from simulation of biennial carbon amendments in those treatments (Table 1). Uncertainty bounds decreased with more selective (i.e. lower RMSE thresholds) GLUE criteria (Fig. 5). Approximately 90% of the 1956–2007 data ($n = 94$) were covered by GLUE uncertainty bounds using an RMSE < 0.27 kg/m^2 acceptance threshold, (Figs. 4 and 5). This acceptance threshold ($T < 0.27$ kg/m^2) was greater than the mean uncertainty in the observed data between treatment replicates (0.14 kg/m^2), suggesting contributions to predictive uncertainty beyond stochasticity in data observations from

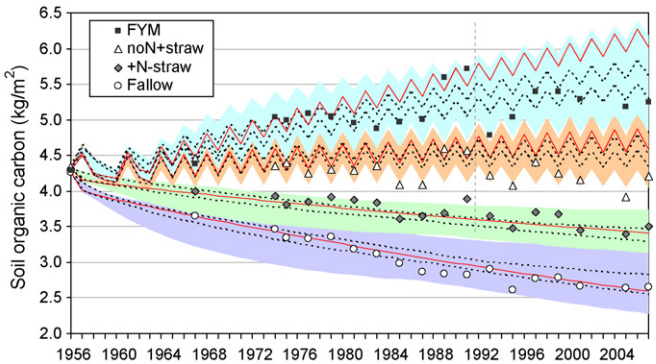

Fig. 4. Time series simulations of the SLU Frame Trial data with uncertainty bounds. For clarity, responses for only four of six treatments are shown. The four shaded bands indicate respective uncertainty intervals for 90% data coverage (RMSE < 0.27 kg/m^2) from the 1956–2007 Case 2 calibration. The dashed line pairs within these shaded bands indicate tightened uncertainty bounds for respective treatments with the GLUE acceptance threshold set at RMSE < 0.19 kg/m^2. Red lines indicate the best observed simulation for the 1956–1991 Case 1 calibration, including forward prediction (1992–2007). Uncertainty bounds for the Case 1 calibration are not shown for clarity, but were approximately as wide as the Case 2 bounds.

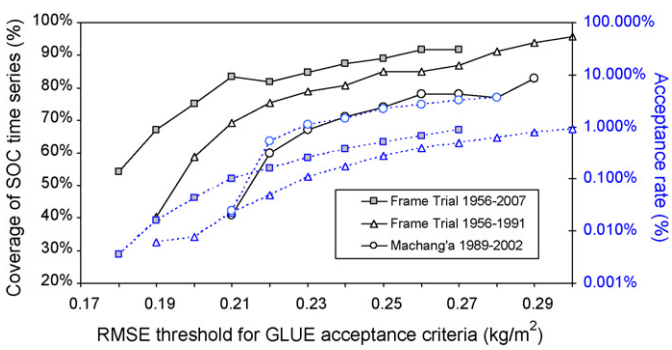

Fig. 5. Coverage of SOC data by GLUE uncertainty bounds as a function of RMSE thresholds for Cases 1, 2, and 4 calibrations, and acceptance rate of retained simulations and parameter sets (secondary *y*-axis). The blue dotted lines refer to the secondary *y*-axis, where symbols used are the same as in the legend.

model structure and/or input data. After 51 years, simulated predictive bands for 90% data coverage ranged between 0.6 and 1.3 kg/m^2 in width. Thus, a prediction of a subsequent year's value of SOC (e.g. 2008) could be estimated with this uncertainty.

An acceptance threshold of RMSE < 0.19 kg/m^2 was within 0.006 kg/m^2 of the lowest identified error in any Case 2 simulations (Table 3). The uncertainty bands at this threshold were indeed narrower than at RMSE < 0.27, but still demonstrated approximately 0.25 kg/m^2 of predictive uncertainty after 51 years simulation (Fig. 4). A smaller uncertainty band such as this would be more appropriate to use to predict the central tendency of long-term time series predictions, with the knowledge that most annual predictions would likely lie outside these bounds.

Residuals from ICBM simulations of the Frame Trial data were not random in structure (Fig. 6). Most notably, simulation of three treatments (i.e. noN, +N-straw, noN + straw) exhibited mean bias on the order of 0.10–0.14 kg/m^2. Autocorrelation in residuals was generally small (R^2 < 0.14). However, there were some notable commonalities in residual structure amongst treatments. In particular, residuals in the noN and +N-straw treatments were highly correlated (R^2 = 0.62) and most treatments shared a similar cyclic pattern in residuals during 1989, 1991, and 1993. Uncertainty estimation with GLUE is robust to complexities such as these in residuals, as the method is fundamentally non-statistical. But these structures could impact the use of alternative methods for uncertainty

