Sensitivity analysis of agricultural inputs for large-scale Soil Organic Matter modelling

Julius Diel[a]*, Uwe Franko[a]

[a] Helmholtz Centre for Environmental Research GmbH – UFZ

*corresponding Author:

Mail: julius.diel@ufz.de

Tel: 0049 345 558 5468

Theodor-Lieser-Straße 4,

06120 Halle

Key words:

uncertainty assessment; ‘4 per mille’; regional data;

Software availablity

R and all its packages are freely available and licenced under the GNU General Public License (www.r-project.org). CCB can freely be downloaded from https://www.ufz.de/index.php?en=39729.

Highlights

- First Sensitivity Analysis of regionalized cultivation data for SOM modelling
- Estimation of possible soil carbon stock changes for the federal state of Saxony
- Estimation of nitrogen mineralisation as source for ground water pollution
- Future data improvement should focus more on carbon stocks

Abstract

The dynamics of Soil Organic Matter (SOM) and its relation to the carbon and nitrogen cycle affect many environmental problems (e.g. climate change, food security and water quality). The development of adaptation strategies requires model predictions, but for the necessary large-scale SOM dynamic studies, the quality of the input data is often limiting the reliability of the results.

So we performed a uncertainty and sensitivity analysis at different sites of the federal state of Saxony, Germany, and assessed the importance of aggregated agricultural data, namely organic amendments, crop yields, area share of by-product incorporation, area share of conservation tillage and initial soil organic carbon (SOC) concentration (p_oram, p_yield, p_bp, p_cons and p_soc respectively) on the result uncertainty by assuming an uniform error of ±10%. The agricultural data was regionalized from 717 long-term observation fields throughout the study region.

We assessed the uncertainties of relative SOC stock change (ΔC_{rel}) and total nitrogen mineralisation from the organic matter (OM-N_{min}) and explored the changing sensitivities over the model period (1998-2014) .

Our results show that p_soc was the most important source of uncertainty for all sites of this study. For ΔC_{rel}, it is over the whole time constantly the by far most sensitive input parameter, with p_bp being the only factor of agricultural practice with some substantial influence on almost all sites. In the mountainous regions, p_cons ranks equal to p_bp, while for the sandy heathlands, none of them mark a substantial influence besides p_soc.

For OM-N_{min}, p_soc loses its importance over time, being outranked by p_oram in the heathlands after 8 years and in the mountainous regions after 13 years. p_oram furthermore places second for all others but one other region, where p_cons is slightly more important. We therefore see the initial carbon content, the share of by-product removal, and the amount of organic amendments as those factors, where improved data quality would bring the highest effect to reduce the uncertainty in regional SOM modelling.

1 Introduction

For regional to global soil organic matter (SOM) dynamic studies, the quality of input data is often a main issue for the plausibility of the results (e.g. Grosz et al., 2017; Hoffmann et al., 2016; Luo et al., 2016). While data sets of different detail are available for soil and climate, the necessary agricultural data must either be modelled itself (with usually only one or two reference crops (Hoffmann et al., 2016; Luo et al., 2013)) or be aggregated from existing survey data.

The inherent uncertainties of input data naturally lead to an uncertainty of the model outputs as well. A sensitivity analysis can help to rank uncertainties of different data classes, which we will refer to as factors from now on, to identify where a reduction of uncertainty would lead to the biggest reduction of the respective output uncertainty, i. e. where further effort for data quality would be put to best use (Saltelli et al., 2007).

The relation between SOM accumulation and agricultural management has gained a lot of interest through the ongoing discussion about the '4 per mille: Soils for Food Security and Climate' initiative (see for example Minasny et al., 2018 and references therein). Considering uncertain model inputs, we investigated the possibility to reach the 4‰ goal, as a relative carbon stock change (ΔC_{rel}). Dynamics of C and N in soil have a strong interaction and must not be assessed separately. Threfore we also investigated the development of overall mineralisation of organic nitrogen (OM-N_{min}) as an N source that reacts over several years on management changes and needs special consideration in long term developments, like when building up SOM stocks.

