https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

A robust DayCent model calibration to assess the potential impact of integrated soil fertility management on maize yields, soil carbon stocks and greenhouse gas emissions in Kenya.

Moritz Laub[1], Magdalena Necpalova[1,2], Marijn Van de Broek[1], Marc Corbeels[3,4], Samuel Mathu Ndungu[4], Monicah Wanjiku Mucheru-Muna[5], Daniel Mugendi[6], Rebecca Yegon[6], Wycliffe Waswa[4], Bernard Vanlauwe[4], and Johan Six[1]

[1]Department of Environmental Systems Science, ETH Zurich, 8092 Zürich, Switzerland
[2]University College Dublin, School of Agriculture and Food Science, Ireland
[3]CIRAD, Avenue d'Agropolis, F-34398 Montpellier, France
[4]International Institute of Tropical Agriculture (IITA), c/o ICIPE Compound, P. O. Box 30772-00100, Nairobi, Kenya
[5]Department of Environmental Sciences and Education, Kenyatta University, P.O. Box 43844-00100, Nairobi, Kenya
[6]Department of Water and Agricultural Resource Management, University of Embu, P.O. Box 6-60100, Embu, Kenya

Correspondence: Moritz Laub (moritz.laub@usys.ethz.ch)

Abstract. Sustainable intensification schemes that increase crop production and soil fertility, such as integrated soil fertility management (ISFM), are a proposed strategy to close yield gaps and achieve food security in sub-Saharan Africa while maintaining soil fertility. However, field trials are insufficient to estimate the potential impact of such technologies at the regional or national scale. Upscaling via biogeochemical models, such as DayCent, from the field-scale to a larger region can be a suitable and powerful way to assess the potential of such agricultural management practices at scale, but they need to be calibrated to new environments and their reliability needs to be assured. Here, we present a robust calibration of DayCent to simulate maize productivity under ISFM, using data from four long-term field experiments. The experimental treatments consisted of the addition of low- to high-quality organic resources to the soil, with and without mineral N fertilizer. We assess the potential of DayCent to represent the key aspects of sustainable intensification, including 1) yield, 2) changes in soil carbon, and 3) global warming potential. The model was calibrated and cross-evaluated with the probabilistic Bayesian calibration technique.

The standard parameters of DayCent led to poor simulations of maize yield (Nash–Sutcliffe modeling efficiency; EF 0.33) and changes in SOC (EF -1.3) for different ISFM treatments. After calibration of the model, both the simulation of maize yield (EF 0.51) and the change in SOC (EF 0.54) improved significantly compared to the model with the standard parameter values. A leave-one-site-out cross-evaluation indicated the robustness of the approach for spatial upscaling (i.e., the significant improvement, described before, was achieved by calibrating with data from 3 sites and evaluating with the remaining site). The SOC decomposition parameters were altered most severely by the calibration. They were an order of magnitude higher compared to the default parameter set. This confirms that the decomposition of SOC in tropical maize cropping systems is much faster than in temperate systems and that the DayCent temperature function is not suitable to capture this with a single parameter set. Finally, the global warming potential simulated by DayCent was highest in control -N treatments (0.5-2.5 kg CO_2 equivalent per kg grain yield, depending on the site) and could be reduced by 14 to 72% by combined application of

mineral N and manure at a medium rate. In three of the four sites, the global warming potential was largely (> 75%) dominated by SOC losses. In summary, our results indicate that DayCent is suitable for estimating the impact of ISFM from the site to the regional level, that trade-offs between yields and global warming potential are stronger in low-fertility sites, and that the reduction of SOC losses is a priority for the sustainable intensification of maize production in Kenya.

1 Introduction

Similar to many countries of Sub-Saharan Africa (SSA), the low maize yields in Kenya, on average 1.7 t ha^{-1} compared to the global average of 5.6 t ha^{-1} over the last decade (2011-2021) (FAO, 2023), are in part responsible for low self-sufficiency of food production. Consequently, around 20% of the Kenyan population lives with severe food insecurity (World-Bank, 2021b). If yields are not improved, increased demand due to population growth will further deteriorate self-sufficiency and in general food security in the coming decades (Zhai et al., 2021): yield declines resulting from more frequent extreme weather events could further exacerbate this situation (Lobell et al., 2011). One of the key limitations to sustainable maize production in SSA is the lack of mineral fertilizer and organic inputs. Integrated soil fertility management (ISFM) is a sustainable intensification practice that can alleviate these limitations by combining the use of mineral fertilizers with organic inputs (Vanlauwe et al., 2010). Different studies reported that ISFM has the potential to more than double maize yields in Kenya, especially in currently unfertile soils, due to the effects on soil fertility, including soil organic matter (SOM) content (Chivenge et al., 2009, 2011). However, in addition to the amount and quality of fertilizer and organic inputs, the effectiveness of ISFM in increasing yields strongly depends on local site conditions, such as soil and climate (Chivenge et al., 2011). Furthermore, increasing SOM also contributes to minimizing the negative effects of climate change, and there may be considerable potential in carbon-depleted soils in SSA (Corbeels et al., 2019).

Hence, to close yield gaps in a resource-efficient way and to assess the climate change mitigation potential of ISFM, we need to understand the effects of long-term implementation of ISFM at larger scales. Ideally, this would be facilitated by implementing a large number of long-term experiments (LTE) across a representative range of soil and climatic conditions. However, due to the high demands in terms of cost, labor, and time associated with maintaining LTE, the number of LTE that evaluate variation in the effects of ISFM practices due to site-specific conditions is very limited. In addition, using statistical predictive techniques to upscale results obtained with a low number of sites could lead to low predictive power and large predictive errors, because it is unlikely that the effects of variation in soils and climate on yield and the development of soil organic matter (SOM) would be fully captured by the statistical models.

Biogeochemical process-based ecosystem models, such as DayCent (Parton et al., 1998; Del Grosso et al., 2001), are able to simulate the influence of the important driving variables of yield and SOM formation using semi-mechanistic functions developed through decades of agronomic and soil research. Because they (partly) embed our current understanding of the complex ecosystem processes, they are more robust for upscaling of the yield potential (Saito et al., 2021) and the SOM building capacity (Lee et al., 2020), compared to statistical predictive techniques. However, to avoid bias in model output, it

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

is best practice that models are calibrated and validated to local conditions (Necpálová et al., 2015) whenever they are applied outside the range of usual conditions, for example, in a new climate zone with different soils.

Although DayCent has been used to estimate the development of SOM stock in Kenya on a national scale (Kamoni et al., 2007) and recently the efficiency of conservation agriculture in Ethiopia (Lemma et al., 2021), DayCent's modules of SOM and maize crop growth have never been rigorously calibrated and validated for tropical agroecosystems in SSA. A recent test of DayCent in Kenyan maize growing systems showed that SOM turnover is underpredicted by DayCent (Nyawira et al., 2021). Because SOM in biogeochemical models is coupled to nitrogen (N) mineralization, there is the potential that this translates into biased crop responses to N and biased crop growth rates in any upscaling exercise. A potential solution to this issue is the joint calibration of soil and crop parameters in DayCent, using long-term data from local trials. Ideally, it would include the uncertainty in the model parameters and outputs (Clifford et al., 2014), so a propagation of errors is possible in upscaling exercises (Stella et al., 2019). This is especially relevant because a recent study showed considerable uncertainty in DayCent SOM turnover rates, even when using a range of long-term experiments for calibration (Gurung et al., 2020).

In order to use DayCent to assess the potential of ISFM to contribute to closing the yield gap while minimizing environmental impact in Kenya and other SSA countries, the aim of this study was to use Bayesian calibration to derive robust DayCent parameters of SOM cycling and maize growth in Kenya. We used the experimental data of four long-term ISFM trials conducted in Kenya for almost two decades (Laub et al., 2022, 2023). Of these, two sites were in humid western Kenya and two in subhumid to semi-arid central Kenya. Their treatments were different additions of organic resources (1.2 and 4 t C ha^{-1} yr^{-1} including farmyard manure, high-quality *Tithonia diversifolia* and *Calliandra calothyrsus* residues as well as low-quality low quality stover of *Zea mays* and sawdust), combined with (+N) or without (-N) 120 kg mineral N fertilizer ha^{-1} $season^{-1}$. They thus contain a range of experimental conditions and inputs. The goal was to evaluate to what extent DayCent can replicate the differences in yields and SOM development in response to different qualities and rates of organic resources combined with different rates of N fertilizer and between sites. Furthermore, because ISFM can simultaneously be a source of N_2O to the atmosphere (Leitner et al., 2020) and mitigate CO_2 emissions from the soil (Laub et al., 2022), the second objective was to evaluate the total global warming potential of different rates of organic material in ISFM to find the optimal balance between limiting greenhouse gas emissions from the soil and optimizing the yield (that is, sustainable intensification).

Thus, the specific objectives of this study were (i) to test the capability of an uncalibrated version of DayCent to simulate yield and SOC development of the different ISFM practices, (ii) to calibrate DayCent to represent ISFM under Kenyan conditions using experimental data from the 4 LTE, displaying the confidence in model parameters by Bayesian calibration, and (iii) to use the calibrated model to gain an understanding of possible trade-offs between yield and SOM increases under ISFM, and to understand the global warming potential of the different ISFM treatments.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

2 Material and Methods

2.1 The experimental sites

The present study used data pooled from four experimental sites. The experiments tested different organic resource treatments at different rates in combination with the application of mineral N fertilizer in the context of ISFM. The experimental setup was identical between the sites. The sites are located in agriculturally important areas in western and central Kenya. The Embu and Machanga sites are both located in Embu County, in the central part of Kenya. The Aludeka site is located in Busia County in western Kenya, while Sidada is located in the adjacent Siaya County, just south of Busia (Table A1). The experiments in Embu and Machanga were initiated in early 2002 and the experiments in Aludeka and Sidada were initiated in early 2005. Therefore, 19 years of data were available in central and 16 years in western Kenya (2 sites x 16 years + 2 sites x 19 years = 70 site years = 140 site seasons). The sites cover a range of altitudes, temperatures, and precipitations. Embu, with a mean annual temperature (MAT) of 20 °C and 1200 mm annual precipitation, is the coolest of all sites, while Machanga has a MAT of 24°C and receives the lowest amounts of annual precipitation (800 mm). Sidada (23°C, 1700 mm) and Aludeka (24°C, 1700 mm) have a high MAT and receive significantly more precipitation than the sites in central Kenya. There are two rainy seasons at each site, which corresponds to two maize growing seasons per year. The long rainy season runs from March to August/September, the short rainy season runs from October until January/February. In terms of soil texture, Machanga and Aludeka have low clay content (13% clay at both sites), while Sidada and Embu are rich in clay (56 and 60%, respectively).

All experiments were set up as a split plot design with three replicates, with different qualities and quantities of organic resources as main plots and the presence or absence of N fertilizer as subplots. Maize was grown continuously in all experiments, with two crops per year, one in the long rainy season and one in the short rainy season. The design was similar in all four sites and has been described in detail for the Embu site (Chivenge et al., 2009; Gentile et al., 2011). Organic resource treatments consisted of high quality *Tithonia diversifolia* (TD) green manure and *Calliandra calothyrsus* (CC) prunings, low quality stover of *Zea mays* (MS) and sawdust from *Grevillea robusta* trees (SD), locally available farmyard manure (FYM) and a control treatment (CT) without any organic resource additions. Organic resources differed in quality by the contents of N, lignin and polyphenols (Table A2). Each organic resource was applied at two rates, 1.2 and 4 t C ha^{-1} yr^{-1}, while only one amount of applied mineral fertilizer, 120 kg N ha^{-1} ($CaNH_4NO_3$) in each of the two growing seasons was tested. Of that, 40 kg N ha^{-1} were applied with planting, and the remaining 80 kg N ha^{-1} about six weeks after planting. Organic resources were applied only once a year, prior to planting for the long rainy season in January or February. They were incorporated to a depth of 15 cm with hand hoes. Furthermore, a blanket application of 60 kg P ha^{-1} as triple superphosphate and of 60 kg K ha^{-1} as muriate potash at planting was provided to all plots once each season. The plots were kept weed free by hand weeding, between two and three times per season, and selective application of pesticides was used when necessary to protect against armyworm, stemborer, and/or termites.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

2.2 The DayCent model

The DayCent model (version 2020 of DD_EVI) is a terrestrial ecosystem model of intermediate complexity (Del Grosso et al., 2001). It simulates C and N fluxes within the soil-plant-atmosphere continuum on a daily basis and has been parameterized for many crops and ecosystems (Necpalova et al., 2018). It has submodules to simulate plant growth, organic resource and soil organic matter (SOM) decomposition, mineralization of N, soil water and temperature, N gas fluxes, and CH oxidation$_4$. The net primary productivity of plants is a function of their genetic potential, a simplified phenology, solar radiation, temperature, and stresses, such as reduced water or N availability. Here, we used the non-growing degree day version of the DayCent crop module, which does not simulate phenology but has a seedling stage with reduced growth until a certain biomass (full canopy) is reached. SOC and soil N in the upper 20 cm are represented by active, slow, and passive SOM pools, while litter and organic resources are represented by structural and metabolic litter pools (Parton et al., 1987). All soil pools are conceptual and have no measurable counterparts, whereas the litter pools are semiquantitative, that is, their division is based on the measurable ratio of lignin to N in organic resources and plant litter. For temperate conditions, DayCent has been shown to adequately simulate crop yields, SOC and soil N dynamics and N_2O emissions (Del Grosso et al., 2005; Necpálová et al., 2015; Necpalova et al., 2018; Gurung et al., 2020, 2021). Yet, for tropical conditions the performance has not been studied in similar detail. A recent paper showed that the general model fit of the uncalibrated model was suboptimal (Nyawira et al., 2021).

2.3 Data used for the DayCent model evaluation/calibration

In a process that was repeated four times, a large data set was used based on three of the four sites for the model calibration and the validation was performed based on the fourth site. The plot-scale yield of maize grain and the aboveground biomass, both on a dry matter basis (t ha^{-1}), were available for each cropping season between 2002 and 2020 (further details in Laub et al., 2023). In addition to that, plot-scale SOC and total N contents in the top 15 cm were available for several points over time. In Embu and Machanga, soil samples were taken every two to three years since initiation in 2002 until 2021, but in Sidada and Aludeka, only in 2005, 2018, 2019 and 2021 (further details in Laub et al., 2022). Because bulk density was not available for most soil samples and in 2021 there was no significant difference in topsoil bulk density between treatments at any site, the mean bulk density per site was used to calculate SOC stocks of the top 15 cm of soil depth. All simulations were conducted at the site scale, so the plot-scale (i.e. replicate) measurements were aggregated to the site scale, calculating means and variances. DayCent calculates SOC to a depth of 20 cm, so we rescaled the SOC stocks for the top 15 cm to the top 20 cm, using the formula of Jobbágy and Jackson (2000):

$$SOC_{20}(kg\ ha^{-1}) = \frac{1-\beta^{20}}{1-\beta^{15}} * SOC_{15} \quad (1)$$

Here, SOC_{20} and SOC_{15} are SOC stocks in kg ha^{-1} in the top 20 and 15 cm soil depth, respectively. The parameter β is the relative decrease of SOC stocks with depth, for which we took the mean values across sites (0.9725), calculated from the 2021 sampling, where samples from 0-15, 15-30, and 30-50 cm were available for all of the sites.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Data on N_2O emissions were used in the model evaluation phase, but not for model calibration, because the data were scarce and subject to high variability. The N_2O measurements were conducted after N fertilization in 2005 (weekly measurements form March to June in Embu and Machanga and daily measurements in Machanga in November), in 2013 and 2018 (weekly measurements form March to beginning of May in Sidada and Aludeka), and in 2021 (weekly measurements form mid-March to mid-May in Sidada). They were conducted with the static chamber method (Hutchinson and Mosier, 1981). Measuring frames were permanently installed in the plots for a whole rainy season. The chambers (0.27 × 0.375 × 0.11 m) were made of polyvinylchloride and equipped with a vent tube and a fan to homogenize gas inside them before gas sampling. The measured N_2O emissions were evaluated at two levels of aggregation. First, as site means per measurement day (from the three replicates, similar to all other data) and second as cumulative emissions over the whole season, for which we first used the trapezoid method at the plot scale (Levy et al., 2017), specifically, the trapz function of R (Tuszynski, 2021). Site-scale means and variances were then computed for these cumulative N_2O emissions, similar to all other measurements.

Furthermore, data on soil mineral N (N_{min}), measured as NH_4^+ and NO_3^-, and measured moisture content at soil depths of 0-15 cm, were available from several measurement campaigns in the years 2012, 2013, 2018, 2019 and 2020. Finally, in the control and the 1.2 t C plots of the *Calliandra*, farmyard manure and maize stover treatments, continuous soil moisture measurements from sensors placed in each replicate at 10 cm soil depth (EC-5 Soil Moisture Sensor, Meter, Germany), were available for Sidada and Aludeka sites (March 2018 to December 2020). These soil moisture measurements were used to determine the optimal pedotransfer functions for soil hydraulic conductivity before the actual model calibration phase.

2.3.1 Model driving variables and model assumptions

The site-specific crop management data used to run the DayCent model was obtained from field management operation records for each season at each site. This included specific dates for the yearly application of organic resources (in all but the control plots), manual plowing before planting, maize planting, split application of mineral N at rates of 40 kg and 80 kg N ha^{-1} per season, weeding and harvest dates. Pesticide application events and gap filling plus maize thinning are not part of standard DayCent management and were therefore not included in the DayCent schedule file. Our model runs therefore, assumed no occurrence of pests or diseases and an optimal plant density from the start, which in practice was ensured by manual thinning and gap filling.

The climate data used to run DayCent consisted of recorded data at each of the sites. However, filling the data gaps was necessary due to unavailability and loss of recorded data. In Embu and Machanga, manual recordings of daily minimum and maximum temperature and precipitation were available from 2002 until the end of 2007, but from 2008 until 2017, only measured precipitation was available. After 2017, high-quality data from newly installed TAHMO stations (https://tahmo.org/climate-data/) were available near the Machanga and Embu sites, with daily values for temperature and precipitation. In Aludeka and Sidada, manual recordings of daily minimum and maximum temperature and precipitation were available for all years from 2005 to 2017, during which high-quality weather stations were installed (Meter climate station, Meter Environment, Munich, Germany). These data gaps were filled by using the NASA POWER product (https://power.larc.nasa.gov/docs/methodology/). A bias correction for the minimum and maximum temperature of NASA POWER

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

data was performed, using a linear regression with measured data as dependent variable (y) and NASA POWER data as independent variable (x). Specifically, the slope and intercept of the regression equation $y = mx + b$, were used to produce a corrected estimate of these data. In our specific case, the slopes were not significantly different from 1, but intercepts (b) were significantly different from 0. The specific intercepts for maximum temperature were -0.3°C, -0.4°C, +3°C and +6°C for Embu, Machanga, Sidada and Aludeka, respectively. The intercepts for the minimum temperature were -0.25°C, -0.5°C, -3°C and +1°C for Embu, Machanga, Sidada, and Aludeka, respectively. For precipitation, no bias correction was done.

The data on the hydraulic properties of the soil needed in DayCent (volumetric field capacity, wilting point, and saturated hydraulic conductivity K_s) were calculated based on the soil texture measured at each site. We tested two pedotransfer functions to see which one provided a better fit: (1) the widely applied function of Saxton and Rawls (2006) and (2) the function of Hodnett and Tomasella (2002), which was specifically designed for tropical soils. Within the Hodnett and Tomasella (2002) equation, K_s was calculated using the Saxton and Rawls (2006) equation, with values of the water retention curve, α and n (van Genuchten, 1982), taken from Hodnett and Tomasella (2002), because their equation does not provide a way to estimate K_s. In initial test simulations, we compared the observed versus soil moisture contents in the top 15 cm, for which continuous measurements were available from Sidada and Aludeka from 2018 to 2020. In this comparison, the pedotransfer functions of Hodnett and Tomasella (2002) showed better agreement between the measured and simulated soil moisture contents than Saxton and Rawls (2006) and were consequently used in the application of the model at the four LTE sites.