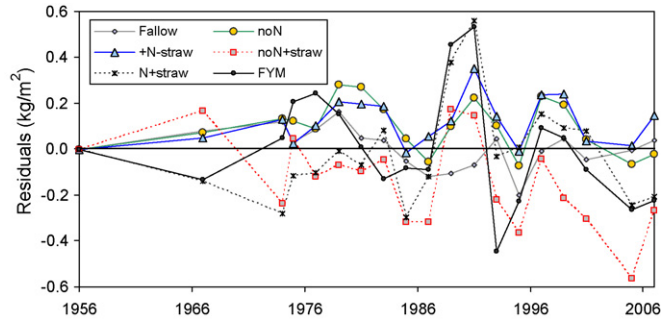

Fig. 6. Time series of model residuals from the Case 2 calibration with lowest RMSE. Average residuals were approximately 0 kg/m^2 in Fallow, +N + straw, and FYM treatments, and 0.10, 0.12, and −0.14 kg/m^2 in noN, +N + straw, and noN + straw treatments, respectively.

Table 4
Covariance structure of ICBM parameters expressed as correlation coefficients (R) from Case 2 (T < 0.19) and Case 3 (T < 0.19) GLUE calibrations. Cells to the upper right of the diagonal show Case 1 covariance; cells to the lower left show Case 3.

R	k_Y	k_O	h_C	h_M	I
k_Y	1.0	0.05	0.58	0.37	0.00
k_O	0.16	1.0	0.78	0.89	0.92
h_C	0.78	0.57	1.0	0.88	0.72
h_M	0.64	0.59	0.77	1.0	0.63
I	0.14	0.87	0.35	0.42	1.0

analysis, which we will discuss later.

Minor differences were evident in conditioned probabilities for ICBM parameters between Cases 1 and 2 calibrations (Fig. 7). Distributions for k_Y, k_O, I, and h_C parameters were largely similar between Cases 1 and 2 across a range of RMSE acceptance thresholds (Fig. 7a–d). The manure humification factor, h_M, seemed to be the parameter most clearly affected by the additional data in the Case 2 calibration. The lower range for h_M in Case 2 calibrations compared to Case 1 (Fig. 7e) were consistent with the lower range of simulated SOC values when ICBM was calibrated to the extended time series (Fig. 4). There was also substantial covariance exhibited between k_O, I, h_C, and h_M parameters in the accepted parameter sets (Table 4). The strongest covariant relationship in the Case 2 calibration existed between k_O and I (R = 0.92). In general, covariant relationships were more defined (in addition to sharper parameter distributions) with more selective threshold criteria (Fig. 7f).

In general, there was high uncertainty in parameter distributions, even at the lowest ranges of accepted RMSE scores. For instance, the covariance of k_O and I from the most selective Case 2 calibrations (RMSE < 0.19 kg/m^2) covered a factor-of-four range in k_O and a plausible range of 0–60% for the inert fraction of SOC at the site (Fig. 7f). Recall that the Case 2 calibration assumed no prior information was available on the inert SOC at the site. Also recall that the RMSE < 0.19 kg/m^2 criteria was near optimal—just 0.006 kg/m^2 greater than the lowest achieved RMSE with any of our Case 2 Monte Carlo realizations (Table 3). This very small differential is arguably well within uncertainties in the SOC data itself (Fig. 1), as well as the uncertainties in carbon inputs (Table 1). Also note that the k_Y parameter was not well defined in any case study (Fig. 7a), suggesting low information value in the multi-year interval data to support calibration of turnover rates in the Young pool in the ICBM.

3.2. Hypothetical impact of supplemental data (Case 3)

If information were available to bound prior knowledge of the inert SOC content of the Frame Trial soils, then the uncertainty in the remaining ICBM model parameters could be reduced. As a