We used the CCB model (Franko et al., 2011) to predict the SOM dynamics on agricultural land depending on the individual site conditions given by soil texture, air temperature and rainfall as well as land management data concerning cropping (crop type, yield and usage of by-products), organic amendments (amount of slurry and farm yard manure) and soil tillage (conventional or conservational). (Witing et al., 2019) showed that the CCB model does not need spatially or temporally explicit data as can work well with aggregated data (several fields/years with the same shares of cultivation). This dramatically simplifies regional studies of SOM dynamics, because it reduces the heterogeneity of the necessary input data.

In our study, we used aggregated data from a scenario simulation of the federal state of Saxony, Germany, covering the years 1990 to 2014. While climate data and soil physical parameters were available at high spatial resolution, data regarding management and initial SOM concentration of agricultural soils resulted from generalizing procedures that were based on 717 long-term observation fields. The predicted variability of the state-wide scenario results from a combination of real data variability (like climate data, soil types and crop yields) and stochastic variability within the aggregated data (like share of conservation tillage and amount of organic amendments). Beside the Sensitivity Analysis of five selected management parameters (see. section 2.3), we therefore also compared the uncertainty given by our assumed data error with that given by “true” regional heterogeneity.

There exists a vast amount of sensitivity measures, with constantly new developments. Overviews are given in (Helton et al., 2006; Iooss and Lemaître, 2014; Iooss and Saltelli, 2016; Nguyen and Reiter, 2015; Saltelli et al., 2007). Depending on the model complexity and available computational time, different measures may suit the question at hand better or worse. However, all of them relate the change of a model output Y to the change of each model input X_i for which a probability distribution or an uncertainty interval is defined. As our chosen model is based on first-order kinetics (Franko et al., 2011 and Appendix), we assume linear relations between changing input quantities and results. Hence we will compare the suitability of the computationally cheap squared Standardized Regression Coefficient (sSRC) and the total sensitivity T index after Sobol’, the latter being often referred to as a standard procedure, but with high computational cost (Saltelli et al., 2010). We chose T over the First-order Sensitivity Index S (that shall mathematically coincide with the sSRC for linear models) to account for unforeseen nonlinearities in terms of higher-order interactions.

Altogether, in this study we analysed the importance of different sources of uncertainty on the assessment of model results concerning SOM-C accumulation and SOM-N mineralisation.

2 Material and methods

2.1 Model description

CCB (Candy Carbon Balance) is a simplified version of the carbon dynamic model in CANDY (Franko et al., 2011). It describes the turnover of soil organic carbon of the topsoil (0-30 cm) with first-order kinetics in annual time steps for specific site conditions depending on crop yields, input rates of fresh organic matter and the initial soil organic carbon (SOC) content (see Appendix for some formulas). The turnover conditions are characterized by biologic active time (BAT) that is calculated from soil texture, tillage system (Franko and Spiegel, 2016), and annual averages of rainfall and air temperature (Franko and Oelschlägel, 1995). The annual BAT sum represents the hypothetical time in days that is needed under optimal conditions in a laboratory to get the same amount of turnover as in the real world over one year. Outputs of CCB include dynamics of SOC, SOM reproduction and coupled SOM-N mineralization. The model was validated for a range of agricultural management options and site conditions (Franko et al., 2011; Franko and Merbach, 2017; Franko and Spiegel, 2016).

The CCB approach uses conceptual pools that subdivide organic matter in soil into four compartments: (1) several pools of fresh organic matter (FOM), (2) biological active soil organic matter (A-SOM) where mineralization takes part and which is filled up with a matter specific part of FOM turnover (C_{rep}), (3) stabilized soil organic matter (S-SOM) where a matter exchange is assumed with the A-SOM pool and (4) long term stabilized soil organic matter (LTS-SOM) that was considered inert in this study. For each year, the resulting amount of C_{org} is represented by the sum of all SOM pools. The nitrogen flows are coupled through pool-specific C:N-ratios.