2.3.2 Initial model parameterization and selection of potentially sensitive parameters for calibration

To parameterize the organic inputs, the mean lignin content and C/N ratio of the different organic materials across sites (Table A2) were used. This was justified because measurements were not available for all sites and years and because an analysis of variance of data from the years 2002, 2003, 2004, 2005 and 2006 for Embu and Machanga and from the year 2018 for all sites did not indicate any significant differences in lignin contents and C/N ratios between the sites or years. The C content in maize grain was assumed to be 42.5% throughout the simulation period. This was the mean value of measured grain C content across sites (standard deviation 1.8%), which were available from the short rainy season 2018 and long rainy season 2019 (data not shown).

The DayCent simulations were conducted at the site scale using averaged values across all three replicate plots for soil texture and SOC and soil N stocks, yield and aboveground biomass. This was done to reduce the computation time of the sensitivity analysis and calibration. Additionally, initial tests with the default DayCent parameterization showed that applying the model to each experimental replicate individually did not lead to better agreement between measured and simulated values than a single simulation of the average soil parameters of the three replicates. Therefore, final model simulations were conducted with averaged values for soil parameters (i.e., texture, bulk density, soil pH) across the three replicates per site. For data used in model calibration and evaluation, the site-specific variance for each type of measurement was used as a measure of uncertainty of the measured data (specifically, the median variance per site and measurement type computed from the three replicates from each time point of each treatment). Site-specific variances were used because a statistical analysis of the data in earlier work

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

had shown that variance heterogeneity existed only between sites, but not between treatments at the same site (Laub et al., 2022, 2023).

The standard parameter values of the DayCent 2020 version were taken as initial model parameters. The exceptions were the decomposition parameters of the SOM pools. For these, we used updated estimates from a recent Bayesian calibration (Gurung et al., 2020) that included data from most of the well-known long-term experiments in Europe and the US (and textures from 15-50% clay). To our knowledge Gurung et al. (2020) provide the most up-to-date decomposition parameters, and hence, we assigned the median of reported parameter values for each SOM pool as initial parameter values in our model simulations. Furthermore, for the parameters determining the minimum and maximum proportion of nitrified N lost as N_2O, we used the most recent values from Gurung et al. (2021), who showed that older parameters overestimate N_2O emissions. To identify which model parameters to include in the global sensitivity analysis (see section 2.4) and in the model calibration, we screened the literature for recently conducted sensitivity analyzes of the DayCent model (Necpálová et al., 2015; Gurung et al., 2020) and additionally used the DayCent manual to identify and add further parameters of potential importance for the processes considered in our study (i.e., plant productivity and soil C and N cycle). This resulted in a selection of 66 parameters (Table 1 and Table A3). Some of these parameters represented groups of the same type of parameters that can each be individually calibrated in DayCent, e.g., the "tillage multiplier" of SOM turnover can have different values for different SOM pools. However, because the tillage multipliers are usually the same for all SOM pools in the standard DayCent parameterization, we decided to have the same tillage multiplier value for active, slow, and litter pools. In addition, some of the parameters can have different values between the SOM pools of the surface and soil (for example, C / N ratios and turnover). For simplicity, we decided to assign the same C/N ratios and a constant ratio to the turnover of surface and soil SOM pools. Because the turnover rates of the SOM pools are typically faster on the surface than in the soil, we defined a new parameter that represents the value ratio of the surface to the soil parameters (i.e., decX(2)/decX(1)). This allowed us to jointly evaluate the sensitivity and calibrate parameters related to surface and soil SOM pools without adding too much complexity. Finally, the parameters for minimum and maximum values of nitrification and loss of nitrified N as N_2O, were reformulated as maximum and the difference between the minimum and maximum parameters (i.e., N2Oadjust_(max-min) and aneref(1)-anaref(2)). This was done to ensure that the maximum was always higher than the minimum value in order to avoid numerical problems.

2.3.3 Spin-up and site history simulation to initialize SOC and soil N contents

As is standard practice in DayCent, the initialization of SOM pools was conducted through a spin-up run, which was followed by a simulation of the history of the site before experiment establishment based on the knowledge of site managers. The spin-up has the aim to initialize the SOM pools to equilibrium using the typical input of biomass of the native vegetation type. The type of native vegetation for each site was determined from an available potential vegetation map (Kamoni et al., 2007) and confirmed by site managers as tropical evergreen forest in Embu, dry savanna in Machanga, and humid savanna in Sidada and Aludeka. A 2000-year spin-up simulation was sufficient to reach a steady state of SOM pools. Site managers had a good knowledge of the type of historical cropping systems, e.g., arable vs. grasslands, types of crop rotation (e.g., maize monoculture vs. crop rotation with legumes), manure inputs and management, but without detailed information on the duration

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

of these systems. Therefore, the duration of cropping systems after native vegetation was adjusted at each site so that the measured initial SOC stocks corresponded to the simulated SOC stocks at the start of the experiment. Additionally, to achieve suitable levels of soil N stocks after the spin-up, the maximum C/N ratio of the SOM pools had to be increased. It was increased from 14 to 20 for the active SOM pool and from 8 to 13 for the passive SOM pool (parameters varat12&11(1,1) and varat3(1,1), respectively). Due to computational time constraints and to ensure a match between simulated and observed initial SOC and soil N levels, the spin-up and site history simulations were not included in the sensitivity analysis and Bayesian calibration.

2.4 Global sensitivity analysis

To reduce the dimensionality of the calibration of the model, we performed a parameter screening (van Oijen, 2020). For this purpose, a global sensitivity analysis was conducted to quantify the relative importance of different model parameters to the relevant model outputs regarding our study focus on maize yield and ISFM greenhouse gas mitigation potential. The goal was to fix less influential model parameters to their default values, reducing the computational cost for performing the consecutive Bayesian calibration (see section 2.5). Global sensitivity analysis was performed using the Sobol method (Saltelli, 2002a, b), which allows for the estimation of the proportion of variance in the model outputs that is explained by each model parameter, considering the interaction terms of first-order and higher-order (Gurung et al., 2020). The "sensitivity" package (function sobolSalt; Iooss et al., 2021) of R version 4.0 (R Core Team, 2020) was applied. This function implements a simultaneous Monte Carlo estimation of first-order and total-effect Sobol indices. The computational cost is $N(p+2)$ model runs, N being the dimension of the two matrices to construct the Sobol sequence, p being the number of parameters (66 in our case). Our tests indicated similar results for N = 500/1000, so we chose a dimension of 1000. The preselection of model parameters to include (Tables 1 and A3) is described above. Independent uniform prior distributions were used for the global sensitivity analysis, with the upper and lower parameter boundaries centered around the default parameter value obtained as described below. We based the global sensitivity analysis parameter ranges on previous sensitivity analyses (e.g. Necpálová et al., 2015; Gurung et al., 2020), on plausible ranges reported in the DayCent manual and on how the parameters varied between different maize parameterizations. The parameters were then grouped by how large the ranges were. The parameters that had very small, small and moderate ranges were varied by $\pm$10, 25 and 50% from the default parameter value, respectively. For parameters with large and very large ranges, the upper/lower boundaries were the default parameter values multiplied/divided by 3 and 10, respectively. Additionally, we assumed that the maize parameters of the second highest production level (C5 in DayCent) would best represent the production levels in the experiment. The parameter sensitivity was independently determined for the mean maize grain yield, aboveground biomass, as well as for the SOC and soil N stocks at the end of the simulation period.

2.5 Combined Bayesian calibration of plant and soil model parameters

Joint Bayesian calibration of the sensitive DayCent parameters was performed using all available data on maize grain yield, aboveground biomass, and SOC stocks. The main reason for only using these data was that the yields, SOC stocks, and their trade-offs were the focus of this study. A second, technical reason, was that the creation and readout of daily simulations outputs, needed to match simulated and measured soil moisture content, mineral N content and /N_2O fluxes, slowed down the

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Table 1. DayCent model parameters and possible range of values. Displayed are parameters considered for calibration due to total sensitivity index > 2.5% (top) and with a total sensitivity index > 1% (bottom). The remainder of parameters (<1%) can be found in the supplementary (Table A3).

Parameter	Description	Possible ranges of values	Units	Initial value	Lower value	Upper value	Calibrated value[+]
Included in calibration (total sensitivity >2.5%)							
himax	Maximum harvest index for maize	moderate	$g\ g^{-1}$ (C)	0.40	0.20	0.60	0.52
ppdf(1)	Optimum temperature for growth of maize	very small	°C	30.00	27.00	33.00	29.10
ppdf(2)	Maximum temperature for growth of maize	very small	°C	45.00	40.50	49.50	48.15
prdx(1)	Potential aboveground production of maize	large	$g\ C\ m^{-2}\ langley^{-1}$	2.25	0.75	6.75	2.45
clteff(1,2&4)	Tillage multiplier for SOM turnover	large	unitless	10.00	3.33	30.00	18.10
aneref(3)	Min. impact of soil anaerobiosis on SOM turnover	fixed	unitless	1.00	0.20	1.00	0.73
dec4	Max. turnover rate of passive SOM pool	very large	$g\ g^{-1}\ yr^{-1}$	0.0035	0.00035	0.035	0.033
dec5(2)	Max. turnover rate of slow SOM pool	large	$g\ g^{-1}\ yr^{-1}$	0.10	0.03	0.30	0.19
fwloss(4)	Scaling factor potential evapotranspiration	moderate	unitless	0.75	0.38	1.125	1.02
pmco2(1&2)	C lost as CO_2 with metabolic litter turnover*	fixed	$g\ g^{-1}$ (C)	0.54	0.30	0.95	0.62
ps1co2(1&2) & rsplig	C lost as CO_2 with structural litter turnover*	fixed	$g\ g^{-1}$ (C)	0.50	0.40	0.99	0.89
Not included in calibration (total sensitivity <2.5% & > 1%)							
frtc(1)	C allocated to roots at planting, without stress	small	fraction of NPP	0.50	0.38	0.62	-
frtc(3)	Time after planting at which maturity is reached	small	number of days	90.00	67.50	112.50	-
pramn(1,2)	Min. aboveground C/N ratio at maturity	small	C/N ratio	62.50	46.88	78.12	-
hiwsf	Max. harvest index reduction with water stress	moderate	$g\ g^{-1}$ (C)	0.60	0.30	0.90	-
teff(1)	Temperature inflection point (effect on SOM turnover)	moderate	unitless	17.05	8.53	25.58	-
varat21&22(2,1)	Min. C/N ratio for material entering slow SOM pool	small	C/N	12.00	9.00	15.00	-
basef	Soil water of bottom layer lost via base flow	moderate	fraction H_2O	0.30	0.15	0.45	-
N2Oadjust_max	Proportion of nitrified N that is lost as N_2O	large	$g\ g^{-1}$ (N)	0.004	0.0013	0.012	-
MaxNitAmt	Maximum daily nitrification amount	large	$g\ N\ m^{-2}$	0.40	0.13	1.20	-

*(1 - microbial carbon use efficiency); [+]highest likelihood parameter set across all four sites

whole calibration by a factor of 5, which made Bayesian calibration unfeasible (one iteration would have taken more than three months on our virtual machine with 64 cores). Following (Gurung et al., 2020), model parameters that had a total sensitivity index of at least 2.5% for either yield, aboveground biomass, or SOC were considered influential and thus were altered in Bayesian calibration. This resulted in 11 parameters to be calibrated (Table 1).

Bayesian calibration is a probabilistic inverse modeling/data assimilation technique, which is used to estimate the joint posterior distribution of model parameters (θ) given the measured data (D) and the model structure (M), expressed as $p(\theta|D,M)$. It uses the proportionality form of Bayes‘ theorem, where $p(\theta|D,M)$ is proportional to the prior belief about model parameters, $p(\theta)$ times the likelihood function, $p(D,M|\theta)$:

$$p(\theta|D,M) \propto p(\theta) * p(D,M|\theta) \quad (2)$$

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

While the prior, $p(\theta)$, is chosen based on previous knowledge of the model parameters, the likelihood function, $p(D,M|\theta)$, measures how well the model and the data match. In practice it is derived for a given set of parameter sampled from the prior, by running and evaluating the model using the measured data, the simulated counterpart and the variance-covariance matrix of the data. We used the median variances per site for each type of measurement (computed from the three replicates) as the diagonal elements of the variance-covariance matrix. Due to the large number of observations and the mostly balanced dataset, the off-diagonal elements were set to 0.

The sampling importance resampling method, which was used in this study, is a direct form of Bayesian calibration, which has recently been used by Gurung et al. (2020), to calibrate the parameters of the SOM module of DayCent using a large collection of temperate long-term experiments. It samples the prior by running the model for a large sample of parameter sets of size *I* from the prior, computing the likelihood for each sample, and filtering the prior based on importance weights $w(\theta_z)$

$$w(\theta_z) = \frac{p(D,M|\theta_z)}{\sum_{i=1}^{I} p(D,M|\theta_i)} \tag{3}$$

where $p(D,M|\theta_z)$ is the likelihood function of the zth parameter set and $w(\theta_z)$ is the corresponding importance weight. It is consistent with the proportionality form of Bayes‘ theorem in that it uses the importance weights $w(\theta_z)$ as probabilities for sampling from the prior, without replacement, to derive the posterior. Overall, a total of 200000 simulations were performed, from which, as in 0.1% (200) of the parameter sets were sampled to derive the posterior distribution through resampling (Gurung et al., 2020). It was assured that this number of simulations was sufficient by splitting the simulations into two halves and visually assessing the similarity of derived posteriors for these subsets. In our experience, the sampling importance resampling algorithm is ideal for DayCent, which is prone to model crash when using inappropriate parameter combinations. This method does not depend on chains, but rather on model runs that are independent of each other. This means that an erroneous run does not stop the algorithm, as would be the case in chain-dependent methods such as Markov chain Monte Carlo. In addition, this strategy allows for an efficient cross-validation of the posterior parameter set.

To evaluate the model by comparing measured vs. modeled data, we conducted a leave-one-site-out cross-validation. This means that each of the sites is left out one by one for an independent evaluation, while the other three sites were used to compute the resampling weights and derive the posterior parameter distributions. Hence, the evaluation is representative of up-scaling exercises of the DayCent model to other sites, because it involves evaluation of the model performance across different sites. Here, the SIR algorithm was also advantageous. We stored the model results for each parameter set by site, which meant that the cross-evaluation could be done by a simple filter by site, without rerunning the model for each iteration. In contrast to model cross-validation, the parameter posterior distributions are not displayed by site. They are displayed for all four combined sites; hence the full dataset without leaving any site out was used to derive them. This was done to present the most representative posterior distributions and to make efficient use of all available data.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

2.6 Model evaluation

We used the following standard model evaluation statistics (Loague and Green, 1991):

$$MSE_y = \frac{1}{n}\sum_{z=1}^{n}(O_{yz} - P_{yz})^2 \tag{4}$$

$$RMSE_y = \sqrt{MSE_y} \tag{5}$$

$$EF_y = 1 - \frac{\sum_{z=1}^{n}(O_{yz} - P_{yz})^2}{\sum_{z=1}^{n}(O_{yz} - \bar{O}_y)^2} \tag{6}$$

Here, MSE_y is the mean-squared-error and *RMSE* is its root. EF_y is the Nash-Sutcliffe modeling efficiency. O_{yz} is the measured value of the *z-th* measurement of the *y-th* type of measurement, $\bar{O}_y$ the mean of the *y-th* type of measurement and P_{yz} the simulated value corresponding to O_{yz}. We further divided MSE_y into squared bias (*SB*), nonunity slope (*NU*) and lack of correlation (*LC*), as suggested by Gauch et al. (2003). We expressed them as a percentage of the MSE_y:

$$SB_y(\%) = \frac{(\bar{O}_y - \bar{P}_y)^2}{MSE_y} * 100 \tag{7}$$

$$NU_y(\%) = \frac{(1 - b_y)^2 * (\frac{\sum_{z=1}^{n}(O_{yz}^2)}{n})}{MSE_y} * 100 \tag{8}$$

$$LC_y(\%) = \frac{(1 - r_y)^2 * (\frac{\sum_{z=1}^{n}(P_{yz}^2)}{n})}{MSE_y} * 100 \tag{9}$$

Here, $\bar{P}_y$ is the mean predicted value of the *y-th* measurement type, *b* the slope of the regression of *P* on *O* and *r* the correlation coefficient between *O* and *P*. The indicators *LC*, *SB* and *NU* show the nature of model errors, that is, a high *LC* shows that it is mostly random, a high *SB* a systematic bias, while a high *NU* shows issues of model sensitivity.

2.7 Net global warming potential

To compare different ISFM treatments in terms of their greenhouse gas emissions, their net global warming potential (GWP) on a yearly basis (CO_2eq ha^{-1} yr^{-1}) was derived from the outputs over the whole simulation period. It was calculated from changes in the SOC content and cumulative emissions of N_2O using 100-year time horizon of global warming potentials (Necpalova et al., 2018):

$$GWP = \frac{44}{12} * \Delta SOC + 265 * N_2O \tag{10}$$

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

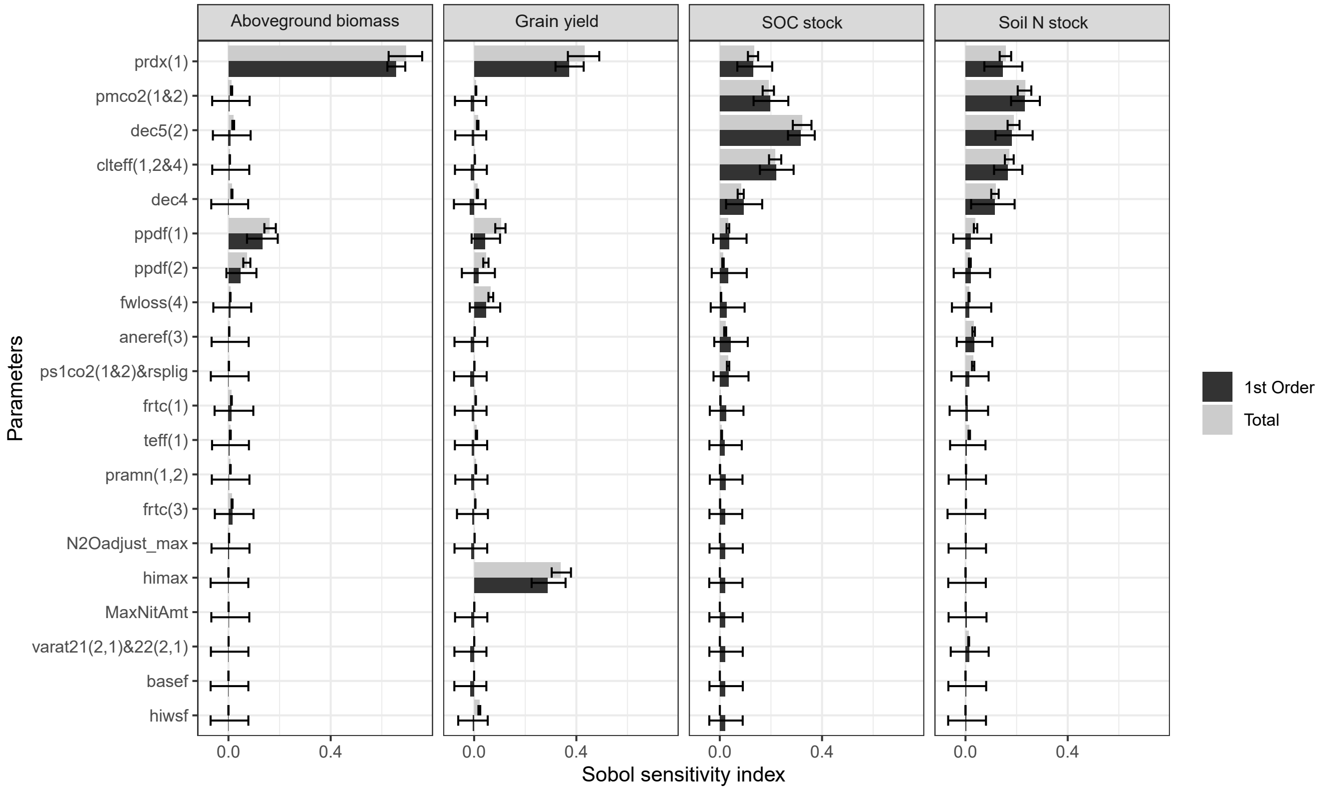

Figure 1. Results of the uncertainty-based global sensitivity analysis of the most relevant DayCent model parameters. Parameter sensitivity was independently determined for the mean maize aboveground biomass, grain yield, and SOC and soil N stocks at the end of the simulation period. Only parameters with a Sobol sensitivity index >1% are displayed.