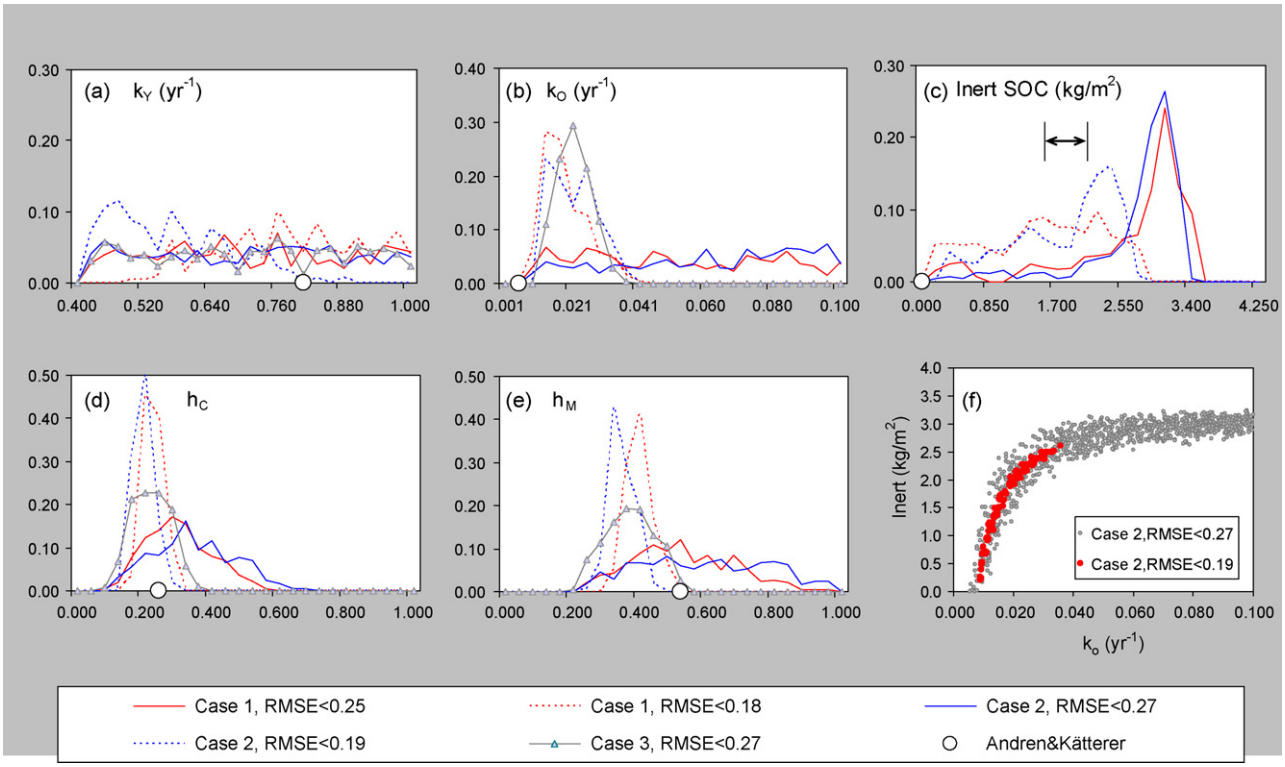

Fig. 7. Probability densities for ICBM parameters after conditioning to the Frame Trial data (Cases 1–3) with the GLUE method. Two distributions each are shown for Cases 1 and 2 calibrations over a range of RMSE thresholds. The Case 1 RMSE < 0.25 and Case 2 RMSE < 0.27 results represent parameter distributions for approximately 90% cover of respective data observations. The Case 1 RMSE < 0.18 and Case 2 RMSE < 0.19 results represent the upper echelons of the respective parameter spaces, in close proximity to the lowest achieved RMSE values (Table 3). The Case 3 RMSE < 0.27 results bracket 90% of data and can be compared to the Case 2 RMSE < 0.27 results.

hypothetical example, if the model parameter I was believed to lie roughly in the vicinity of 1.5–2.0 kg/m^2 (range indicated with arrows in Fig. 7c), then resultant parameter distributions tightened (Fig. 7b, d and e) and covariant relationships changed (Table 4) in the remaining model parameters. Although not shown in Fig. 7b, the corresponding range for k_O at RMSE < 0.19 resulting from this hypothetical information would be 0.14–0.22 year^{-1}, a substantial reduction in uncertainty compared to the Case 2 calibration that assumed no prior information on I.

3.3. Machang'a calibrations (Case 4)

Results of calibration to the Machang'a dataset using the same methods were largely complimentary to the Frame Trial results. Before detailing comparisons though, it is important to keep in mind that the Machang'a dataset is shorter than the Frame Trial data presented here (13 years compared to 51), and that all of the Machang'a data (12 observations after startup, Fig. 2) occur within an interval where there was only one observation in the Frame Trial (Fig. 1).

The lowest achieved (near optimal) RMSE errors in the Machang'a and Frame Trial simulations were within 10% of each other (Table 3). It follows then that the widths of uncertainty bands in time series simulations were comparable between Cases 2 and 4 calibrations at comparable threshold criteria (compare the lowest RMSE threshold bands in Fig. 8 at 2002 to those in Fig. 4 at 1969). Residuals for the Machang'a time series simulations also exhibited bias in some treatment simulations (e.g. Control treatment, Fig. 8). In terms of model parameter distributions, the distribution of k_O from the Machang'a calibration was longer tailed but largely coincident to the k_O distribution from the Case 2 Frame Trial calibration (Fig. 9b); the h_C parameter distributions were also very close (Fig. 9d); the covariant relationship between k_O and I in accepted

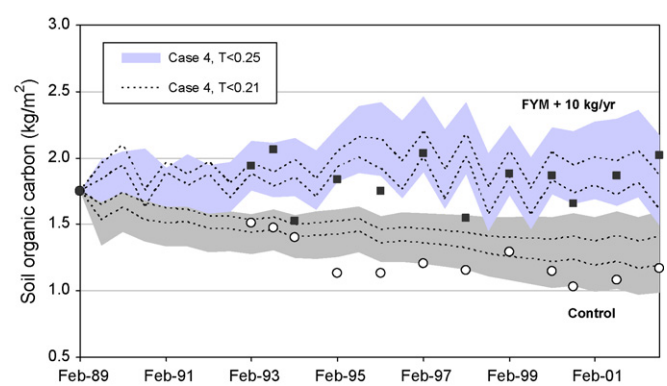

Fig. 8. Time series simulations with uncertainty bounds for two of the six treatments at Machang'a. The remaining four treatments were overlapping and not shown for clarity.

solutions was almost equally strong (R^2 = 0.79, Fig. 9f) to Case 4; and the plausible range of inert SOC was also quite wide (0–70% of the assumed initial condition).