Fluxes are calculated with a first order approach depending on pool sizes and BAT. The aforementioned C_{rep} summarizes the effect on SOM by: crop residues like stubble and roots, by-products that were left on the field and org. amendments like slurry and manure and therefore, it can be used as an area specific measure for the impact of agricultural practice on SOM.

2.2 Used data

The data sets for our study were available from a survey aiming at a quantification of land use on the nitrogen load of surface waters. Soil texture was taken from the soil map BK50, provided by the Saxon State Office for the Environment, Agriculture and Geology (*Landesamt für Umwelt, Landwirtschaft und Geologie*, LfULG) on a 500 x 500 m grid, on which we performed our calculations. Climate data (temperature and precipitation) was provided at a 1x1 km grid by ReKIS (www.rekis.org, LfULG) and interpolated to fit our 500x500 m grid. The share of each crop was taken from the Integrated Administration and Control System IACS of the EU and - according to as the primary goal of this scenario data - averaged annually for each catchment area of surface water bodies.

The initial SOC stock was calculated according to soil type, climate conditions and agricultural management, as described by (Witing et al., 2019). We let the model run for seven years (1990-1997) to account for the error of the initial steady-state assumptions and started evaluating the results from 1998 onwards.

Agricultural data comprised yields of main crops, crop specific amount of organic amendments (with a constant slurry /manure ratio of 67/33) considering their abundance, a constant share of by-product incorporation (40%) and the share of catch crops, which were assumed to be always left on the field. This data set from 717 long term observation fields (so called *Dauerbeobachtungsflächen,* provided by the LfULG) from all over Saxony, was averaged for five sub regions. Thus, although the general spatial resolution was high all management data had a more general character.

For our sensitivity analysis we focused on five factors, as described in Chapter 2.3, and their influence on a) the long-term change of SOM-C stocks over time (average change of stock between start year and model year, ΔC_{rel}), to see if and when the 4 per mille goal might be reached and b) on the corresponding yearly mineralisation of organic nitrogen (OM-N_{min}), as a measure for possible nitrate excess. Furthermore, we assessed the sensitivity onto these parameters regarding the length of the model period and the environmental conditions.