Here, ΔSOC is the change in SOC content (kg C ha^{-1} yr^{-1}), N_2O the cumulative N_2O flux (kg N_2O ha^{-1} yr^{-1}). The CH_4 oxidation capacity was not considered, because it usually makes a very limited contribution to GWP in rainfed maize cropping systems and we did not have data to evaluate the reliability of this simulated flux. In addition to GWP we calculated yield-scaled GWP (in CO_2eq kg^{-1} maize grain yield) by dividing the cumulative GWP over the entire simulation period by cumulative simulated yields (dry matter base).

3 Results

3.1 Most sensitive DayCent parameters

The results of the global sensitivity analysis showed that of the 66 model parameters included, only 20 parameters had a Sobol total sensitivity index >1% for either maize grain yield, aboveground biomass, SOC or soil N stocks (Fig. 1). Of these, only 11 parameters had a Sobol total sensitivity index >2.5%, a threshold that captures the most influential parameters and represents

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

a suitable selection of parameters for model calibration (Gurung et al., 2020). The parameters that turned out to be the most sensitive, with a Sobol total sensitivity index >10% for at least one type of measurement were radiation use efficiency (prdx(1); for all measurement types); the optimal and maximum temperature for maize growth (ppdf(1) and ppdf(2), respectively; only for grain yield of maize and aboveground biomass), and maximum harvest index (himax; only for crop yield). Further, the turnover rate of the slow and passive SOM pools (dec5(2) and dec4, respectively; only for SOC and soil N), the decomposition multiplier after tillage events (clteff(1,2&4); only for SOC and soil N) and the fraction lost as CO_2 upon metabolic litter pool turnover (pmco2(1&2), i.e., 1 - microbial carbon use efficiency (CUE); only for SOC and soil N) belonged to the most sensitive model parameters. The parameters of further importance, with a Sobol total sensitivity index <10% and >2.5%, were the minimum value for the impact of anaerobic soil conditions (aneref(3); only for SOC and soil N), the scaling factor for potential evapotranspiration (fwloss(4); only for maize grain yield), and the fraction lost as CO_2 upon structural litter and lignin turnover (ps1co(1&2)&resplig, i.e., 1-CUE; only for SOC and soil N). The fact that the Sobol 1st order and total sensitivity indexes were similar for most parameters suggested only a limited number of interactions between the parameters identified by the global sensitivity analysis.

3.2 Posterior parameter distributions from the Bayesian model calibration

Following the global sensitivity analysis, 11 selected parameters were calibrated using the same ranges of possible values as defined for the sensitivity analysis. Four parameters were fully constrained by the Bayesian calibration, one was partly constrained and tended towards the upper boundary. Four showed a tendency, but values from the whole prior range were present in the posterior distribution. Finally, two showed almost no difference to the uniform prior distribution (Fig. 2). Very clearly constrained were the potential maximum maize production and the maximum maize harvest index (2.5 g C m^{-2} $langley^{-1}$ for prdx(1) and 0.48 g g^{-1} for himax), with values slightly higher than the default values and somewhere in between the maize with the highest and second highest production levels (default DayCent maize parameters, named C6 and C5). Furthermore, the scaling factor for potential evapotranspiration was clearly constrained and slightly higher than the default value (fwloss(4), 0.9 vs 0.75). The turnover rate of the slow SOM pool was also clearly constrained, but was centered around a value twice the default (0.2 vs 0.1 g g^{-1} yr^{-1} for dec5(2)). The turnover rate of the passive SOM pool was clearly constrained at the lower end of possible values, it was centered around a much higher value than the default value (0.03 vs 0.0035 g g^{-1} yr^{-1} for dec4) and tended towards the upper boundary. A test with even broader ranges (up to 0.1 g g^{-1} yr^{-1}) showed that the value around 0.03 g g^{-1} yr^{-1} was in fact the center of the posterior distribution of this parameter (Fig. A1), but allowing the parameters to vary that much reduced the model performance to unfeasible levels (model output not shown). The optimal production temperature for maize (ppdf(1)) tended to have values lower than the default value, while the opposite was true for the maximum production temperature (ppdf(2)). Also the parameters representing CO_2 loss upon turnover of metabolic and structural litter pools (pmco2(1&2) and ps1co2(1&2)&rsplig) tended towards higher values than the default values of these parameters (i.e., lower carbon use efficiencies, because the parameters represent 1-CUE), but they were not clearly constrained. Finally, the parameters representing the increase of the SOM pools turnover after tillage (clteff(1,2,&4)) and the maximum rate limitation of soil under anaerobic conditions (aneref(3)), were poorly constrained.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

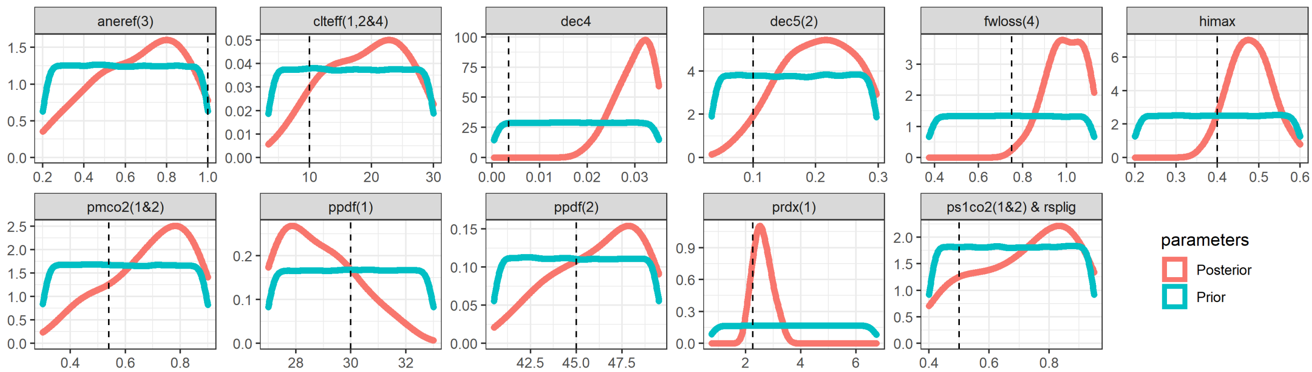

Figure 2. Prior compared to the posterior model parameter distribution resulting from the uncertainty-based Bayesian model calibration of DayCent. Dashed vertical lines represent the values of the default parameter set. The posterior distributions are based on all four study sites combined. For the description of the parameters see Table 1.

Only a few strong correlations existed between the parameters of the posterior parameter set (Fig. A2). Namely, there was a strong positive correlation (r = 0.63) between the potential maximum production of maize (prdx(1)) and the optimal temperature for maize growth (ppdf(1)), the latter also being positively correlated with the maximum temperature for maize growth (ppdf(2); r = 0.45). Furthermore, there was a negative correlation (r = -0.43) between the effect of tillage on SOM turnover and the amount of CO_2 loss upon turnover of the metabolic litter pool (pmco2(1&2)). The other correlations of the parameters were weak (i.e., below ± 0.4) and therefore were of low importance.

3.3 Simulation of maize grain yields and aboveground biomass at harvest

Although the overall variation of maize grain yields between sites and treatments could be captured with the set of default model parameters, the comparison of model results obtained with the default parameters compared to the set of parameters chosen for each site by leave-one-site-out cross-validation showed that the Bayesian calibration significantly improved the fit of the model for both maize grain yield and aboveground biomass (Fig. 3). The calibration improved the simulation of grain yield for all sites and for aboveground biomass for all sites except Machanga (Fig. A9). It should be noted that despite excluding the evaluation site in the calibration step, the calibrated model was unbiased and errors mostly random for both yield (i.e., lack of correlation (LC) increased from 90% to 97%) and aboveground biomass (LC increased from 71% to 91%).

Furthermore, while DayCent was inclined to overestimate the lowest and underestimate the highest values of yields and aboveground biomass, the mean yields and aboveground biomass per treatment throughout the simulation period were simulated well for most treatments (Fig. A4). The exception to this was that DayCent poorly distinguished the mean yields and aboveground biomass of treatments with high compared to very high rates of N inputs (i.e., differences between the different organic resources and the control within the +N treatment). An additional test of the model sensitivity of mean yields to different levels of mineral N fertilizer in the control provided further insights into this (Fig. A6). In this test, the yields stopped

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

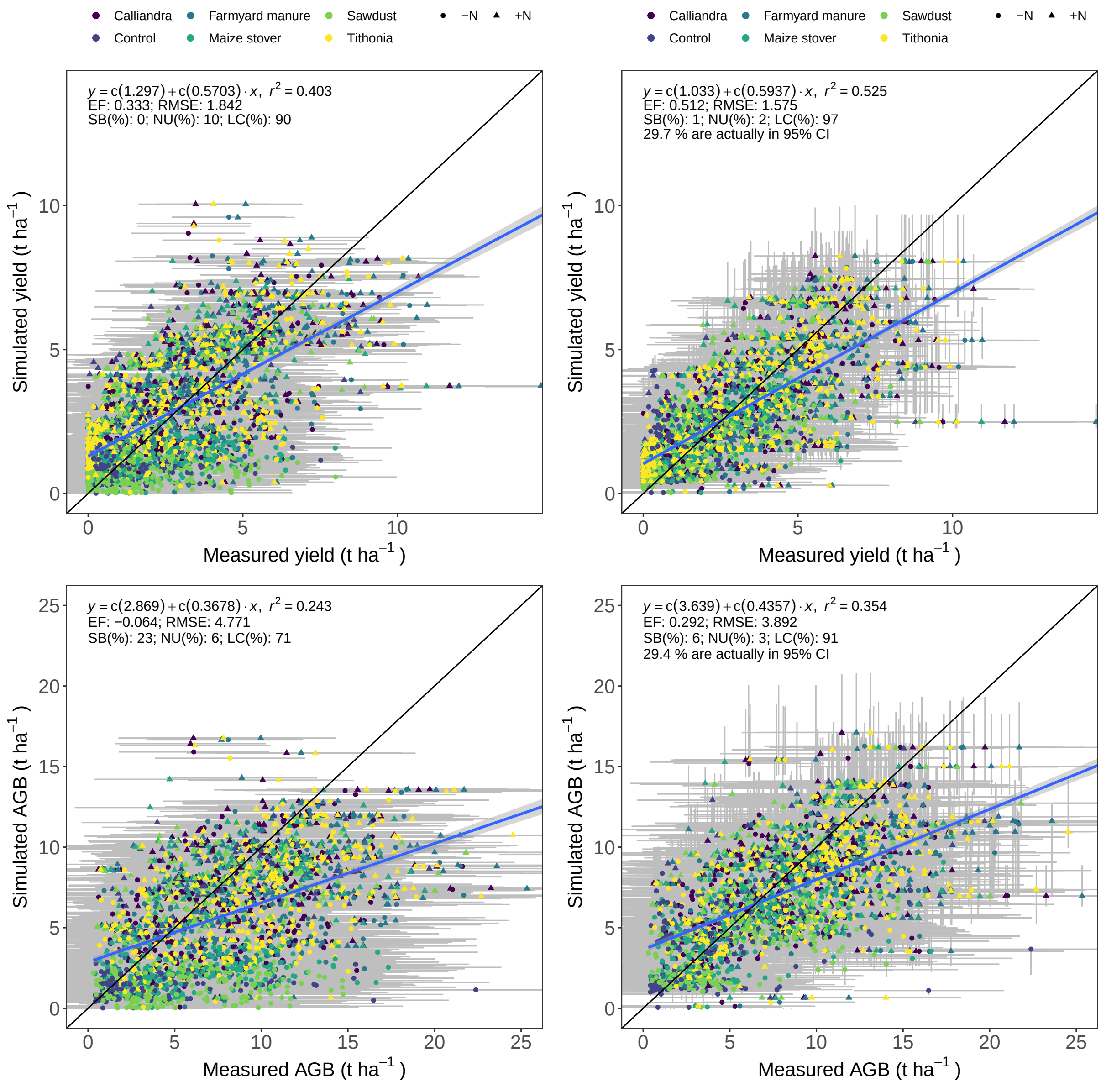

Figure 3. Simulated compared to measured maize grain yields at the four study sites for the default DayCent parameter set (top-left) versus the calibrated parameter set by leave-one-site-out cross-validation (top-right). The same is displayed for maize aboveground biomass (AGB), showing the default DayCent parameter set (bottom-left) versus the calibrated parameter set (bottom-right). The 2985 data points correspond to the observations from the experimental treatments over 32 to 38 seasons, depending on the site. Symbols represent the different organic resource and chemical nitrogen fertilizer treatments. Grey bands show the 95% confidence intervals of measured (horizontal) values and the 95% credibility intervals of posterior distribution (vertical). Abbreviations: EF, Nash-Sutcliffe modeling efficiency; RMSE, root mean squared error; SB, squared bias; NU, non-unity slope; LC, lack of correlation.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

increasing at mineral N rates that were lower than the maximum N rates provided in the organic resource +N treatments by mineral N and organic resources combined (up to >500 kg N per year or > 250kg N per season). In Machanga and Embu, simulated mean yields stopped increasing at around 100 kg N ha^{-1} per season, which is less the 120 kg N ha^{-1} per season that the control +N received. In Aludeka and Sidada, simulated mean yields stopped increasing at 200 to 250 kg N ha^{-1} per season, but most of the response to N was below 120 kg N ha^{-1} per season (Fig. A6).

Although the mean yields of the high-quality inputs in the -N treatments were well simulated, some of the low-quality input treatments in Aludeka and Sidada, namely maize stover at 1.2 t C ha^{-1} yr^{-1} , and sawdust at 1.2 and 4 t C ha^{-1} yr^{-1}, had lower simulated yields than the observed mean yields in their -N treatments (Fig. 4). The same was true for the control -N in Aludeka, but the yields for sawdust -N at 4 t C ha^{-1} yr^{-1} in Machanga were overpredicted compared to measurements. Interestingly, the 95% credibility intervals for yield and aboveground biomass produced by the leave-one-site-out cross-validation contained only about 30% of measured data and also tended to be considerably smaller than variance-based 95% confidence intervals for the measured data.

3.4 Simulated SOC stocks in response to integrated soil fertility management

In contrast to the simulation of maize yields, the change in SOC stocks following the application of organic resources at different rates (1.2 and 4 t ha^{-1} yr^{-1}) was poorly simulated by DayCent with the set of default parameters. The simulated SOC losses were too low compared to the observations (Fig. 5). The posterior parameter set obtained through Bayesian calibration with strongly increased SOM turnover rates showed a tendency for a higher loss of CO_2 from the turnover of metabolic and structural litter pools. With the posterior parameter set, the model, in the leave-one-site-out cross-validation, simulated the change in SOC much more accurately than with the default parameter set (EF of 0.54 vs -1.3; LC of 88% vs 42%; Fig. 5). The 95% credibility intervals of the simulated values contained 47% of the measured data. It should be noted that under- or overestimation of the change in SOC stocks from the start to the last year of the experiment was rather related to site than to treatment. In Embu, for example, all treatments except the control +N treatment tended to have lower observed than simulated SOC losses, while in Aludeka most treatments except the control -N treatment showed weaker predicted than observed SOC losses (Fig. A3). These tendencies are also evident when comparing the temporal dynamics of measured versus simulated SOC stocks (Fig. 6). Most of the treatments were well simulated, but it is noteworthy that there is considerable variability in the measured SOC stocks between different time points.

3.5 Simulated N_2O emissions and global warming potential

The simulations and measurements of N_2O emissions on a daily basis were not well aligned. This is clear from the negative modeling efficiencies and the absence of a clear correlation between observed and measured values (Fig. 7). Peaks of N_2O emissions were often simulated on different dates than the measurements, but treatments with higher N loads tended to have higher simulated and higher measured N_2O fluxes. This was most noticeable in +N compared to -N treatments (Fig. A10). For cumulative N_2O emissions per season on the other hand, there was a much better alignment between the simulated and measured values. All sites showed positive modeling efficiency (highest EF 0.34 for Aludeka, lowest EF 0.21 for Machanga;

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Figure 4. Barplots of mean simulated and mean measured yield and aboveground biomass (AGB) from cross-validation. Error bars represent standard deviation.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

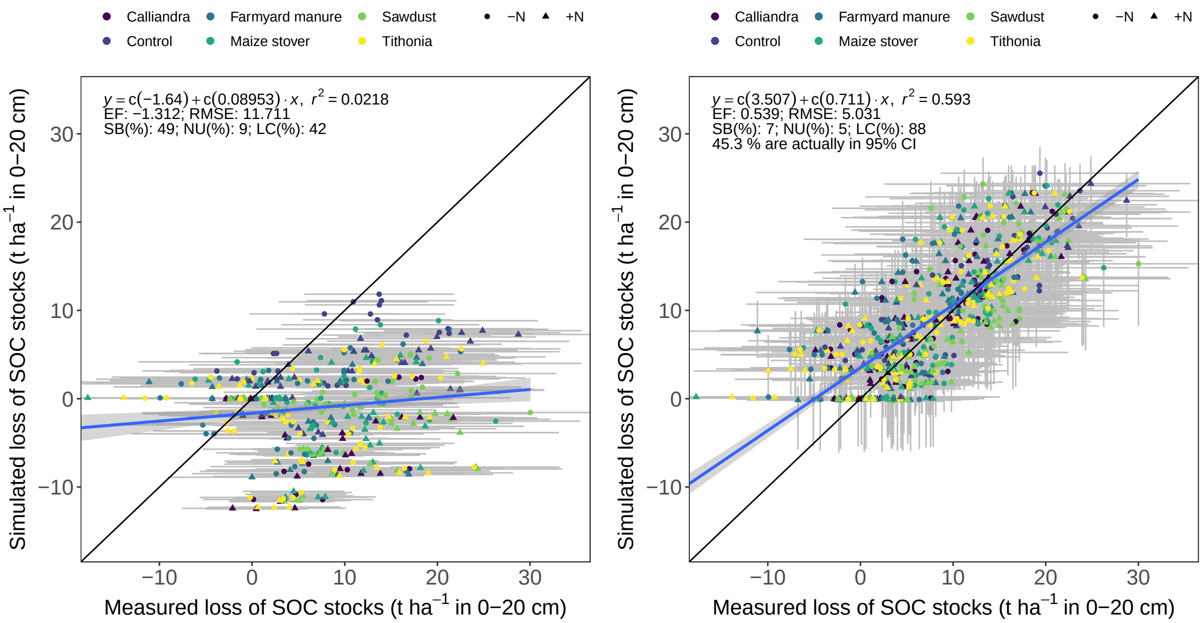

Figure 5. Simulated compared to measured changes in SOC stocks since the start of the experiment at the four study sites for the default DayCent parameter set (left) versus the calibrated parameter set by leave-one-site-out cross-validation (right). The 724 data points correspond to the observations from the experimental treatments over 32 to 38 seasons, depending on the site. Symbols represent the different organic resource and chemical nitrogen fertilizer treatments. Grey bands show the 95% confidence intervals of measured (horizontal) values and the 95% credibility intervals of posterior distribution (vertical). Abbreviations: EF, Nash-Sutcliffe modeling efficiency; RMSE, root mean squared error; SB, squared bias; NU, non-unity slope; LC, lack of correlation.