There were also several notable differences between Cases 2 and 4 results. As we expected, the calibrated range for the humification coefficient, h_M, was lower when ICBM was conditioned to the Machang'a dataset (Fig. 9e). Recall that there was no significant difference in response between 5 and 10 t/year manure treatments in the Machang'a trials (Kihanda et al., 2005). In the context of modeling these data with the ICBM, GLUE results suggested that reducing additions of manure inputs to the *Old* storage (or essentially increasing oxidation) was the most effective pathway to jointly simulating treatment responses with common model parameters. It was also not possible to bracket 90% of the Machang'a data in an effective way with GLUE uncertainty bands (Fig. 5). This

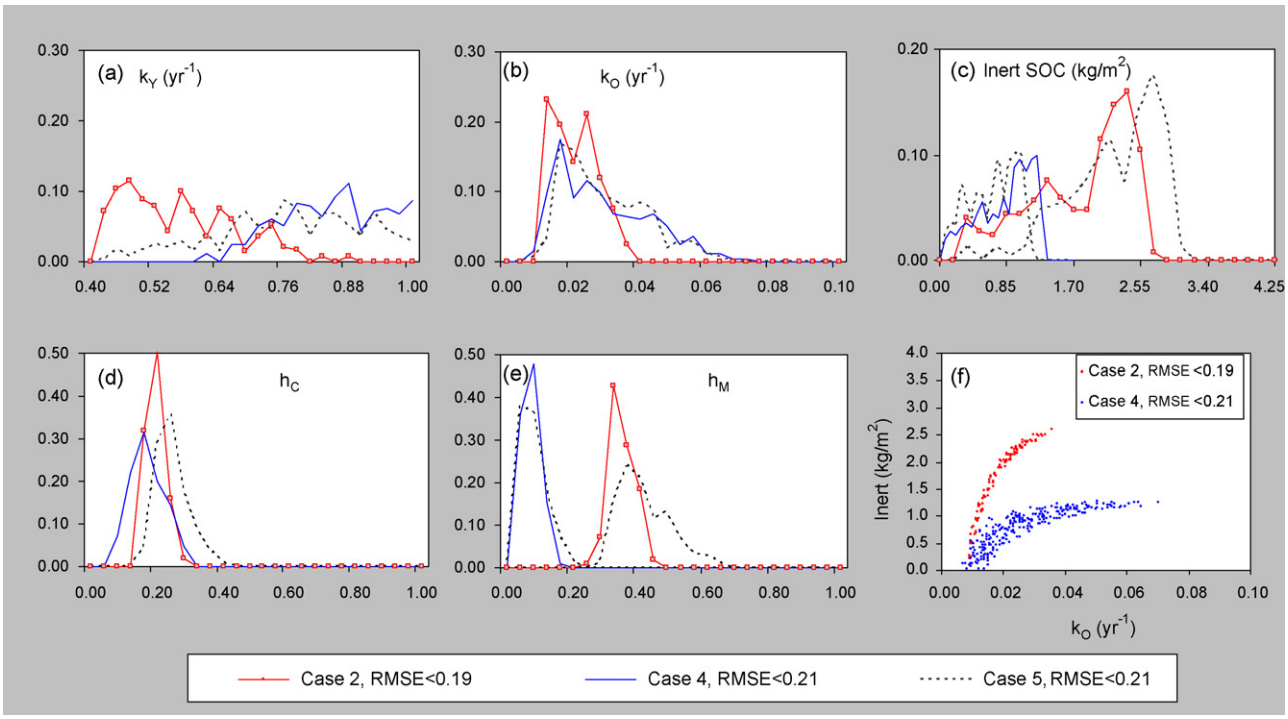

Fig. 9. Parameter distributions for lowest RMSE thresholds tested for Cases 2 (RMSE < 0.19), 4 (RMSE < 0.20), and 5 (RMSE < 0.21). The Case 2 distributions are identical to those shown in Fig. 7 and are repeated here to establish a relative reference. Two distributions are shown for Case 5 in frames (c) and (e), as these were treated as site-specific parameters in the joint Case 5 calibration.

seemed to be due largely to the relatively high biases in model residuals in the C and B1 treatment (Table 1) simulations, compared to the low overall differentiation between Machang'a treatment responses (Fig. 2). However, it is important to note that the manure treatment at the Frame Trial was also not differentiated from other treatments 10-year after startup (1966 data in Fig. 1), and that a notable difference was not observed in the Frame Trial until the next data point at 18 years, 5 years later than the whole the duration of the Machang'a field trial.