Fig. 7). Additionally, the correlation between cumulative simulated and measured N_2O emissions was much higher than for daily emission fluxes (R^2 between 0.26 for Sidada and 0.81 for Aludeka, compared to R^2 close to 0).

The simulated changes in SOC and simulated cumulative N_2O emissions, showed positive global warming potentials (GWP) for all treatments (Fig. 8). No treatment acted as a greenhouse gas sink, but the amount of emissions, as well as the contribution of N_2O and CO_2 differed strongly between sites and treatments. In the control -N treatment, it ranged from 1.4 t CO_2 equivalent ha^{-1} yr^{-1} in Aludeka to 5.1 t CO_2 equivalent ha^{-1} yr^{-1} in Sidada and Embu. The relative contribution of N_2O from site to site also differed strongly by site. In Aludeka, for example, between 32 (CT+N) and 82% (FYM4+N) of the simulated GWP in the +N treatments came from N_2O, while at all other sites, none of the treatments had more than 25% of GWP associated with N_2O emissions. Compared to the control -N treatment, which is closest to the low-input agriculture practiced by most smallholder farmers, all organic resource input treatments in the -N treatments were projected to have lower emissions (Fig. 8). The reductions ranged from -0.2 t CO_2 equivalent ha^{-1} yr^{-1} to 1 t CO_2 equivalent ha^{-1} yr^{-1}. Apart from that, the only site where the addition of mineral N (+N treatment) could lead to a higher simulated GWP per ha than the control -N treatment was Embu, and this was only the case for the treatments of *Calliandra, Tithonia* and farmyard manure in both 1.2 and 4 t of C

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

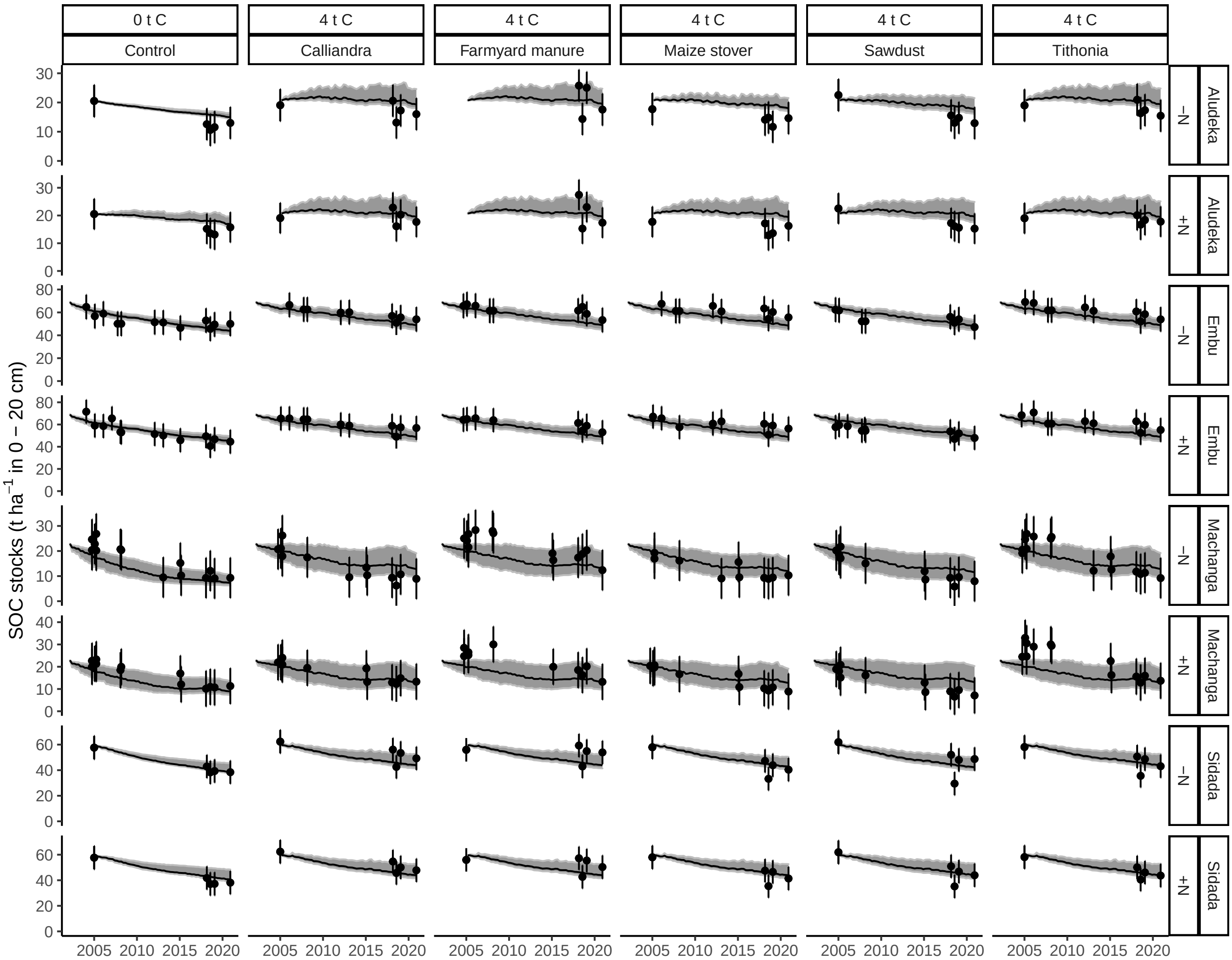

Figure 6. Measured (dots) versus simulated SOC stocks over time at the four study sites for the different organic resource and chemical nitrogen fertilizer treatments. Error bars represent 95% confidence intervals for measured data, the black solid line the simulation by the best parameter set for each site. Grey bands represent the 95% credibility intervals of the model posterior simulations, calibrated by leave-one-site-out cross-validation.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

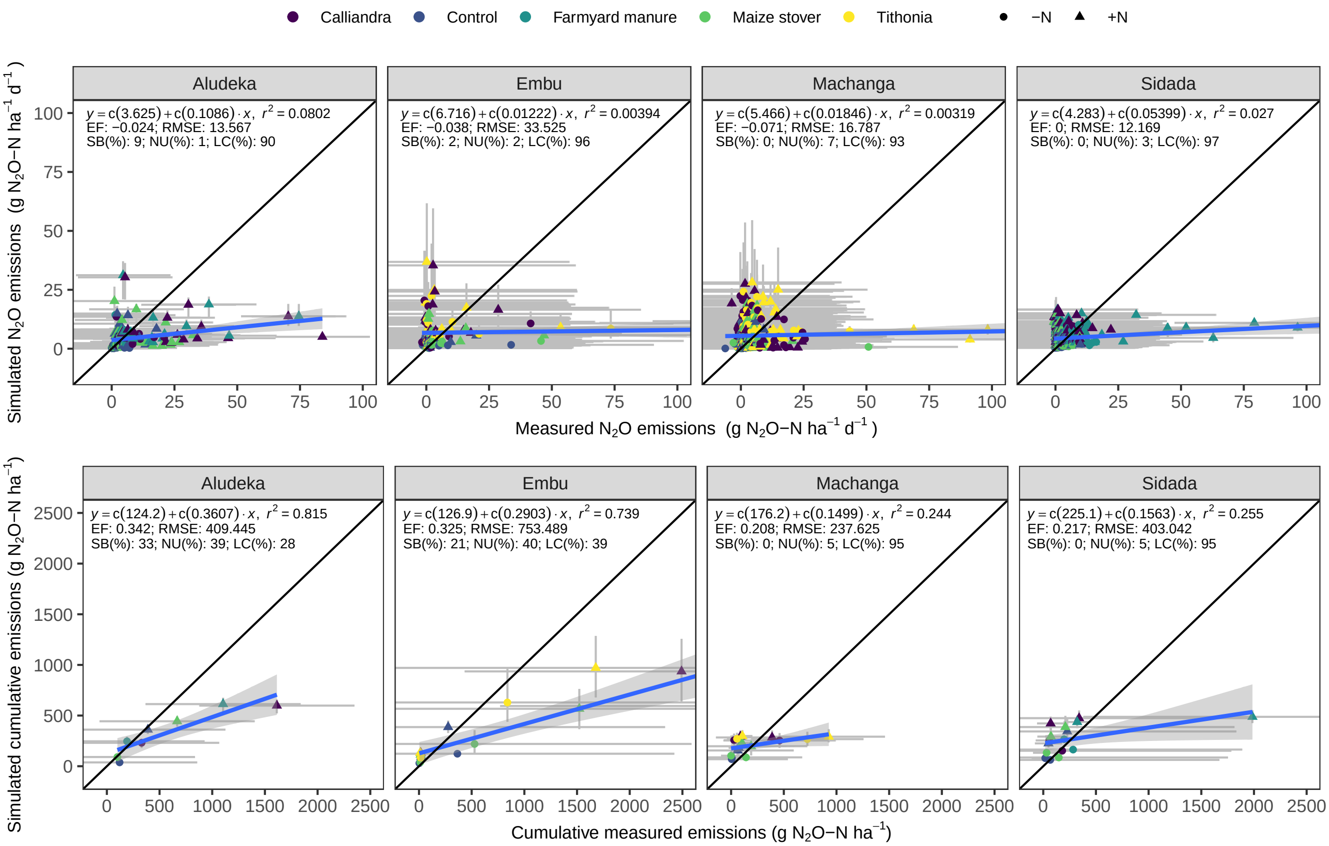

Figure 7. Simulated compared to measured N_2O emissions at the four study sites for the different organic resource and chemical nitrogen fertilizer treatments, based on the calibrated parameter set using leave-one-site-out cross-validation. Displayed are the measured versus modelled per treatment for the days where measurements were conducted (top) and for the mean of cumulative flux measurements per season using the trapeziod method (bottom). The 808 data points (top) correspond to the daily measurements from the experimental treatments over one to two seasons, depending on the site. Symbols represent the different organic resource and chemical nitrogen fertilizer treatments. Error bars represent 95% confidence intervals based on BC (simulations) and variance (measurements). Abbreviations: EF, Nash-Sutcliffe modeling efficiency; RMSE, root mean squared error; SB, squared bias; NU, non-unity slope; LC, lack of correlation.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

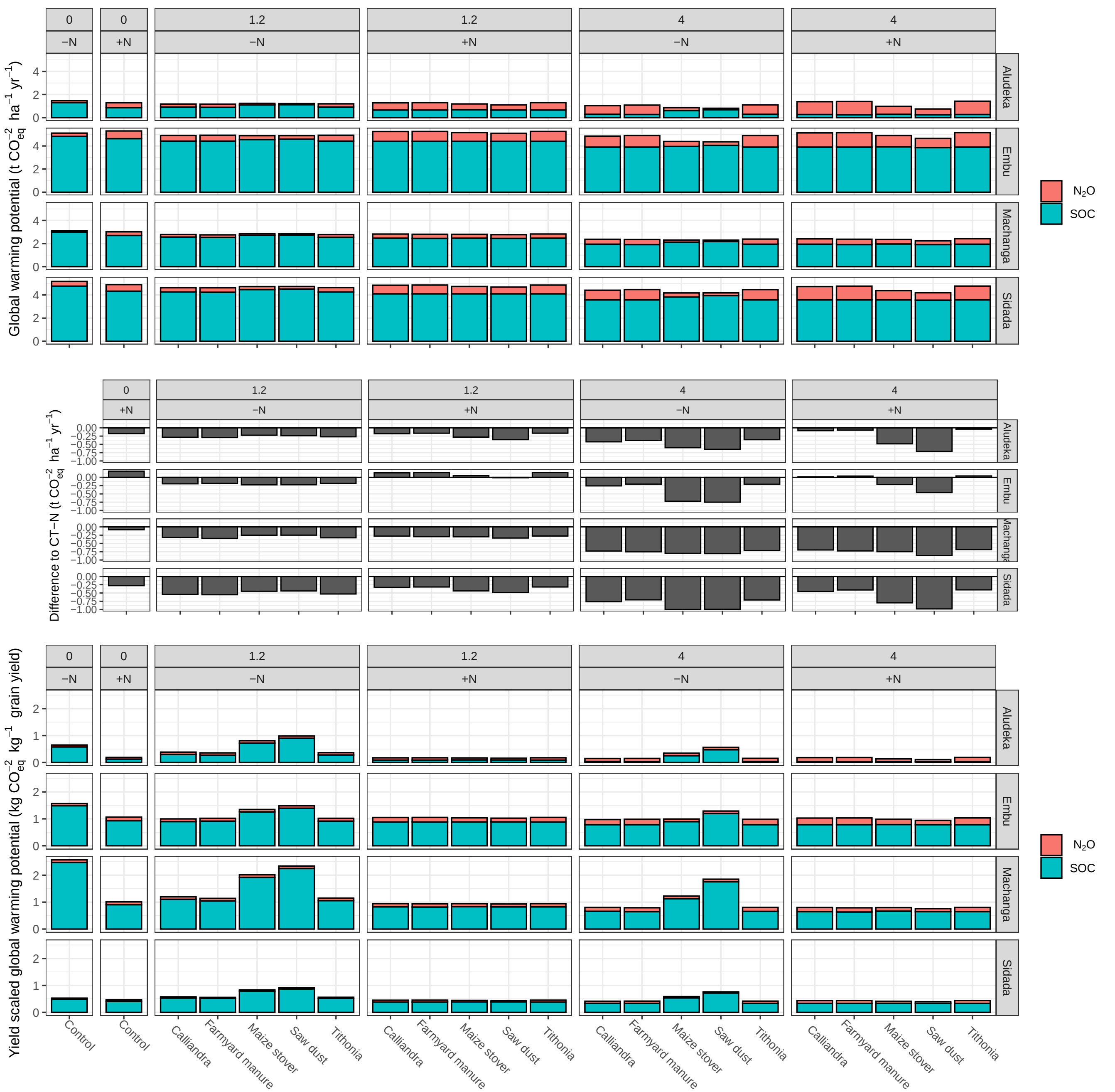

Figure 8. Cumulative simulated global warming potential of N_2O emissions and CO_2 emissions due to loss of SOC at the four study sites for different organic resource and chemical nitrogen fertilizer treatments, combined throughout the simulated period (16 years for Aludeka/Sidada; 19 years for Embu/Machanga). The global warming potential is expressed in CO_2 equivalent over a 100-year horizon.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

input ha^{-1} yr^{-1}. Furthermore, Embu and Machanga tended to have high simulated emissions per kg of maize yields (0.9 to 1.6 and 0.8 to 2.5 kg CO_2 equivalent per kg of yield, respectively, due to lower yield). On the contrary, the simulated emissions ranged between 0.1 and 0.7 kg CO_2 equivalent per kg of yield in Aludeka and between 0.4 and 0.9 kg CO_2 equivalent per kg of yield in Sidada. Notably, none of the +N treatments in Sidada and Aludeka had higher emissions than 0.5 kg CO_2 equivalent per kg of yield.

4 Discussion

4.1 Robustness of the Bayesian calibration shown by cross evaluation

As shown by our leave-one-site-out cross-validation (Figs. 3 and A7), the Bayesian calibration considerably improved the capability of DayCent to predict maize grain yield, aboveground biomass and changes in SOC stocks at the four long-term ISFM experimental sites. The model evaluation statistics from this calibration exercise were comparable to those of other recent publications that combined the predictions of yield and SOC (Necpalova et al., 2018; Levavasseur et al., 2021; Nyawira et al., 2021). However, these studies generally showed a better simulation of crop yield than SOC, which is in contrast to our results, where crop yield and SOC were simulated equally well. Of the three types of measurements, simulated changes in SOC stocks related to different treatments were most improved after model calibration, reducing the strong bias towards simulated gains in SOC, where instead losses were measured. In general, the screening of the model parameters with the global sensitivity analysis proved valuable in reducing the complexity of the calibration problem. The most sensitive parameters (Table 1) were well constrained, while the least sensitive parameters (aneref(3), ppdf(2) and ps1co2) were difficult to constrain. This suggests that a reasonable threshold for including a parameter in model calibration may even be greater than a total Sobol sensitivity index of 2.5%. This result is similar to the study of Gurung et al. (2020), where the parameters with a total sensitivity index <10% were also poorly constrained. In our study, it was possible to constrain the most sensitive parameters (prdx(1), himax, dec(5)); exceptions were the two parameters pmco2(1&2) and clteff(1,2&4), but their correlation (r = -0.43) is a reasonable explanation of why they could not be well constrained despite high sensitivity. Possibly, the algorithm cannot distinguish whether organic resources are poorly stabilized due to tillage disturbance, a fast turnover rate of the slow SOM pool, or a low carbon use efficiency of metabolic and/or structural litter pools. Such a correlation between SOM stabilization and turnover parameters is common in soil models (e.g. Ahrens et al., 2014; Laub et al., 2021), especially if the SOM pools do not represent measurable fractions (Laub et al., 2020).

Robust predictions of crop yields and changes in SOC stocks, as achieved by our cross-validation, are an important basis for potential future upscaling exercises of crop yield and SOC model predictions to national scales. While many model calibration exercises often display an overly optimistic picture of the capacity of models to represent SOC dynamics, by displaying how well the models represent absolute levels of SOC (e.g., in Nyawira et al., 2021; Ma et al., 2022; Levavasseur et al., 2021), the changes in SOC since the start of an experiment, shown in this study, are much more robust evaluation. Since models are typically parameterized to fit the observed SOC stock at the start of the experiment, absolute SOC will always closely resemble measurements, with the possible exception of experiments that run for many decades. The fallacy of this approach was shown

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

by Necpalova et al. (2018); Gurung et al. (2020), where the absolute SOC overoptimistically masks the bias of change in SOC stocks. This is also the case for the present study, as is clear from the results of the model simulations for the absolute SOC stocks compared to the changes in the SOC stocks (Fig. A7). Since our calibration shows a good fit and is free from serious bias even for changes in SOC stocks (that is, 88% of the errors are LC and not systematic; EF is 0.55), the calibrated DayCent model can be used in a robust manner to estimate the change of SOC stocks under different ISFM management in Kenya. Furthermore, because the model evaluation and calibration were performed at different sites, it seems reasonable to use DayCent for other sites with similar climate and soil conditions, even beyond Kenya. In that respect, the leave-one-site-out cross-validation is using the data most efficiently: for evaluation, we leave one site out, but for further model upscaling exercises, ideally the full posterior model parameter set including all sites (Fig. 2) should be used. A computationally less expensive alternative is to use only the single parameter set with the highest likelihood (Table 1).

4.2 Bayesian calibration suggests that SOC is lost at higher rates in tropical than in temperate soils

To estimate the potential yield and long-term sustainability of cropping systems without major bias using biogeochemical models, regionalized model calibrations are needed (Rattalino Edreira et al., 2021; Yang et al., 2021). Therefore, while previous studies have simulated crop productivity under ISFM and similar practices with default parameter sets (e.g. Nezomba et al., 2018; Nyawira et al., 2021), the results of our study indicate that especially to estimate the effect on SOC, local calibration is very much needed (Fig. 5). In fact, the turnover of the slow and passive SOM pools was approximately two and eight times higher than in a recent Bayesian calibration of DayCent for temperate systems (Gurung et al., 2020), indicating a faster SOM turnover under tropical conditions, as expected. However, since the SOM pools in DayCent do not correspond to measurable fractions and we did not have additional data (e.g., ^{14}C), constraining the related parameters is expected to be subject to high uncertainty (e.g. Ahrens et al., 2014). Nevertheless, our results of high SOM turnover in Kenya are in line with the results of two recent studies, which also found that the default parameterizations of the DayCent (Nyawira et al., 2021) and the LPJGUESS model (Ma et al., 2022) underestimated the loss of SOC in two other long-term experiments in Kenya. In our study, the low Sobol total sensitivity indices of the maize productivity parameters for the SOC stocks suggest a limited importance of root input in the soil system for the storage of SOC. These results align with a statistical analysis of SOC stocks across depth at the same experimental sites (Laub et al., 2022), which found limited differences between cropped and bare plots, and with another recent tropical long-term experiment (Cardinael et al., 2022), which estimated that less than 1 t of carbon input ha^{-1} per year came from maize roots.

Overall, the considerably higher simulated SOM turnover rates in this study compared to those in temperate regions (Gurung et al., 2020) suggest that the lower SOM stabilization potential of highly weathered minerals in tropical soils (Doetterl et al., 2015) is not adequately accounted for in DayCent. DayCent only includes an effect of clay content. Possibly, the effect of weathered minerals is large for both slow and passive SOM pools (increased turnover by 2^1 and 2^3 in this study, respectively) and is not adequately represented by texture alone. It was recently shown that the Millenial model, which has measurable pools, could better simulate the global SOC distribution than the CENTURY model, which has conceptual pools (Abramoff et al., 2022). Recent attempts to move DayCent in that direction have also shown some promising results (e.g. Dangal et al.,

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

2022). Furthermore, it is possible that hand tillage twice a year and potential erosion also played a role in the high SOC losses that informed these fast turnover rates. Because DayCent does not include erosion and erosion was not measured in the experiments, the exact role of these processes is hard to determine. Therefore, the different calibrations of the DayCent model in our study and in Gurung et al. (2020), suggest that the calibration developed in this study should only be used for soils of similar mineralogy and under a similar climate and for a similar maize cropping system.

4.3 Maize crop module is more universally applicable than soil carbon module

Although the mean yields per treatment were well represented and the provisioning of N via organic resource mineralization was well captured, as demonstrated by a good agreement between the mean simulated and measured yields in the -N treatments (Fig. 4), the improvement in the simulated maize grain yield through the joint Bayesian model calibration of SOM pools and the maize crop highlights the interconnectedness of both. Thus, even if the crop parameterization is correct, as indicated by the low difference between initial and calibrated maize parameters, too low SOM pool turnover rates, by underestimating mineralization of N from soil, results in too high demand of maize for mineral N addition to produce a suitable yield. This can be seen in the poor simulation of the low input treatments of Sidada before calibration and their improvement after calibration (Fig A5). Interestingly, in contrast to SOC turnover rates, the values of the maize crop parameters in this study were close to the set of default parameter values of DayCent, especially the maximum production capacity (prdx(1)). This suggests that the traits of the maize are universally represented in DayCent, whereas the turnover of the SOM is not. In this context, the changes in the values for optimal and maximum temperature for maize growth (ppdf(1) and ppdf(2)) could be attributed to the adaptation of local maize species to the higher temperatures in Kenya. For example, Yang et al. (2021) conducted region-specific Bayesian model calibration of DayCent maize growing parameters and found ppdf(1) to very between 26 and 32 °C. Therefore, the different optimal temperatures (33° C) and highest temperatures (40° C) for the calibration of maize to US conditions by Necpálová et al. (2015) do not contradict the range of parameters of our study.

An interesting observation is that the model bias for the mean yield of maize appears to be treatment specific (i.e., the mean yields of +N treatments of farmyard manure at 4 t C ha^{-1} yr^{-1} at all sites and of *Tithonia* at the same rate in all but Sidada, were underpredicted by DayCent), while the bias for SOC stocks was mostly site specific (i.e., SOC loss in Aludeka was underpredicted, while overpredicted in Embu). As discussed above, the site bias of SOC is likely due to the fact that soil texture alone is insufficient to explain the mineralogy-driven storage potential of SOC (e.g. Reichenbach et al., 2021; Mainka et al., 2022). On the other hand, our sensitivity test to mineral N input suggests that the yield bias at high N is due to DayCent's inability to capture yield increases above 100-150 kg N per ha and season (Fig. A6); the +N treatments of *Tithonia*, *Calliandra* and farmyard manure at 4 t C ha^{-1} yr^{-1} supplied on average >250 kg N per ha and season. Another reason may be that DayCent does not include other potential beneficial effects of treatments, such as increased pH from farmyard manure application (Xiao et al., 2021; Mtangadura et al., 2017) and higher nutrient retention and water infiltration of treatments that maintained SOC stocks compared to those that showed a decline of SOC.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

4.4 N_2O emissions and global warming potential

In general, the poor match between daily observed and measured N_2O emissions (Fig. A10) illustrates the difficulty of simulating the timing of microbial processes, in which nitrate (NO_3^-) is converted to N_2 and N_2O gasses, with intermediate complexity models such as DayCent. Several recent studies show that poor simulation of N_2O emissions by process models is common (Zhang and Yu, 2021; Wang et al., 2020), which is attributed, among other factors, to a poor representation of soil hydraulics. For example, Sommer et al. (2016) found mainly overestimated N_2O emissions in Kenya. Gaillard et al. (2018) reported that there is a bias in simulated N_2O emissions by most models, postulating an underestimation of strong N_2O flushes. This also seems to be the case in our study, at least for treatments with high N inputs such as farmyard manure +N. Without resorting to more intricate water flow equations, it could be difficult to make further improvements to the simulation of soil water content and N_2O emissions, particularly in terms of accurately predicting daily and hourly events. DayCent is driven by inputs that are only resolved at daily time steps and fed with pedotransfer functions that are poorly represented in the tropics (Van Looy et al., 2017). However, the fact that cumulative N_2O emissions were better captured than daily emissions and that there was no systematic under- or over-prediction of cumulative N_2O emissions, does suggest that simulated N_2O emissions were generally reasonably well predicted with this current DayCent calibration. This is important for the predictions of the GWP. Because the simulation of SOC change showed low bias, we can conclude that this part of the GWP is well represented. Depending on the site and treatment, N_2O emissions contributed between 80% (Aludeka) and up to 20% of the GWP (other sites; Fig. 8). However, the larger confidence intervals of the measured compared to the simulated cumulative N_2O emissions suggest that the DayCent model cannot fully represent the variability. Although DayCent's simulation of N_2O emissions is superior to using emission factors (dos Reis Martins et al., 2022), simulating N_2O emissions remains challenging due to the complexity of the processes involved and the high temporal and spatial variability.

Despite this uncertainty, which has not yet been resolved, our analysis of yield-scaled and area-based GWP showed the importance of the unit of reference. Our modeling results show that ISFM with maize monocropping cannot be seen as a negative emission technology. All treatments had a positive absolute GWP, which is consistent with the reported losses of SOC in long-term maize cultivation in SSA (Sommer et al., 2018; Laub et al., 2022). It is also consistent with the postulate that closing SSA yield gaps will boost absolute N_2O emissions (Leitner et al., 2020). However, the strong differences in the yield-scaled GWP between treatments, such as a 72, 32, 63 and 14 % lower yield-scaled GWP in the FYM 1.2+N treatment compared to the control-N treatment, show that the yield-scaled GWP is highly relevant in practical terms. Low-emission and high-yield ISFM treatments, such as FYM 1.2+N, producing between 2 and 4 t of yield per season at emissions between 0.2 and 1 kg CO_2 equivalent per kg of yield, are a suitable mitigation practice compared to standard practice, control with little or no inputs of organic or chemical fertilizer. The difference in the yield-scaled GWP ranges of +N treatments that existed between western Kenyan sites (Aludeka and Sidada; 0.1 to 0.5 kg CO_2 equivalent per kg of yield) compared to central Kenyan sites (Embu and Machanga; 0.8 to 1.1 kg CO_2 equivalent per kg of yield), furthermore, show that N fertilizer should only be applied to responsive soils (Sileshi et al., 2022) to avoid high N_2O emissions at minimal yield. Consequently, sustainable intensification and mitigation of greenhouse gases can go hand in hand.

4.5 DayCent suitable to upscale simulations of "real" ISFM, but not sensitive for very high N inputs

Because mean yields were represented well by the calibrated version of DayCent, it can be used to upscale to national levels and predict the potential of ISFM to lower yield gaps. However, the levelling off of mean yields at very high N loads (Fig. A6) indicates that it may not be suitable to estimate the maximum achievable yields (e.g., Ittersum et al., 2016) and it should be restricted to predictions for medium input levels of N. Given that the historical rates of N fertilizer application in Kenya were less than 50 kg of N ha^{-1} (World-Bank, 2021a), the model seems suitable to simulate the effect of applying real ISFM, which aims at maximum N use efficiency (Vanlauwe et al., 2010), with N rates considerably below the maximum N rates of the simulated field trials (e.g., 80 kg N per season; Mutuku et al., 2020). The prediction of mean yields worked well for *Calliandra* and farmyard manure treatments at 1.2 and 4 t C ha^{-1} in the -N treatment and was also acceptable for CT+N, i.e. all treatments that supply N at the desired rate for ISFM. Hence, at these N-levels, simulated mean yields are likely representative of the ISFM yield potential. The poorer performance of sawdust is less relevant for upscaling, because this is a treatment for scientific understanding (testing if SOC really forms more efficient from lignin rich materials; Palm et al., 2001b). In summary, the mean yield potential and the change in the SOC stocks were well represented, so the calibration seems suitable for assessing the long-term effect of relevant ISFM treatments on soil fertility, yield, and greenhouse gas emissions as well as their trade-offs. Since year-to-year variations of yield were not captured very well, it is questionable how well the current calibration could represent scenarios of climate change, where temperature and precipitation patterns will become more erratic. In the absence of major pests (which in the trials were controlled), the yearly variation in precipitation and temperature should be responsible for the differences, and if these are not well represented, we cannot trust that we can apply DayCent outside of the climatic range that it was calibrated for.

5 Conclusions

In this study, we demonstrated the effectiveness of jointly calibrating the SOM and plant modules of DayCent to simulate maize productivity under integrated soil fertility management (ISFM) in Kenya, using a Bayesian calibration. The calibration successfully constrained the most sensitive model parameters, which were identified by a global sensitivity analysis. Although the default DayCent maize plant parameterization represented the tropical conditions in Kenya well (i.e., the highest probability posterior parameter values were close to the default parameterization), the highest probability posterior SOM turnover rates were two and eight times faster than the default for the passive and slow SOM pool, respectively. This indicates that there is the potential for large losses of SOC in highly weathered tropical soils. However, SOM turnover was subject to high uncertainty, showing that the current module structure suboptimally captures SOM dynamics in highly weathered tropical soils. Nevertheless, our leave-one-site out cross validation showed that the calibration-derived parameter set is robust for upscaling of the model to larger areas in Kenya. At the same time, mean maize grain yields were much better represented than the year-to-year variability of yields, posing the question whether the effects of climate change can be adequately captured by DayCent. Finally, while no ISFM treatment was predicted to act as an absolute sink of greenhouse gases, yield-scaled emissions were lowest in treatments with high and intermediate yields.

Code availability. An earlier version of DayCent is freely available under https://www.nrel.colostate.edu/projects/century/century-downloads.php. For the latest version, we suggest to contact DayCent developers directly, who in our case kindly provided the latest DayCent version.

Data availability. The data sets used for the calibration of this study are available under the IITA data repository. For SOC: https://doi.org/10.25502/wdh5-6c13/d. For yields and biomass: https://doi.org/10.25502/be9y-xh75/d.

Author contributions. JS, MN and ML designed the modeling exercise. ML summarized the data, conducted the modeling exercise and prepared the original draft. MWMM, DM, RY, SMN and WW managed and maintained the long-term trials. ML, SMN, MN, WW, MvdB, MC and JS were involved in the various sampling campaigns. MC, MN, BV and JS acquired funding for the long-term trials. All coauthors contributed in writing and editing of the final submitted article.

Competing interests. All authors declare that they have no conflict of interest.

Acknowledgements. This study was supported by funds from the European Union‘s Horizon2020 framework (LANDMARC; Grant agreement ID 869367) the Swiss National Science Foundation (SNSF; grant number 172940) and by the DSCATT project “Agricultural Intensification and Dynamics of Soil Carbon Sequestration in Tropical and Temperate Farming Systems” (N° AF 1802-001, N° FT C002181), supported by the Agropolis Foundation (“Programme d’Investissement d’Avenir” Labex Agro, ANR-10-LABX-0001-01) and by the TOTAL Foundation within a patronage agreement. We further acknowledge funding and technical support from the Tropical Soil Biology and Fertility Institute of CIAT (TSBF-CIAT), the International Institute of Tropical Agriculture (ITTA), and ETH Zurich in maintaining the experiments throughout many years. The AI language model "Writefull for Overleaf" has been used to improve the grammar of the manuscript.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Appendix A: Appendix

A1 Site and organic resource characteristics

Table A1. Locations, soil properties and climatic conditions of the study sites. Soil properties are given for the 0 - 15 cm depth layer. Coordinates are given in the WGS 84 reference system.

Soil characteristics	Embu	Machanga	Sidada	Aludeka
Latitude	-0.517	-0.793	0.143	0.574
Longitude	37.459	37.664	34.422	34.191
Initial soil C (%)	2.9	0.3	1.54	0.83
Initial N (%)	0.3	0.02	0.12	0.08
Initail bulk density (g cm^{-3})	1.26	1.51	1.3	1.45
pH (H_2O)	5.43	5.27	5.4	5.49
Clay (%)	59.8	13.2	55.7	13.4
Soil type (FAO, 1998)	Humic Nitisol	Ferric Alisol	Humic Ferralsol	Acrisol
Altitude (m)*	1380	1022	1420	1180
Annual rainfall (mm)*	1175	795	1730	1660
Mean annual temperature (°C)	20.1	23.7	22.6	24.4
Months of long rainy season	3 - 8	3 - 8	3 - 9	3 - 9
Months of short rainy season	10 - 01	10 - 01	10 - 01	10 - 01

*Means calculated based on measured data from 2005 to 2020

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Table A2. Mean measured chemical characteristics (and 95% confidence intervals) of organic resources applied at all sites. Measurements were available from Embu and Machanga from 2002 to 2004, all sites from 2005 to 2007 and in 2018. Significant differences in residue properties were found between the different organic resources, but not between sites and years. Same letters within the same row indicate the absence of significant differences for that property ($p < 0.05$). Abbreviations: n.c. = not classified * according to Palm et al. (2001a).

Measured property	*Tithonia*	*Calliandra*	Maize stover	Sawdust	Farmyard manure
C (g kg^{-1})	345^b (333-357)	396^c (383-409)	397^c (386-408)	433^d (416-449)	234^a (213-255)
N (g kg^{-1})	33.2^d (28.9-38.2)	32.5^d (28.3-37.3)	7.2^b (6.5-8)	2.5^a (2.1-2.8)	18.1^c (15-21.8)
C/N ratio	12.4^a (10.8-14.1)	13.6^a (11.9-15.5)	58.7^b (52.8-65.2)	199.1^c (174.1-227.7)	12.3^a (9.9-15.4)
P (g kg^{-1})	2.3^d (1.8-2.9)	1.1^c (0.8-1.5)	0.4^b (0.3-0.6)	0.1^a (0-0.2)	3.1^d (2.3-3.9)
K (g kg^{-1})	37.2^c (21.2-65.2)	8.7^b (5-15.3)	9^b (6-13.5)	2.8^a (1.6-4.9)	19.4^{bc} (7.8-48.6)
Lignin (g kg^{-1})	90^{ab} (62-117)	105^b (77-133)	48^a (37-60)	172^c (144-199)	198^c (154-242)
Polyphenols (g kg^{-1})	19^c (14.9-24.3)	108.7^d (85.3-138.6)	11.3^b (9.5-13.6)	4.9^a (3.8-6.2)	7.8^{ab} (5.2-11.5)
Lignin/N ratio	2.6^a (1.8-3.7)	3.1^{ab} (2.2-4.3)	6.2^c (4.8-8)	58.3^d (41.1-82.8)	6.9^{bc} (3.9-12.3)
Quality / turnover rate*	High / fast	High / slow	Low / fast	Low / slow	n.c.
Class*	1	2	3	4	n.c.
kg N in 4.0 t C ha^{-1} yr^{-1}, -N [+N]	323 [563]	295 [535]	68 [308]	20 [260]	324 [564]
kg N in 1.2 t C ha^{-1} yr^{-1}, -N [+N]	97 [337]	88 [328]	20 [260]	6 [246]	97 [337]

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Table A3. DayCent model parameters (and feasible ranges) of parameters which were not included in the Bayesian model calibration due to a Sobol total sensitivity index < 1%.