3.4. Combined calibration (Case 5)

Here, we tested one GLUE run using an RMSE < 0.21 kg/m^2 threshold to demonstrate the potential for conditioning model parameters to multiple datasets simultaneously. Joint calibration to the Machang'a and Frame Trial datasets was successful when some parameters were treated as global (k_Y, k_O, h_C) and others treated as site-specific (h_M, I) (Fig. 9). Distributions were wider in some of the joint parameters, such as k_O, since the RMSE criteria was less limiting in the joint calibration than for the separate dataset calibrations (Fig. 9b).

4. Discussion

Calibration of the ICBM with uncertainty analysis has provided numerous insights. We have demonstrated an alternative method for calibration of SOC or similar models using a selective Monte Carlo procedure. This methodology has provided information about the model parameter space (Figs. 7 and 9) and uncertainty bounds on future model predictions (Figs. 4 and 8), an effective method for evaluating impacts of different data durations and types on parameter uncertainties (Fig. 7), a means to evaluate model performance across different climatic regions (Fig. 9), and a means to jointly and in principle more robustly calibrate model parameters to multiple datasets simultaneously (Fig. 9). It seems likely this method would have benefits for use with other SOC models and datasets as well.

In terms of time series simulations, the results presented herein (Figs. 4 and 8) compare favorably to past efforts with both datasets. The Frame Trial data (1956–1991) were originally simulated with ICBM assuming zero inert storage (Andrén and Kätterer, 1997). Petersen et al. (2005) revisited calibration using extended time series (1956–1999), the five-pool CN-SIM model, and an assumption of a 41% inert fraction in SOC storage. Their results (Petersen et al., 2005) appeared to exhibit similar simulation errors and residual structures as our results here (Fig. 6), despite the more-complicated model structure. Previous simulations of the Machang'a field trials using APSIM (Micheni et al., 2004) and RothC (Kamoni et al., 2007a,b) models assumed inert carbon fractions of 50% and 17%, respectively. Simulation results from these efforts appeared adequate, but seemed to demonstrate larger errors than shown here with ICBM. In general, these observations fall in line with previous observations that more-complicated model structures do not necessarily improve simulations of soil carbon (i.e. decrease errors) in field trial data (Traore et al., 2008). However it is important to note that systematic patterns in error residuals that we observed may still be the result of inherent errors in model structure. The sub-model for the relative climate factor contains several parameters and a different model for estimating r may have reduced this pattern in the residuals. Additionally, various microbial efficiency parameters may also be hidden behind the remaining discrepancies. We fully welcome future studies with a combination of model structures and number of parameters to investigate additional possible explanations to current available long-term data sets using available uncertainty methods.

Modeling and initializing inert pools in SOC models has been a topic of considerable interest in recent years (e.g. Falloon et al., 1998, 2000; Puhlmann et al., 2006; Zimmermann et al., 2007). Falloon et al. (1998,2000) provided a means to estimate the size of inert pools based on size of the initial condition of soil organic carbon at a site. However those regressions were not so useful here, as their 95% confidence intervals covered a range of approximately 95% and 65% of the initial soil organic content at the Frame Trail and

Machang'a sites, respectively. There has also been some debate as to if initial soil carbon pools from laboratory fractionation can be precisely related to initial conditions in SOC models (Paul et al., 2006). Here, we have shown that even uncertain information on an inert pool (albeit less uncertain than estimates according to Falloon et al., 1998) were valuable to reduce uncertainties in model parameters. Our use of the GLUE method provides a template for meaningfully incorporating this type of fuzzy information into other and future SOC model calibrations.