Parameter	Description	Range width	Units	Initial value	Lower value	Upper value	Model file
frtc(2)	C allocated to roots at time frtc(3) without stress	small	fraction of NPP	0.20	0.15	0.25	crop.100
frtc(4)	Max. increase in C going to roots under stress	small	fraction of NPP	0.10	0.08	0.12	crop.100
frtc(5)	Max. increase in C going to roots under stress (maturity)	small	fraction of NPP	0.10	0.08	0.12	crop.100
biomax	AGB at which min.and max. C/E ratios of plant increases	small	g biomass m^{-2}	700.00	525.00	875.00	crop.100
pramx(1,2)	Max. aboveground C/N ratio with biomass > biomax	small	C/N ratio	125.00	93.75	156.25	crop.100
prbmn(1,1)	For computing min. C/N ratio for belowground matter	small	C/N ratio	45.00	33.75	56.25	crop.100
efrgrn(1)	Fraction of above ground N which goes to grain.	small	fraction	0.75	0.56	0.94	crop.100
flig(1,1)	Intercept for annual rainfall effect on lignin content	small	fraction of lignin	0.12	0.09	0.15	crop.100
ppdf(3)	Right curve shape for temperature effect on growth curve	very small	unitless	1.00	0.90	1.10	crop.100
ppdf(4)	Right curve shape for temperature effect on growth curve	very small	unitless	2.50	2.25	2.75	crop.100
favail(1)	Fraction of N available per day to plants	moderate	fraction of N	0.15	0.07	0.22	crop.100
(aneref(1)-anaref(2))	Rain/ET ratio below which, no effect of anaerobiosis	small	unitless	1.00	0.50	1.50	fix.100
aneref(2)	Rain/ET ratio with max. anaerobiosis effect	moderate	unitless	3.00	1.50	4.50	fix.100
damr(1,1)&(2,1)	Fraction of surface N and soil N absorbed by residue	large	fraction of N	0.02	0.01	0.06	fix.100
damrmn(1)	Min. C/N ratio allowed in residue after direct absorption	moderate	C/N	15.00	7.50	22.50	fix.100
dec1(2)	Max. structural litter turnover	small	g g^{-1} yr^{-1}	4.90	3.68	6.12	fix.100
dec2(2)	Max. metabolic litter turnover	small	g g^{-1} yr^{-1}	18.50	13.88	23.12	fix.100
dec3(2)	Max. active pool turnover	small	g g^{-1} yr^{-1}	7.30	5.47	9.12	fix.100
(decX(2)/decX(1))	Ratio soil to surface turnover (newly defined parameter)	small	unitless	1.25	0.94	1.56	fix.100
fwloss(1)	Scaling factor; interception & evaporation by biomass	moderate	unitless	1.00	0.50	1.50	fix.100
fwloss(2)	Scaling factor; bare soil precipitation evaporation	moderate	unitless	1.00	0.50	1.50	fix.100
fwloss(3)	Scaling factor; transpiration water loss	moderate	unitless	1.00	0.50	1.50	fix.100
pabres	Residue amount which results in max. direct N absorption	moderate	g C m^{-2}	100.00	50.00	150.00	fix.100
teff(2)	Y location of temperature inflection point (decomposition)	large	unitless	11.75	3.92	35.25	fix.100
teff(3)	Step size of temperature effect on decomposition	moderate	unitless	29.70	14.85	44.55	fix.100
teff(4)	Inflection point slope of temperature effect (decomposition)	very large	unitless	0.25	0.02	2.46	fix.100
varat11&12(1,1)	Max. C/N ratio for material entering active pool	small	C/N	20.00	15.00	25.00	fix.100
varat11&12(2,1)	Min. C/N ratio for material entering active pool	small	C/N	3.00	2.25	3.75	fix.100
varat21&22(1,1)	Max. C/N ratio for material entering slow pool	small	C/N	20.00	15.00	25.00	fix.100
varat3(1,1)	Max. C/N ratio for material entering passive pool	small	C/N	13.00	9.75	16.25	fix.100
varat3(2,1)	Min. C/N ratio for material entering passive pool	small	C/N	6.00	4.50	7.50	fix.100
drain	Fraction of excess water lost by drainage		fraction of H_2O	0.80	0.20	0.90	site.100
dmp_st	Damping factor for calculating soil temperature by layer	large	unitless	0.00	0.00	0.01	sitepar.in
N2Oadjust_(max-min)	Proportion of nitrified N that is lost as N_2O (difference)	large	fraction of N	0.00	0.00	0.01	sitepar.in
Ncoeff	Min water/temperature limitation coefficient (nitrification)	large	unitless	0.03	0.01	0.09	sitepar.in
dmpflux	The damping factor for soil water flux	large	unitless	0.00	0.00	0.00	sitepar.in
astlig_TD	lignin fraction content of organic matter	small	g g^{-1} biomass	0.09	0.07	0.11	omad.100
astrec(1)_TD	C/N ratio of added organic matter	very small	C/N ratio	12.40	11.16	13.64	omad.100
astlig_CC	lignin fraction content of organic matter	small	g g^{-1} biomass	0.10	0.08	0.13	omad.100
astrec(1)_CC	C/N ratio of added organic matter	very small	C/N ratio	13.60	12.24	14.96	omad.100
astlig_MS	lignin fraction content of organic matter	small	g g^{-1} biomass	0.05	0.04	0.06	omad.100
astrec(1)_MS	C/N ratio of added organic matter	very small	C/N ratio	58.70	52.83	64.57	omad.100
astlig_SD	lignin fraction content of organic matter	small	g g^{-1} biomass	0.17	0.13	0.21	omad.100
astrec(1)_SD	C/N ratio of added organic matter	very small	C/N ratio	199.10	179.19	219.01	omad.100
astlig_FYM	lignin fraction content of organic matter	small	g g^{-1} biomass	0.20	0.15	0.25	omad.100
astrec(1)_FYM	C/N ratio of added organic matter	small	C/N ratio	12.30	11.07	13.53	omad.100

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

A1 Posterior, when allowing even larger ranges

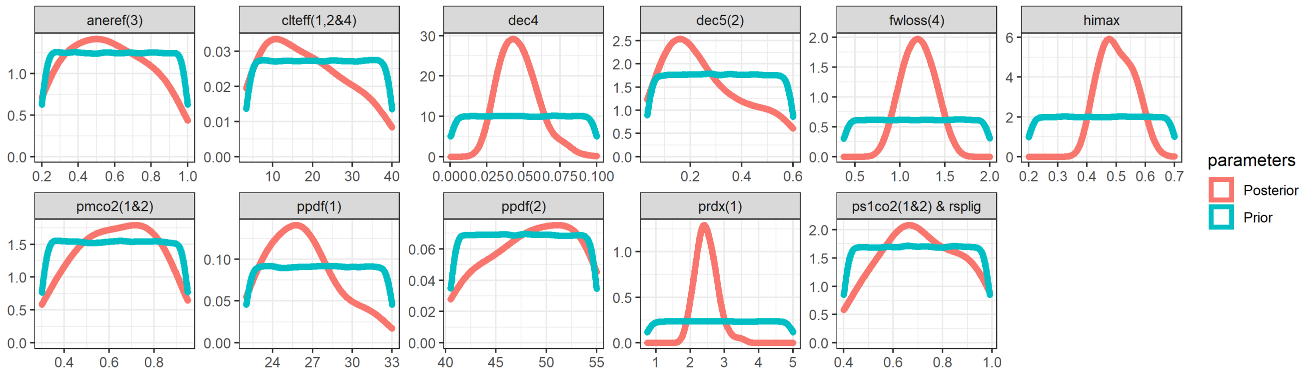

Figure A1. Prior compared to posterior parameter distribution resulting from increasing the ranges of the uncertainty based Bayesian calibration. Dashed vertical lines represent the default parameter sets. The posterior distributions are based on all four sites combined.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

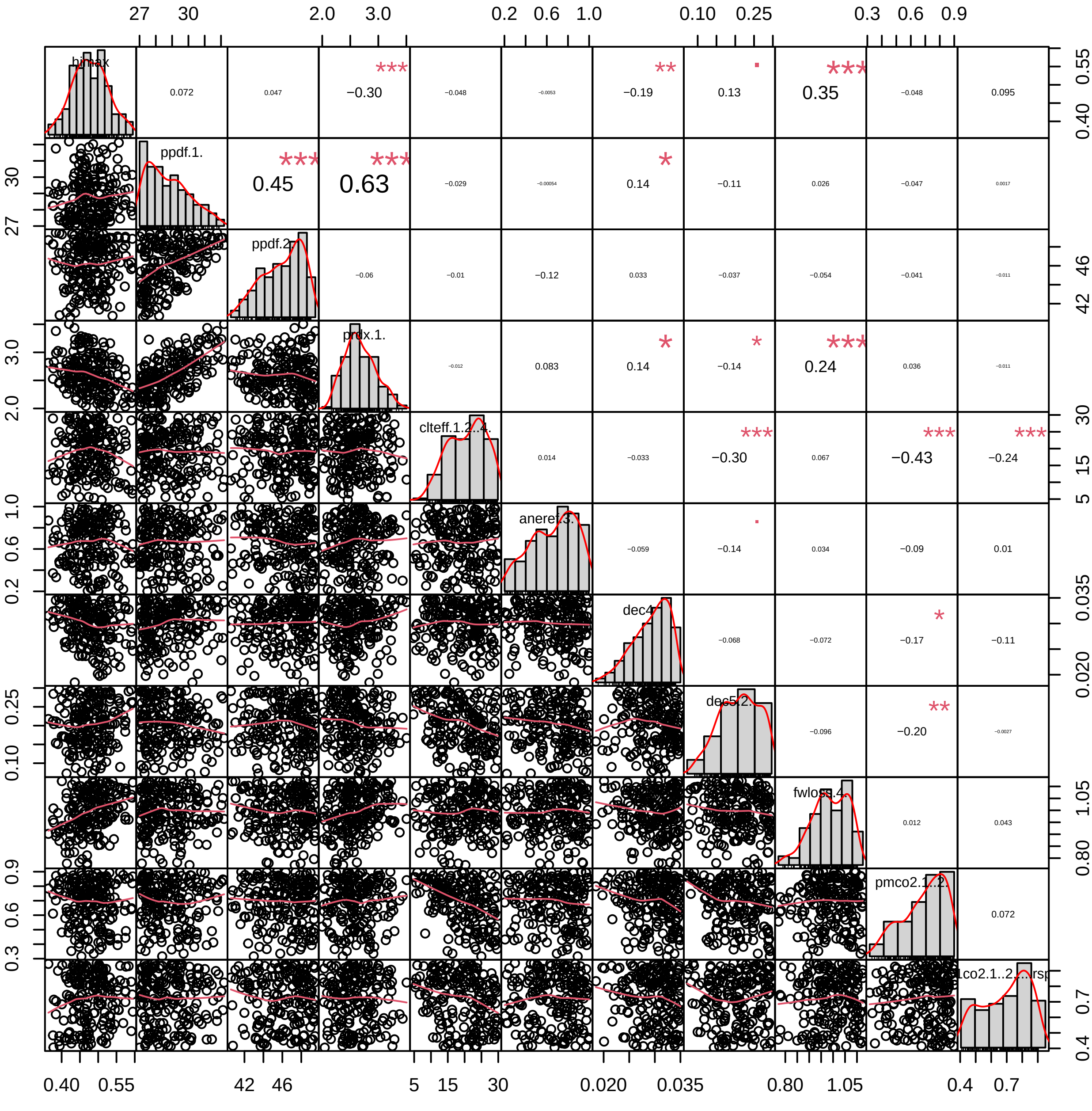

Figure A2. Correlation of parameters from the posterior parameter sets. The posterior distributions are based on all four sites combined.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

A2 Barplots of SOC and comparing measured and simulated mean yield

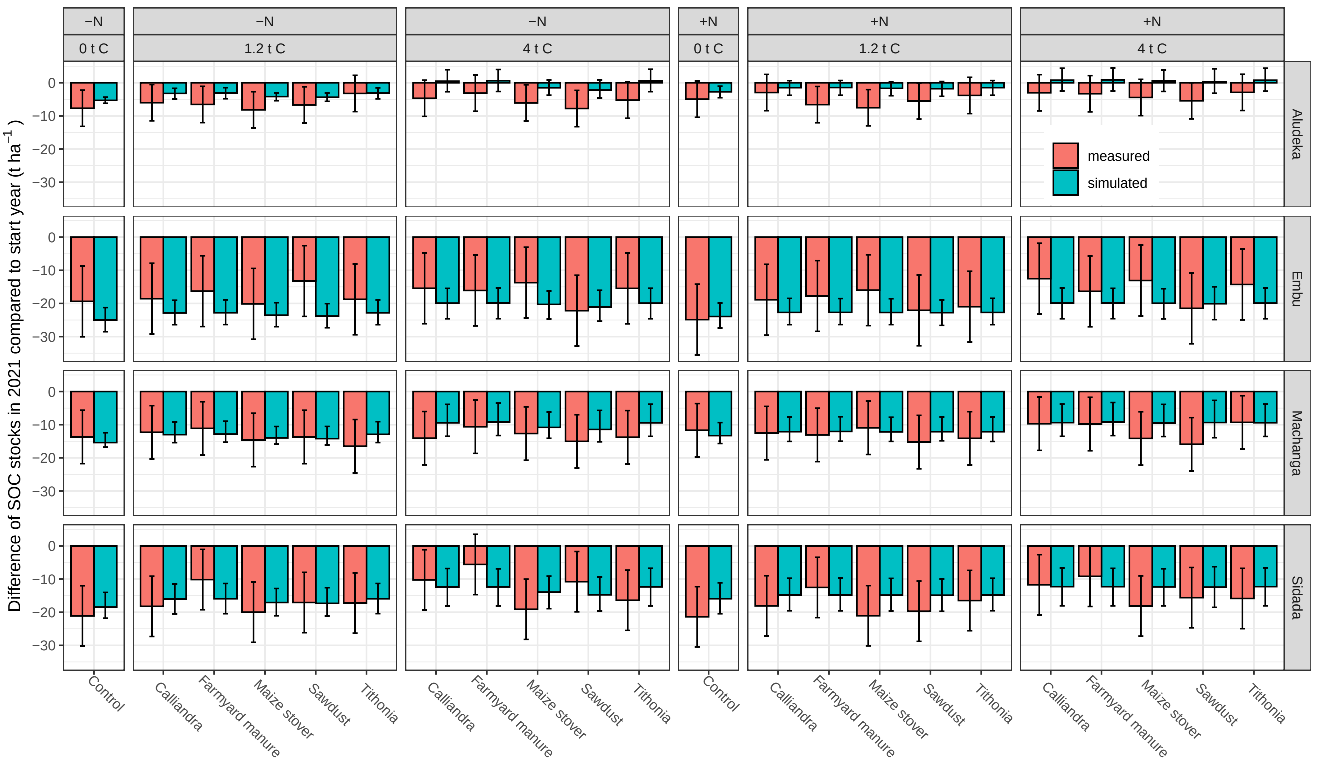

Figure A3. Barplots of simulated and measured change of SOC stocks until 2021 from cross-validation, at the four study sites for the different organic resource and chemical nitrogen fertilizer treatments. Error bars represent 95% confidence intervals based on BC (simulations) and variance (measurements).

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

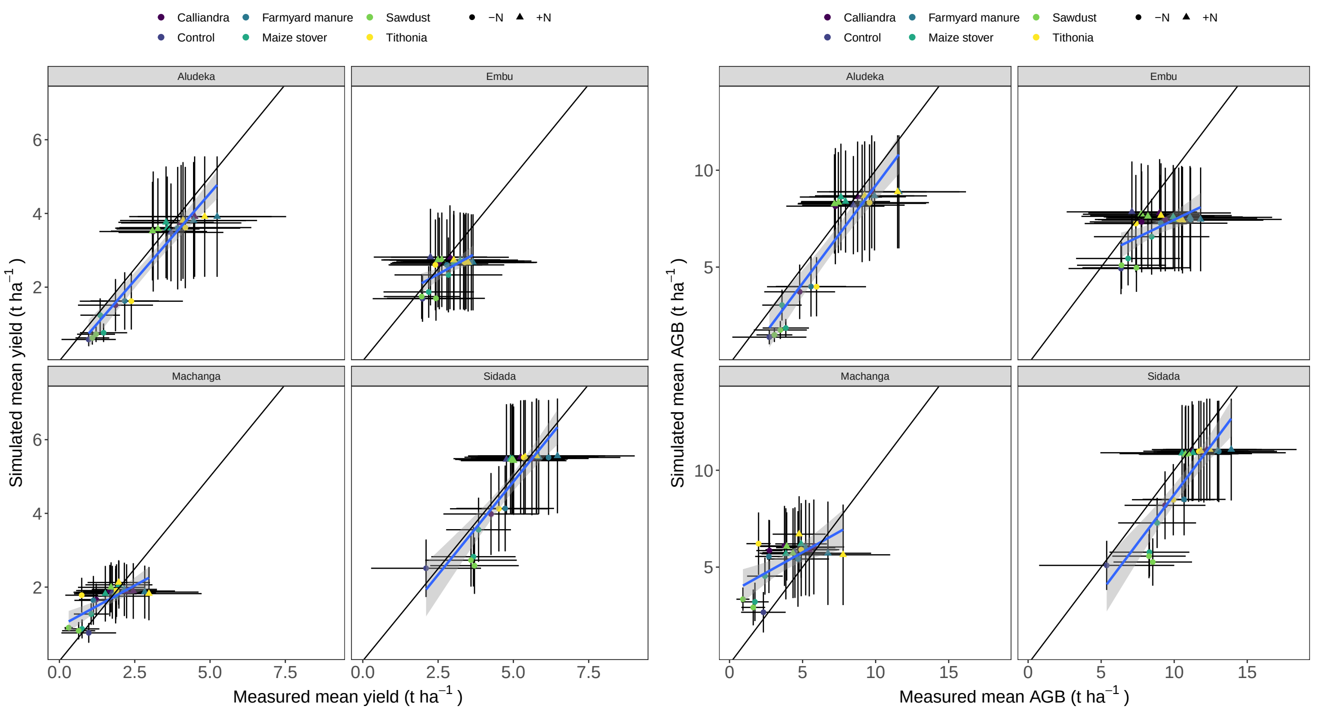

Figure A4. Mean simulated versus mean measured yield and aboveground biomass (AGB) from the leave-one-site-out cross-validation. Error bars represent the standard deviation of measured and simulated values over all years.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

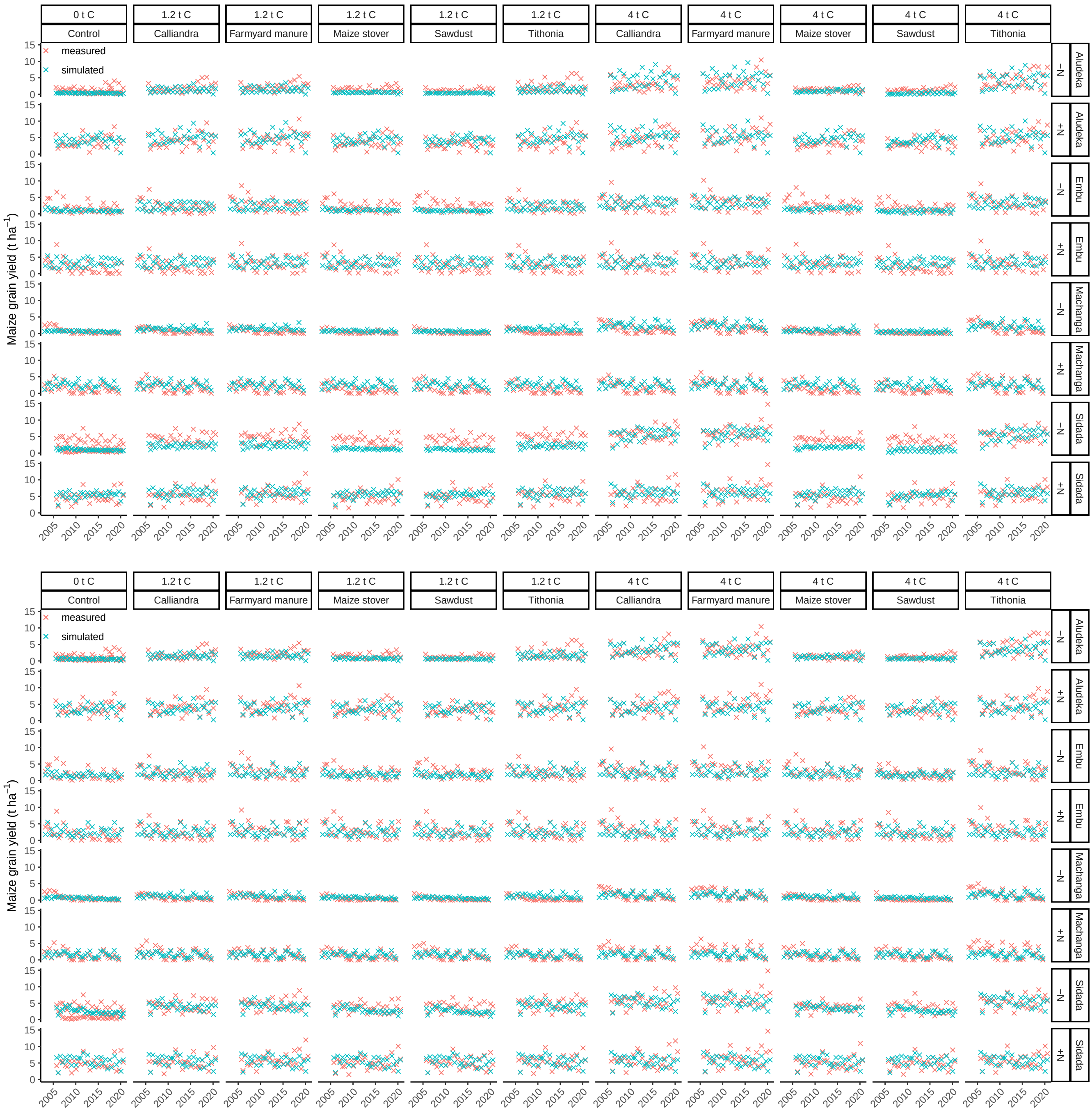

Figure A5. Measured and simulated maize grain yields over time, at the four study sites for the different organic resource and chemical nitrogen fertilizer treatments for the uncalibrated (top) and calibrated DayCent model (bottom).

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

635 **A3 Site specific sensitivities of yield to N fertilizer**

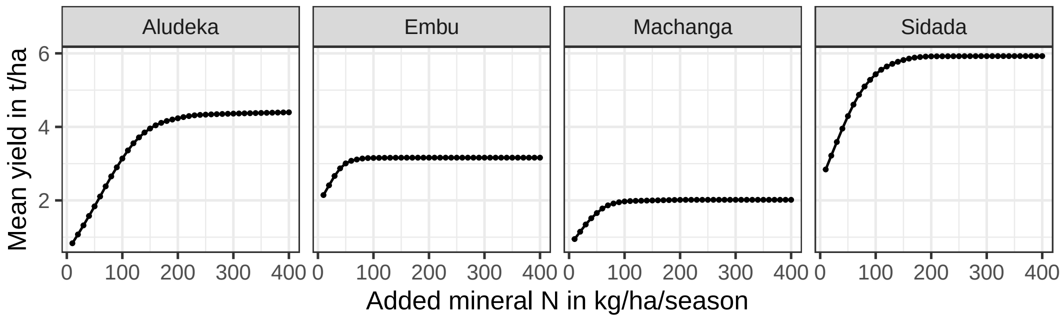

Figure A6. N response curve by site using the calibrated DayCent parameters. This was done to explain why DayCent insensitive at high N levels. N was given in 50/50 split application, at the two real dates.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

A4 SOC stocks

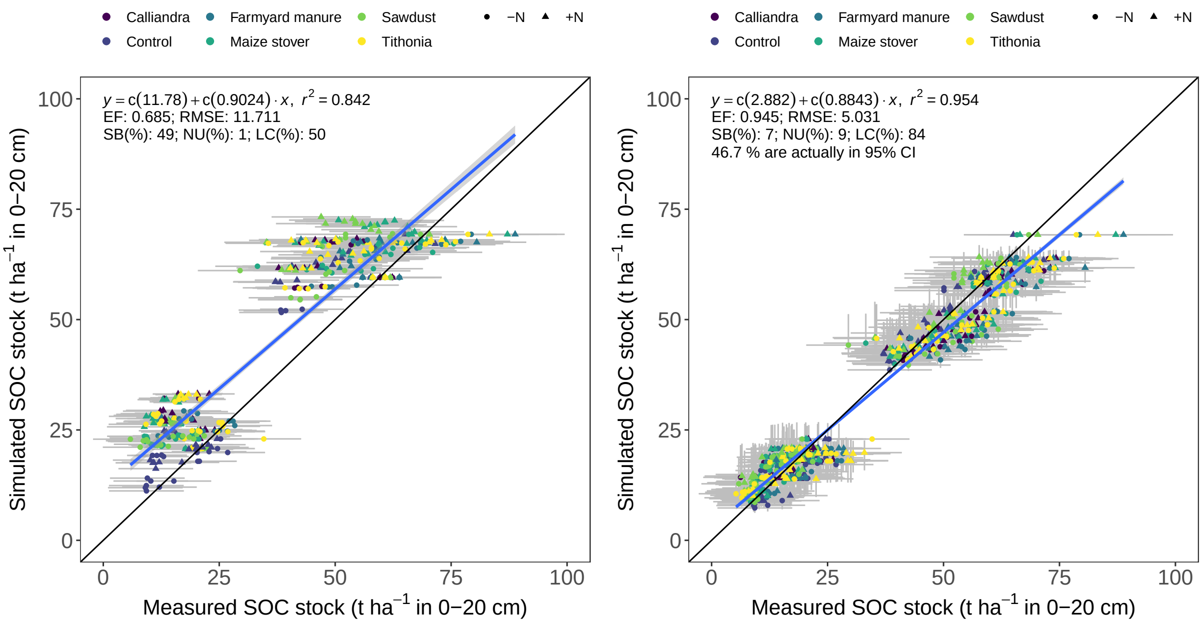

Figure A7. SOC: 1st DayCent outputs initial (left) vs calibrated by leave-one-site-out cross-validation (right). Grey bands denote the 95% confidence intervals of measured (horizontal) and posterior simulated (vertical) values.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

A5 Calibration improvement by Site

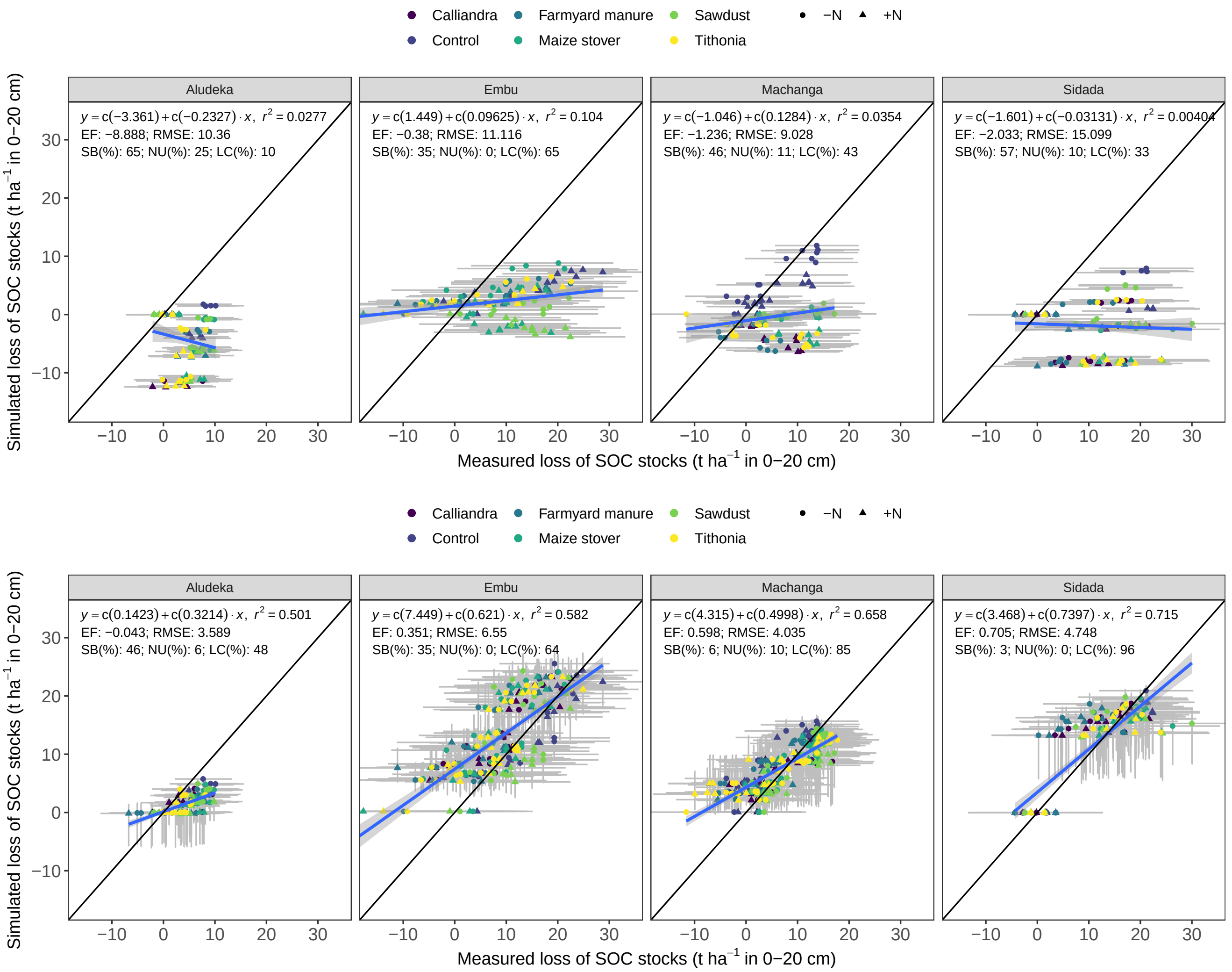

Figure A8. Simulated compared to measured changes in SOC stocks since experiment start for the default DayCent parameter set (top) vs the calibrated parameter set by leave-one-site-out cross-validation (bottom). Grey bands denote the 95% confidence intervals of measured (horizontal) values and the 95% credibility intervals of posterior distribution (vertical).

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

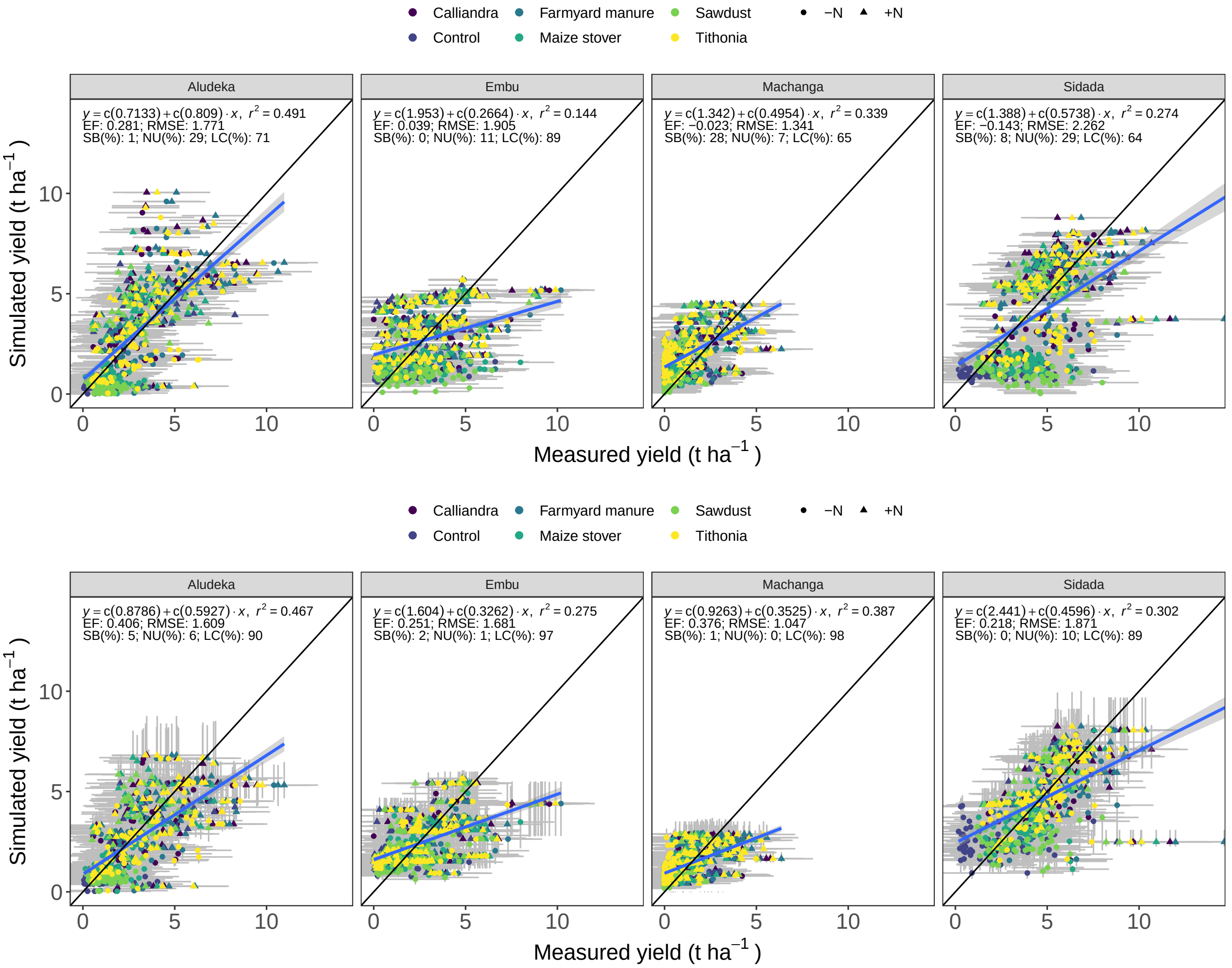

Figure A9. Simulated compared to measured yield for the default DayCent parameter set (top) vs the calibrated parameter set by leave-one-site-out cross-validation (bottom). Grey bands denote the 95% confidence intervals of measured (horizontal) values and the 95% credibility intervals of posterior distribution (vertical).

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

A6 N_2O emissions

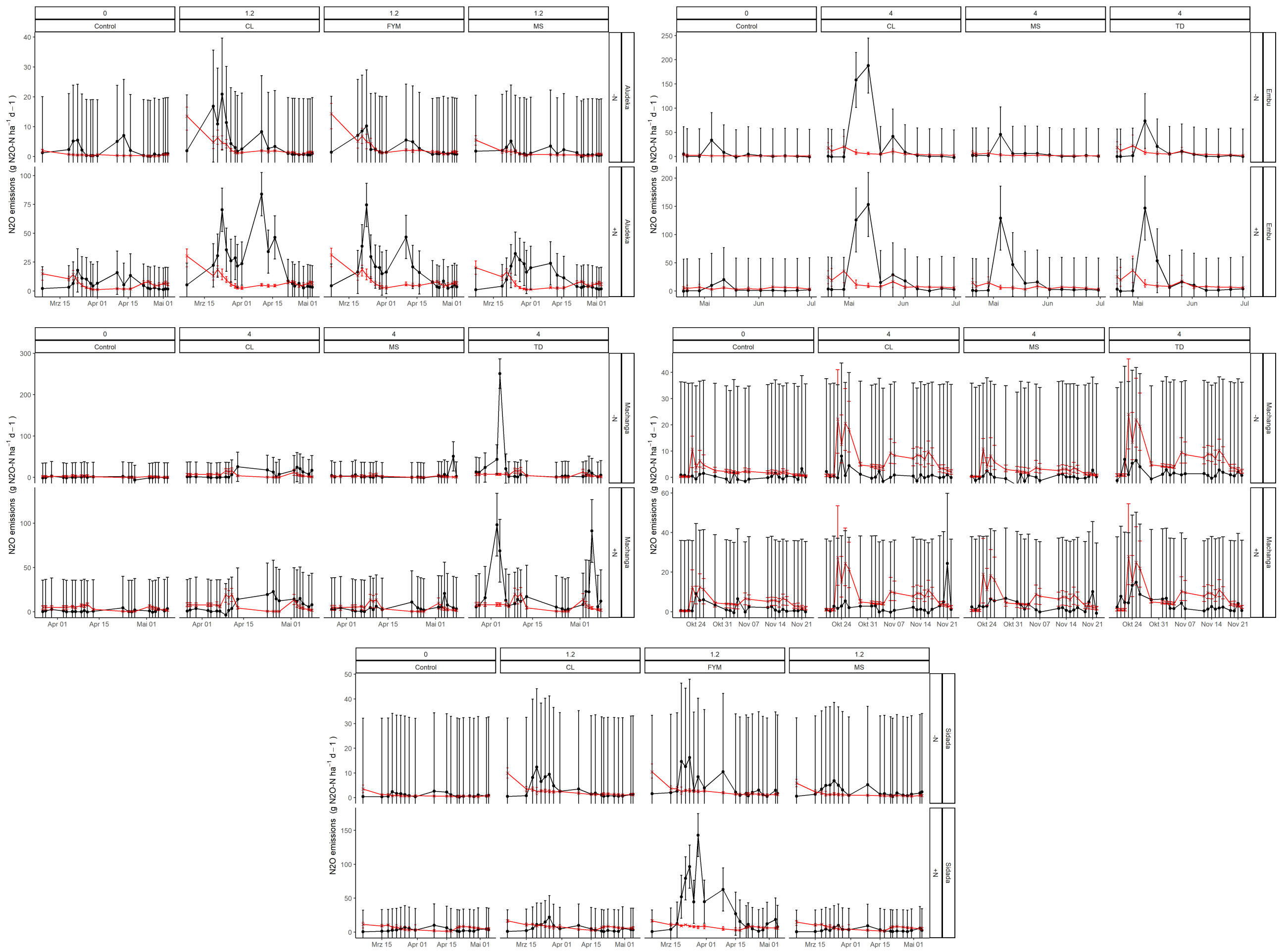

Figure A10. Temporal development of measured (black) vs simulated (red) N_2O emissions by site. The black error bars represent the 95% confidence intervals due to spatial replication error, the red error bars represent the 95% credibility intervals of simulated N_2O emissions resulting from parameter distribution of the posterior parameter set.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

References

Abramoff, R. Z., Guenet, B., Zhang, H., Georgiou, K., Xu, X., Viscarra Rossel, R. A., Yuan, W., and Ciais, P.: Improved global-scale predictions of soil carbon stocks with Millennial Version 2, Soil Biology and Biochemistry, p. 108466, https://doi.org/10.1016/j.soilbio.2021.108466, 2022.

Ahrens, B., Reichstein, M., Borken, W., Muhr, J., Trumbore, S. E., and Wutzler, T.: Bayesian calibration of a soil organic carbon model using Δ14C measurements of soil organic carbon and heterotrophic respiration as joint constraints, Biogeosciences, 11, 2147–2168, https://doi.org/10.5194/bg-11-2147-2014, 2014.

Cardinael, R., Guibert, H., Kouassi Brédoumy, S. T., Gigou, J., N'Goran, K. E., and Corbeels, M.: Sustaining maize yields and soil carbon following land clearing in the forest–savannah transition zone of West Africa: Results from a 20-year experiment, Field Crops Research, 275, 108 335, https://doi.org/10.1016/j.fcr.2021.108335, 2022.

Chivenge, P., Vanlauwe, B., Gentile, R., Wangechi, H., Mugendi, D., Kessel, C. v., and Six, J.: Organic and Mineral Input Management to Enhance Crop Productivity in Central Kenya, Agronomy Journal, 101, 1266–1275, https://doi.org/10.2134/agronj2008.0188x, 2009.

Chivenge, P., Vanlauwe, B., and Six, J.: Does the combined application of organic and mineral nutrient sources influence maize productivity? A meta-analysis, Plant and Soil, 342, 1–30, https://doi.org/10.1007/s11104-010-0626-5, 2011.

Clifford, D., Pagendam, D., Baldock, J., Cressie, N., Farquharson, R., Farrell, M., Macdonald, L., and Murray, L.: Rethinking soil carbon modelling: a stochastic approach to quantify uncertainties, Environmetrics, 25, 265–278, https://doi.org/10.1002/env.2271, 2014.

Corbeels, M., Cardinael, R., Naudin, K., Guibert, H., and Torquebiau, E.: The 4 per 1000 goal and soil carbon storage under agroforestry and conservation agriculture systems in sub-Saharan Africa, Soil and Tillage Research, 188, 16–26, https://doi.org/10.1016/j.still.2018.02.015, 2019.

Dangal, S. R. S., Schwalm, C., Cavigelli, M. A., Gollany, H. T., Jin, V. L., and Sanderman, J.: Improving Soil Carbon Estimates by Linking Conceptual Pools Against Measurable Carbon Fractions in the DAYCENT Model Version 4.5, Journal of Advances in Modeling Earth Systems, 14, e2021MS002 622, https://doi.org/10.1029/2021MS002622, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2021MS002622, 2022.

Del Grosso, S., Parton, W., Mosier, A., Hartman, M., Brenner, J., Ojima, D., Schimel, D., and others: Simulated interaction of carbon dynamics and nitrogen trace gas fluxes using the DAYCENT model, in: Modeling carbon and nitrogen dynamics for soil management, pp. 303–332, CRC Press, Boca Raton, 1st edition edn., https://www.taylorfrancis.com/books/edit/10.1201/9780367801373/modeling-carbon-nitrogen-dynamics-soil-management-liwang-ma-shaffer-soren-hansen, 2001.

Del Grosso, S., Parton, W., Mosier, A., Holland, E., Pendall, E., Schimel, D., and Ojima, D.: Modeling soil CO2 emissions from ecosystems, Biogeochemistry, 73, 71–91, https://doi.org/10.1007/s10533-004-0898-z, 2005.