One subject in this study was the effective RMSE threshold value for each case study that provided coverage of 90% of data observations with GLUE-estimated time series uncertainty bounds. Another subject was in parameter distributions at the upper echelons of the model parameter space, which we functionally defined here as the first RMSE increment (within 0.01 kg/m^2) above the lowest achieved RMSE scores in a particular case study. One of the principal characteristics of the model parameter space that was demonstrated herein was the non-uniqueness and in fact wide-ranging values of model parameter sets in the vicinity of optimal RMSE scores (Figs. 7 and 9, Table 3). In general, most parameter distributions were not sharply peaked in the regions of lowest achieved RMSE scores, but instead rather flat with high covariance. This is not a unique characteristic to ICBM or these data, and in fact has been previously demonstrated in numerous studies with models of similar complexity and beyond (e.g., Beven and Freer, 2001; Juston et al., 2009; Liu et al., 2009). Beven (2006) has termed this condition *equifinality*, and defined it as an inability to meaningfully distinguish one single "best" parameter set given inherent uncertainties and errors in available data and model structures and typical over-parameterization in model structures. The principle of equifinality leads directly to the notion that multiple parameter sets provide equally feasible representations of the system. Here, we demonstrated feasible representations of the Frame Trail and Machang'a datasets with rate constants varying by a factor-of-four or more and inert soil carbon content varying between 0 and 60% (Fig. 9). Indeed, this equifinality is also apparent in closed-form steady-state solutions to the ICBM governing equations (Appendix A). Covariance between similar model parameters has also been previously identified by Bostick et al. (2007) in a dynamic two-storage model but the GLUE results shown here seem to more clearly elucidate this relationship. Petersen et al. (2005) also discussed high sensitivity in CN-SIM model to inert pool assumptions. Falloon et al. (2000) investigated the impacts of inert pool assumptions on RothC model predictions, but they did not consider that rate constants and inert pool assumptions might be intertwined. It seems likely to us that a strong covariance would be discovered between inert pool assumptions and longer-term rate constants in other more-complicated SOC models as well, including RothC, Century, APSIM, and CANDY, if the calibration of these models were re-evaluated with Monte Carlo methods such as GLUE.

The addition of 16 years additional data in the Frame Trial time series beyond the "original" 35-year time series provided only marginal refinement to model parameter relationships (Fig. 7). It appeared the most substantial refinement was related to simulating the manure (FYM) treatment. Interestingly, it was also the manure treatments at Machang'a that raised some flags in our simulations due to their non-significant treatment responses (Kihanda et al., 2005). Although we cannot conclude with certainty due to differing data set lengths and data densities, it seems that these findings suggest a possibility for improving the manure response in the ICBM.

Lastly, statistical methods for uncertainty analysis provide a counterpoint to the non-statistical GLUE method and should be a topic of future research for calibrating soil carbon models to field trial data. Markov Chain Monte Carlo (MCMC) analysis using Bayesian inference is a formal statistical technique that has gained popularity in hydrological and forest modeling applications (e.g. Kuczera and Parent, 1998; Van Oijen et al., 2005; Svensson et al., 2008; Klemedtsson et al., 2008). One complication is that formal error models used in these methods require assumptions on model residuals such that they are independent and with Gaussian distribution. There has been recent research focused on achieving these conditions with time series data and autocorrelated residuals in hydrological models using Box–Cox transformations and an autoregressive error model (e.g., Yang et al., 2007). Here however, we identified bias in residuals that were not consistent amongst parallel time series simulations of experimental treatments (Figs. 6 and 8). This is an interesting condition (i.e., constructing a formal error model for parallel simulations each with different error structure) and one that may require further investigation before probabilistic uncertainty methods could be used with confidence for calibrating SOC models to multi-treatment field trial data.

Acknowledgement

The authors would like to acknowledge Geoff Warren for assistance with the Machang'a dataset. We would also like to acknowledge the Swedish International Development Agency (SIDA) Department of Research Cooperation (SAREC) for partial funding of this work.

Appendix A.

The governing differential equations for ICBM are (Andrén and Kätterer, 1997):

$$\frac{dY}{dt} = i - k_Y rY \tag{A.1}$$

$$\frac{dO}{dt} = hk_Y rY - k_O rO \tag{A.2}$$

where state variable and parameter definitions are given in Fig. 3 and Table 2.

The total SOC in the upper 20 cm is the sum of young, old and inert pools (Fig. 3):

$$T = Y + O + I \tag{A.3}$$

Using long-term average values for carbon inputs (i) and climate-dependent scale factor (r) and a weighted mean for the humification factor (h), the closed-form solution for total SOC in the upper 20 cm can be expressed as:

$$T_{SS} = \frac{i(1/k_Y + h/k_O)}{r} + I \tag{A.4}$$

The relative proportions of Y and O in steady-state are:

$$\frac{Y_{SS}}{O_{SS}} = \frac{k_O}{h\,k_Y} \tag{A.5}$$

References

Andrén, O., Kätterer, T., 1997. ICBM: the introductory carbon balance model for exploration of soil carbon balances. Ecological Applications 7 (4), 1226–1236.

Andrén, O., Kihara, J., Bationo, A., Vanlauwe, B., Kätterer, T., 2007. Soil climate and decomposer activity in sub-Saharan Africa estimated from standard weather station data: a simple climate index for soil carbon balance calculations. Ambio 36 (5), 379–386.

Bationo, A., Kihara, J., Vanlauwe, B., Waswa, B., Kimetu, J., 2007. Soil organic carbon dynamics, functions and management in West African agro-ecosystems. Agricultural Systems 94 (1), 13–25.

Beven, K.J., 2006. A manifesto for the equifinality thesis. Journal of Hydrology 320, 18–36.

Beven, K.J., 2009. Environmental Modeling: An Uncertain Future? Routledge, London.

Beven, K.J., Binley, A.M., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 279–298.