Doetterl, S., Stevens, A., Six, J., Merckx, R., Van Oost, K., Casanova Pinto, M., Casanova-Katny, A., Muñoz, C., Boudin, M., Zagal Venegas, E., and Boeckx, P.: Soil carbon storage controlled by interactions between geochemistry and climate, Nature Geoscience, 8, 780–783, https://doi.org/10.1038/ngeo2516, number: 10 Publisher: Nature Publishing Group, 2015.

dos Reis Martins, M., Necpalova, M., Ammann, C., Buchmann, N., Calanca, P., Flechard, C. R., Hartman, M. D., Krauss, M., Le Roy, P., Mäder, P., Maier, R., Morvan, T., Nicolardot, B., Skinner, C., Six, J., and Keel, S. G.: Modeling N2O emissions of complex cropland management in Western Europe using DayCent: Performance and scope for improvement, European Journal of Agronomy, 141, 126 613, https://doi.org/10.1016/j.eja.2022.126613, 2022.

FAO: World reference base for soil resources, no. 84 in World Soil Resources Report, FAO, Rome, 1998.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

FAO: FAOSTAT Online Database, https://www.fao.org/faostat/en/#data/QCL, 2023.

Gaillard, R. K., Jones, C. D., Ingraham, P., Collier, S., Izaurralde, R. C., Jokela, W., Osterholz, W., Salas, W., Vadas, P., and Ruark, M. D.: Underestimation of N2O emissions in a comparison of the DayCent, DNDC, and EPIC models, Ecological Applications, 28, 694–708, https://doi.org/10.1002/eap.1674, _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1002/eap.1674, 2018.

Gauch, H. G., Hwang, J. T. G., and Fick, G. W.: Model Evaluation by Comparison of Model-Based Predictions and Measured Values, Agronomy Journal, 95, 1442–1442, https://doi.org/10.2134/agronj2003.1442, 2003.

Gentile, R., Vanlauwe, B., and Six, J.: Litter quality impacts short- but not long-term soil carbon dynamics in soil aggregate fractions, Ecological Society of America, 21, 695–703, https://doi.org/10.1890/09-2325.1, 2011.

Gurung, R. B., Ogle, S. M., Breidt, F. J., Williams, S. A., and Parton, W. J.: Bayesian calibration of the DayCent ecosystem model to simulate soil organic carbon dynamics and reduce model uncertainty, Geoderma, 376, 114 529, https://doi.org/10.1016/j.geoderma.2020.114529, 2020.

Gurung, R. B., Ogle, S. M., Breidt, F. J., Parton, W. J., Del Grosso, S. J., Zhang, Y., Hartman, M. D., Williams, S. A., and Venterea, R. T.: Modeling nitrous oxide mitigation potential of enhanced efficiency nitrogen fertilizers from agricultural systems, Science of The Total Environment, 801, 149 342, https://doi.org/10.1016/j.scitotenv.2021.149342, 2021.

Hodnett, M. G. and Tomasella, J.: Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer functions developed for tropical soils, Geoderma, 108, 155–180, https://doi.org/10.1016/S0016-7061(02)00105-2, 2002.

Hutchinson, G. L. and Mosier, A. R.: Improved Soil Cover Method for Field Measurement of Nitrous Oxide Fluxes, Soil Science Society of America Journal, 45, 311–316, https://doi.org/10.2136/sssaj1981.03615995004500020017x, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.2136/sssaj1981.03615995004500020017x, 1981.

Iooss, B., Veiga, S. D., Janon, A., Pujol, G., Broto, w. c. f. B., Boumhaout, K., Delage, T., Amri, R. E., Fruth, J., Gilquin, L., Guillaume, J., Idrissi, M. I., Gratiet, L. L., Lemaitre, P., Marrel, A., Meynaoui, A., Nelson, B. L., Monari, F., Oomen, R., Rakovec, O., Ramos, B., Roustant, O., Song, E., Staum, J., Sueur, R., Touati, T., and Weber, F.: sensitivity: Global Sensitivity Analysis of Model Outputs, https://CRAN.R-project.org/package=sensitivity, 2021.

Ittersum, M. K. v., Bussel, L. G. J. v., Wolf, J., Grassini, P., Wart, J. v., Guilpart, N., Claessens, L., Groot, H. d., Wiebe, K., Mason-D'Croz, D., Yang, H., Boogaard, H., Oort, P. A. J. v., Loon, M. P. v., Saito, K., Adimo, O., Adjei-Nsiah, S., Agali, A., Bala, A., Chikowo, R., Kaizzi, K., Kouressy, M., Makoi, J. H. J. R., Ouattara, K., Tesfaye, K., and Cassman, K. G.: Can sub-Saharan Africa feed itself?, Proceedings of the National Academy of Sciences, 113, 14 964–14 969, https://doi.org/10.1073/pnas.1610359113, publisher: National Academy of Sciences Section: Biological Sciences, 2016.

Jobbágy, E. G. and Jackson, R. B.: The vertical distribution of soil organic carbon and its relation to climate and vegetation, Ecological Applications, 10, 423–436, https://doi.org/10.1890/1051-0761(2000)010[0423:TVDOSO]2.0.CO;2, 2000.

Kamoni, P. T., Gicheru, P. T., Wokabi, S. M., Easter, M., Milne, E., Coleman, K., Falloon, P., and Paustian, K.: Predicted soil organic carbon stocks and changes in Kenya between 1990 and 2030, Agriculture, Ecosystems & Environment, 122, 105–113, https://doi.org/10.1016/j.agee.2007.01.024, 2007.

Laub, M., Demyan, M. S., Nkwain, Y. F., Blagodatsky, S., Kätterer, T., Piepho, H.-p., and Cadisch, G.: DRIFTS band areas as measured pool size proxy to reduce parameter uncertainty in soil organic matter models, Biogeosciences, 17, 1393–1413, https://doi.org/10.5194/bg-17-1393-2020, 2020.

Laub, M., Ali, R. S., Demyan, M. S., Nkwain, Y. F., Poll, C., Högy, P., Poyda, A., Ingwersen, J., Blagodatsky, S., Kandeler, E., and Cadisch, G.: Modeling temperature sensitivity of soil organic matter decomposition: Splitting the pools, Soil Biology and Biochemistry, 153, 108 108–108 108, https://doi.org/10.1016/j.soilbio.2020.108108, 2021.

Laub, M., Corbeels, M., Couëdel, A., Mathu Ndungu, S., Mucheru-Muna, M. W., Mugendi, D., Necpalova, M., Waswa, W., van de Broek, M., Vanlauwe, B., and Six, J.: High quality organic resources are most efficient in stabilizing soil organic carbon: Evidence from four long-term experiments in Kenya, EGUsphere, pp. 1–33, https://doi.org/10.5194/egusphere-2022-1416, publisher: Copernicus GmbH, 2022.

Laub, M., Corbeels, M., Mathu Ndungu, S., Mucheru-Muna, M. W., Mugendi, D., Necpalova, M., Van de Broek, M., Waswa, W., Vanlauwe, B., and Six, J.: Combining manure with mineral N fertilizer maintains maize yields: Evidence from four long-term experiments in Kenya, Field Crops Research, 291, 108 788, https://doi.org/10.1016/j.fcr.2022.108788, 2023.

Lee, J., Necpálová, M., and Six, J.: Biophysical potential of organic cropping practices as a sustainable alternative in Switzerland, Agricultural Systems, 181, 102 822, https://doi.org/10.1016/j.agsy.2020.102822, 2020.

Leitner, S., Pelster, D. E., Werner, C., Merbold, L., Baggs, E. M., Mapanda, F., and Butterbach-Bahl, K.: Closing maize yield gaps in sub-Saharan Africa will boost soil N2O emissions, Current Opinion in Environmental Sustainability, 47, 95–105, https://doi.org/10.1016/j.cosust.2020.08.018, 2020.

Lemma, B., Williams, S., and Paustian, K.: Long term soil carbon sequestration potential of smallholder croplands in southern Ethiopia with DAYCENT model, Journal of Environmental Management, 294, 112 893, https://doi.org/10.1016/j.jenvman.2021.112893, 2021.

Levavasseur, F., Mary, B., and Houot, S.: C and N dynamics with repeated organic amendments can be simulated with the STICS model, Nutrient Cycling in Agroecosystems, 119, 103–121, https://doi.org/10.1007/s10705-020-10106-5, 2021.

Levy, P. E., Cowan, N., van Oijen, M., Famulari, D., Drewer, J., and Skiba, U.: Estimation of cumulative fluxes of nitrous oxide: uncertainty in temporal upscaling and emission factors, European Journal of Soil Science, 68, 400–411, https://doi.org/10.1111/ejss.12432, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejss.12432, 2017.

Loague, K. and Green, R. E.: Statistical and graphical methods for evaluating solute transport models: Overview and application, Journal of Contaminant Hydrology, 7, 51–73, https://doi.org/10.1016/0169-7722(91)90038-3, 1991.

Lobell, D. B., Bänziger, M., Magorokosho, C., and Vivek, B.: Nonlinear heat effects on African maize as evidenced by historical yield trials, Nature Climate Change, 1, 42–45, https://doi.org/10.1038/nclimate1043, 2011.

Ma, J., Rabin, S. S., Anthoni, P., Bayer, A. D., Nyawira, S. S., Olin, S., Xia, L., and Arneth, A.: Assessing the impacts of agricultural managements on soil carbon stocks, nitrogen loss, and crop production – a modelling study in eastern Africa, Biogeosciences, 19, 2145–2169, https://doi.org/10.5194/bg-19-2145-2022, publisher: Copernicus GmbH, 2022.

Mainka, M., Summerauer, L., Wasner, D., Garland, G., Griepentrog, M., Berhe, A. A., and Doetterl, S.: Soil geochemistry as a driver of soil organic matter composition: insights from a soil chronosequence, Biogeosciences, 19, 1675–1689, https://doi.org/10.5194/bg-19-1675-2022, publisher: Copernicus GmbH, 2022.

Mtangadura, T. J., Mtambanengwe, F., Nezomba, H., Rurinda, J., and Mapfumo, P.: Why organic resources and current fertilizer formulations in Southern Africa cannot sustain maize productivity: Evidence from a long-term experiment in Zimbabwe, PLOS ONE, 12, e0182 840, https://doi.org/10.1371/journal.pone.0182840, publisher: Public Library of Science, 2017.

Mutuku, E. A., Roobroeck, D., Vanlauwe, B., Boeckx, P., and Cornelis, W. M.: Maize production under combined Conservation Agriculture and Integrated Soil Fertility Management in the sub-humid and semi-arid regions of Kenya, Field Crops Research, 254, 107 833, https://doi.org/10.1016/j.fcr.2020.107833, 2020.

Necpalova, M., Lee, J., Skinner, C., Büchi, L., Wittwer, R., Gattinger, A., van der Heijden, M., Mäder, P., Charles, R., Berner, A., Mayer, J., and Six, J.: Potentials to mitigate greenhouse gas emissions from Swiss agriculture, Agriculture, Ecosystems & Environment, 265, 84–102, https://doi.org/10.1016/j.agee.2018.05.013, 2018.

Necpálová, M., Anex, R. P., Fienen, M. N., Del Grosso, S. J., Castellano, M. J., Sawyer, J. E., Iqbal, J., Pantoja, J. L., and Barker, D. W.: Understanding the DayCent model: Calibration, sensitivity, and identifiability through inverse modeling, Environmental Modelling & Software, 66, 110–130, https://doi.org/10.1016/j.envsoft.2014.12.011, 2015.

Nezomba, H., Mtambanengwe, F., Rurinda, J., and Mapfumo, P.: Integrated soil fertility management sequences for reducing climate risk in smallholder crop production systems in southern Africa, Field Crops Research, 224, 102–114, https://doi.org/10.1016/j.fcr.2018.05.003, 2018.

Nyawira, S. S., Hartman, M. D., Nguyen, T. H., Margenot, A. J., Kihara, J., Paul, B. K., Williams, S., Bolo, P., and Sommer, R.: Simulating soil organic carbon in maize-based systems under improved agronomic management in Western Kenya, Soil and Tillage Research, 211, 105 000, https://doi.org/10.1016/j.still.2021.105000, 2021.

Palm, C. A., Gachengo, C. N., Delve, R. J., Cadisch, G., and Giller, K. E.: Organic inputs for soil fertility management in tropical agroecosystems: application of an organic resource database, Agriculture, Ecosystems & Environment, 83, 27–42, https://doi.org/10.1016/S0167-8809(00)00267-X, 2001a.

Palm, C. A., Giller, K. E., Mafongoya, P. L., and Swift, M. J.: Management of organic matter in the tropics: Translating theory into practice, Nutrient Cycling in Agroecosystems, 61, 63–75, https://doi.org/10.1023/A:1013318210809, 2001b.

Parton, W. J., Schimel, D. S., Cole, C. V., and Ojima, D. S.: Analysis of Factors Controlling Soil Organic Matter Levels in Great Plains Grasslands, Soil Science Society of America Journal, 51, 1173–1179, https://doi.org/10.2136/sssaj1987.03615995005100050015x, 1987.

Parton, W. J., Hartman, M., Ojima, D., and Schimel, D.: DAYCENT and its land surface submodel: description and testing, Global and Planetary Change, 19, 35–48, https://doi.org/10.1016/S0921-8181(98)00040-X, 1998.

R Core Team: R: A Language and Environment for Statistical Computing, https://www.r-project.org/, 2020.

Rattalino Edreira, J. I., Andrade, J. F., Cassman, K. G., van Ittersum, M. K., van Loon, M. P., and Grassini, P.: Spatial frameworks for robust estimation of yield gaps, Nature Food, pp. 1–7, https://doi.org/10.1038/s43016-021-00365-y, bandiera_abtest: a Cc_license_type: cc_by Cg_type: Nature Research Journals Primary_atype: Research Publisher: Nature Publishing Group Subject_term: Biogeography;Plant ecology Subject_term_id: biogeography;plant-ecology, 2021.

Reichenbach, M., Fiener, P., Garland, G., Griepentrog, M., Six, J., and Doetterl, S.: The role of geochemistry in organic carbon stabilization against microbial decomposition in tropical rainforest soils, SOIL, 7, 453–475, https://doi.org/10.5194/soil-7-453-2021, publisher: Copernicus GmbH, 2021.

Saito, K., Six, J., Komatsu, S., Snapp, S., Rosenstock, T., Arouna, A., Cole, S., Taulya, G., and Vanlauwe, B.: Agronomic gain: Definition, approach, and application, Field Crops Research, 270, 108 193, https://doi.org/10.1016/j.fcr.2021.108193, 2021.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Computer Physics Communications, 145, 280–297, https://doi.org/10.1016/S0010-4655(02)00280-1, 2002a.

Saltelli, A.: Sensitivity Analysis for Importance Assessment, Risk Analysis, 22, 579–590, https://doi.org/10.1111/0272-4332.00040, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/0272-4332.00040, 2002b.

Saxton, K. E. and Rawls, W. J.: Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions, Soil Science Society of America Journal, 70, 1569–1578, https://doi.org/https://doi.org/10.2136/sssaj2005.0117, 2006.

Sileshi, G. W., Kihara, J., Tamene, L., Vanlauwe, B., Phiri, E., and Jama, B.: Unravelling causes of poor crop response to applied N and P fertilizers on African soils, Experimental Agriculture, pp. 1–17, https://doi.org/10.1017/S0014479721000247, publisher: Cambridge University Press, 2022.

Sommer, R., Mukalama, J., Kihara, J., Koala, S., Winowiecki, L., and Bossio, D.: Nitrogen dynamics and nitrous oxide emissions in a long-term trial on integrated soil fertility management in Western Kenya, Nutrient Cycling in Agroecosystems, 105, 229–248, https://doi.org/10.1007/s10705-015-9693-6, 2016.

Sommer, R., Paul, B. K., Mukalama, J., and Kihara, J.: Reducing losses but failing to sequester carbon in soils – the case of Conservation Agriculture and Integrated Soil Fertility Management in the humid tropical agro-ecosystem of Western Kenya, Agriculture, Ecosystems & Environment, 254, 82–91, https://doi.org/10.1016/j.agee.2017.11.004, 2018.

Stella, T., Mouratiadou, I., Gaiser, T., Berg-Mohnicke, M., Wallor, E., Ewert, F., and Nendel, C.: Estimating the contribution of crop residues to soil organic carbon conservation, Environmental Research Letters, 14, 094 008–094 008, https://doi.org/10.1088/1748-9326/ab395c, 2019.

Tuszynski, J.: caTools: Tools: Moving Window Statistics, GIF, Base64, ROC AUC, etc, https://CRAN.R-project.org/package=caTools, 2021.

van Genuchten, M.: A comparison of numerical solutions of the one-dimensional unsaturated—saturated flow and mass transport equations, Advances in Water Resources, 5, 47–55, https://doi.org/10.1016/0309-1708(82)90028-8, 1982.

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y. A., Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y., and Vereecken, H.: Pedotransfer Functions in Earth System Science: Challenges and Perspectives, Reviews of Geophysics, 55, 1199–1256, https://doi.org/10.1002/2017RG000581, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/2017RG000581, 2017.

van Oijen, M.: Bayesian Compendium, Springer International Publishing, Cham, https://doi.org/10.1007/978-3-030-55897-0, 2020.

Vanlauwe, B., Bationo, A., Chianu, J., Giller, K., Merckx, R., Mokwunye, U., Ohiokpehai, O., Pypers, P., Tabo, R., Shepherd, K., Smaling, E., Woomer, P., and Sanginga, N.: Integrated Soil Fertility Management: Operational Definition and Consequences for Implementation and Dissemination, Outlook on Agriculture, 39, 17–24, https://doi.org/10.5367/000000010791169998, 2010.

Wang, Y., Dou, F., Paustian, K. H., Grosso, S. J. D., Storlien, J. O., Wight, J. P., and Hons, F. M.: Simulating impacts of nitrogen fertilization using DAYCENT to optimize economic returns and environmental services from bioenergy sorghum production, Agronomy Journal, 112, 4861–4878, https://doi.org/10.1002/agj2.20390, _eprint: https://acsess.onlinelibrary.wiley.com/doi/pdf/10.1002/agj2.20390, 2020.

World-Bank: Arable land (% of land area) - Kenya | Data, https://data.worldbank.org/indicator/AG.LND.ARBL.ZS?locations=KE, 2021a.

World-Bank: Prevalence of severe food insecurity in the population (%) - Kenya, World, Sub-Saharan Africa, Malawi | Data, https://data.worldbank.org/indicator/SN.ITK.SVFI.ZS?locations=KE-1W-ZG-MW, 2021b.

Xiao, Q., Huang, Y., Wu, L., Tian, Y., Wang, Q., Wang, B., Xu, M., and Zhang, W.: Long-term manuring increases microbial carbon use efficiency and mitigates priming effect via alleviated soil acidification and resource limitation, Biology and Fertility of Soils, https://doi.org/10.1007/s00374-021-01583-z, 2021.

Yang, Y., Ogle, S., Grosso, S. D., Mueller, N., Spencer, S., and Ray, D.: Regionalizing crop types to enhance global ecosystem modeling of maize production, Environmental Research Letters, 17, 014 013, https://doi.org/10.1088/1748-9326/ac3f06, publisher: IOP Publishing, 2021.

Zhai, R., Tao, F., Lall, U., and Elliott, J.: Africa Would Need to Import More Maize in the Future Even Under 1.5°C Warming Scenario, Earth's Future, 9, e2020EF001 574, https://doi.org/10.1029/2020EF001574, _eprint: https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2020EF001574, 2021.

https://doi.org/10.5194/egusphere-2023-1738
Preprint. Discussion started: 21 August 2023

Zhang, Y. and Yu, Q.: Does agroecosystem model improvement increase simulation accuracy for agricultural N2O emissions?, Agricultural and Forest Meteorology, 297, 108 281, https://doi.org/10.1016/j.agrformet.2020.108281, 2021.