Beven, K.J., Freer, J., 2001. Equifinality, data assimilation, and uncertainty estimation in mechanistic modeling of complex environmental systems using the GLUE methodology. Journal of Hydrology 249 (1–4), 11–29.
Bolinder, M.A., Janzen, H.H., Gregorich, E.G., Angers, D.A., VandenBygaart, A.J., 2007. An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada. Agriculture, Ecosystems and Environment 118, 29–42.
Bostick, W.M., Bado, V.B., Bationo, A., Soler, C.T., Hoogenboom, G., Jones, J.W., 2007. Soil carbon dynamics and crop residue yields of cropping systems in the Northern Guinea Savanna of Burkina Faso. Soil & Tillage Research 93, 138–151.
Coleman, K., Jenkinson, D.S., 1996. ROTHC-26. 3: a model for the turnover of carbon in soil. In: Powlson, S., Smith, P., Smith, J.U. (Eds.), Evaluation of Soil Organic Matter Models (I). Springer-Verlag, Berlin, pp. 237–246.
Debreczeni, K., Körschens, M., 2003. Long-term field experiments of the world. Archiv fur Acker und Pflanzenbau und Bodenkunde 49, 465–483.
Diels, J., Vanlauwe, B., Van der Meersch, M.K., Sanginga, N., Merckx, R., 2004. Long-term soil organic carbon dynamics in a subhumid tropical climate: ^{13}C data in mixed C_3/C_4 cropping and modeling with RothC. Soil Biology and Biochemistry 26, 1739–1750.
Falloon, P., Smith, P., 2002. Simulating SOC changes in long-term experiments with RothC and CENTURY: model evaluation for a regional scale application. Soil Use and Management 18, 101–111.
Falloon, P., Smith, P., Coleman, K., Marshall, S., 1998. Estimating the size of the inert organic matter pool from total soil organic carbon content for use in the Rothamsted Carbon Model. Soil Biology and Biochemistry 30 (8/9), 1207–1211.
Falloon, P., Smith, P., Coleman, K., Marshall, S., 2000. How important is inert organic matter for predictive soil carbon modeling using the Rothamsted carbon model? Soil Biology and Biochemistry 32, 433–436.
Guo, L., Falloon, P., Coleman, K., Zhou, B., Li, Y., Lin, E., Zhang, F., 2007. Application of the RothC model to the results of long-term experiments on typical upland soils in northern China. Soil Use and Management 23 (1), 63–70.
Hyvönen, R., Ågren, G.I., Andrén, O., 1996. Modelling long-term carbon and nitrogen dynamics in an arable soil receiving organic matter. Ecological Applications 6 (4), 1345–1354.
Kamoni, P.T., Gicheru, P.T., Wokabi, S.M., Easter, M., Milne, E., Coleman, K., Falloon, P., Pasutian, K., 2007a. Predicted soil organic carbon stocks and changes in Kenya between 1990 and 2030. Agriculture, Ecosystems and Environment 122, 105–113.
Kamoni, P.T., Gicheru, P.T., Wokabi, S.M., Easter, M., Milne, E., Coleman, K., Falloon, P., Pasutian, K., Killian, K., Kihandra, F.M., 2007b. Evaluation of two soil carbon models using two Kenyan long term experimental datasets. Agriculture, Ecosystems and Environment 122, 95–104.
Kihanda, F.M., Warren, G.P., Micheni, A.N., 2005. Effects of manure and fertilizer on grain yield, soil carbon, and phosphorus in a 13-year field trial in semi-arid Kenya. Exploratory Agriculture 41, 389–412.
Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis Algorithm. Journal of Hydrology 211, 69–85.
Juston, J., Seibert, J., Johansson, P.-O., 2009. Temporal sampling strategies and uncertainty in calibrating a conceptual hydrological model for a small boreal catchment. Hydrological Processes 23 (21), 3093–3109.
Klemedtsson, L., Jansson, P.-E., Gustafsson, D., Karlberg, L., Weslien, P., von Arnold, K., Ernfors, M., Langvall, O., Lindroth, A., 2008. Bayesian calibration method used to elucidate carbon turnover in forest on drained organic soil. Biogeochemistry 89 (1), 61–79.
Liu, Y., Freer, J.E., Beven, K.J., Matgen, P., 2009. Towards a limits of acceptability approach to the calibration of hydrological models: extending observation error. Journal of Hydrology 367, 93–103.
Lugato, E., Paustian, K., Giardini, L., 2007. Modelling soil organic carbon dynamics in two long-term experiments of north-eastern Italy. Agriculture, Ecosystems and Environment 120, 423–432.
Micheni, A.N., Kihanda, F.M., Warren, G.P., Probert, M.E., 2004. Testing the APSIM model with experimental data from the long-term manure experiment at Machang'a (Embu), Kenya. In: Delve, R.J., Probert, M.E. (Eds.), Modelling Nutrient Management in Tropical Cropping Systems. Australian Centre for International Agricultural Research, Canberra, Australia.
Nandwa, S.M., 2001. Soil organic carbon (SOC) management for sustainable productivity of cropping and agro-forestry systems in Eastern and Southern Africa. Nutrient Cycling in Agroecosystems 61, 143–158.
Nash, J.E., Sutcliffe, I.V., 1970. River flow forecasting through conceptual models: Part 1—A discussion of principles. Journal of Hydrology 10, 282–290.
Nilsson, K.S., Hyvönen, R., Ågren, G.I., 2005. Using the continuous-quality theory to predict microbial biomass and soil organic carbon following organic amendments. European Journal of Soil Science 56, 397–405.
Ogle, S.M., Fay, B.F., Eve, M.D., Paustian, K., 2003. Uncertainty in estimating land use and management impacts on soil organic carbon storage for US agricultural lands between 1982 and 1997. Global Change Biology 9 (11), 1521–1542.
O'Neill, B.C., Melnikov, N.B., 2008. Learning about parameter and structural uncertainty in carbon cycle models. Climatic Change 89, 23–44.
Paul, E.A., Morris, S.J., Conant, R.T., Plante, A.F., 2006. Does the acid hydrolysis-incubation method measure meaningful soil organic carbon pools? Soil Science Society of America Journal 70, 1023–1035.
Paustian, K., Parton, W.J., Persson, J., 1992. Modeling soil organic matter in organic-amended and nitrogen-fertilized long-term plots. Soil Science Society of America Journal 56, 476–488.
Persson, J., Kirchmann, H., 1994. Carbon and nitrogen in arable soils as affected by supply of N fertilizers and organic manures. Agriculture, Ecosystems and Environment 51, 249–255.
Petersen, B.M., Bernsten, J., Hansen, S., Jensen, L.S., 2005. CN-SIM—a model for the turnover of soil organic matter: I. Long-term carbon an radiocarbon development. Soil Biology & Biochemistry 37, 359–374.
Post, J., Hattermann, F.F., Krysanova, V., Suckow, F., 2008. Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. Environmental Modelling and Software 23, 125–138.
Puhlmann, M., Kuka, K., Franko, U., 2006. Comparison of methods for the estimation of inert carbon suitable for initialization of the CANDY model. Nutrient Cycling in Agroecosystems 74, 295–304.
Richter, D.D., Callaham, M.C., Powlson, D.S., Smith, P., 2007. Long-term soil experiments: Keys to managing Earth's rapidly changing ecosystems. Soil Science Society of America Journal 71, 266–279.
Smith, P., Smith, J.U., Powlson, D.S., McGill, W.B., Arah, J.R.M., Chertov, O.G., Coleman, K., Franko, U., Frolking, S., Jenkinson, D.S., Jensen, L.S., Kelly, R.H., Klein-Gunnewiek, H., Komarov, A.S., Li, C., Molina, J.A.E., Mueller, T., Parton, W.J., Thornley, J.H.M., Whitmore, A.P., 1997. A comparison of the performance of nine soil organic matter models using datasets from seven long-term experiments. Geoderma 81, 153–225.
Svensson, M., Jansson, P.-E., Gustafsson, D., Berggren Kleja, D., Langvall, O., Lindroth, A., 2008. Bayesian calibration of a model describing carbon, water and heat fluxes for a Swedish boreal forest stand. Ecological Modelling 213 (3–4), 331–344.
Traore, P.C.S., Bostick, W.M., Jones, J.W., Koo, J., Goita, K., Bado, B.V., 2008. A simple soil organic-matter model for biomass data assimilation in community-level carbon contracts. Ecological Applications 18 (3), 624–636.
Vanlauwe, B., Giller, K.E., 2006. Popular myths around soil fertility in sub-Saharan Africa. Agriculture, Ecosystems and Environment 116, 34–46.
Van Oijen, M., Rougier, J., Smith, R., 2005. Bayesian calibration of process-based forest models: bridging the gap between models and data. Tree Physiology 25, 915–927.
Xenakis, G., Ray, D., Mencuccini, M., 2008. Sensitivity and uncertainty analysis from a coupled 2-PG and soil organic mater decomposition model. Ecological Modelling 219, 1–16.
Yang, J., Reichert, P., Abbaspour, K.C., Yang, H., 2007. Hydrological modeling of the Chaohe basin in China: statistical model formulation and Bayesian inference. Journal of Hydrology 340 (3–4), 167–182.
Zimmermann, M., Leifled, J., Scmidt, W.I., Smith, P., Fuhrer, J., 2007. Measrued soil organic matter fractions can be related to pools in the RothC model. European Journal of Soil Science 58, 658–667.
Zingore, S., Delve, R.J., Nyamangara, J., Giller, K.E., 2008. Multiple benefits of manure: the key to maintenance of soil fertility and restoration of depleted sandy soils on African smallholder farms. Nutrient Cycling and Agroecosystems 80 (3), 267–282.