2.2.1 Site Description

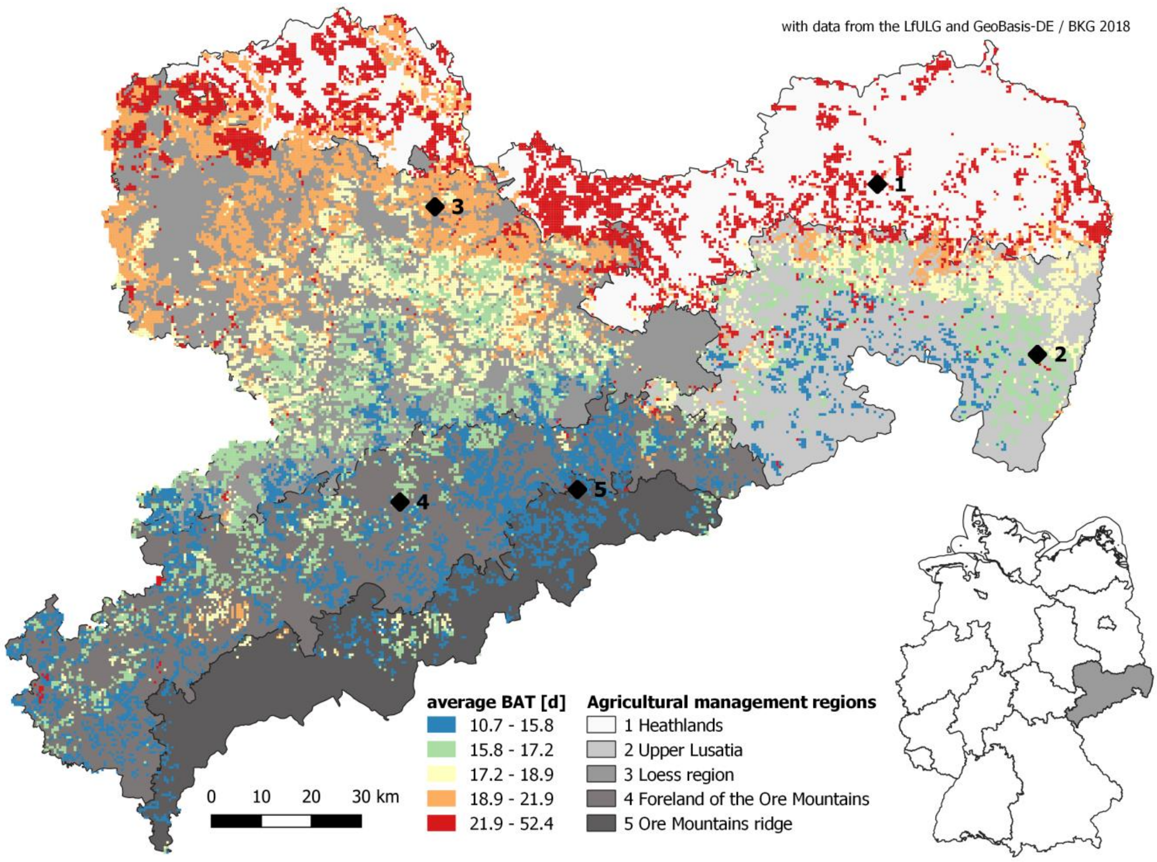

Figure 1: average Biologic Active Time (BAT [d]) for the model period (1998-2014). The classes have equal counts. Black dots show the selected grid cells and the ID of the management region they represent

Saxony is divided into five agricultural management regions (AMR, in german *Agrarstrukturgebiete)* that are considered to be homogeneous. Specific agricultural recommendations are regularly published for them by the local authorities LfULG. The main site characteristics are shown in Table 1. For each agricultural management region we selected one site, i.e. grid cell, with the most representative conditions. We therefore looked for cells that were closest to the mode of first the turnover conditions (BAT), then of the carbon reproduction flux (C_{rep}) and finally of the initial SOC stock. These characteristics are shown in Table 2.

Table 1: Average environmental conditions and standard deviation of the regions for the time period of 1990-2014.

region ID	Temperature [°C/a]	Precipitation [mm/a]	Main soil texture classes	C_{rep} [kg/ha/a]	SR / BP / OA [%]
1	9.6 ± 0.8	715 ± 137	S: 67%, SiL: 16%	841 ± 217	59 / 19 / 21
2	8.9 ± 0.9	825 ± 151	SiL: 84%, L: 8%	944 ± 223	57 / 23 / 20
3	9.5 ± 0.8	750 ± 141	SiL: 88%, L: 5%	966 ± 227	56 / 24 / 20
4	8.3 ± 0.9	903 ± 174	SiL: 77%, L: 18%	1038 ± 252	54 / 19 / 27
5	7.0 ± 0.8	1051 ± 165	SiL: 44%, L: 36%	1046 ± 261	59 / 13 / 28

Soil texture classes from the WRB (IUSS Working Group WRB, 2015). C_{rep}: amount of Carbon that actually enters SOM; SR / BP / OA: share of Stubble and Roots, By-Products and Organic Amendments to C_{rep} (only averages)

For region 1, the Heathlands ("Heidegebiet") we selected a pixel with sandy soil which is the dominant soil type here. It is both the warmest and driest region in Saxony and has good turnover conditions. It is characterized by low yields and low SOM stocks.

Silty loams dominate in region 2, Upper Lusatia ("Oberlausitz"), followed by loams. Together with region 3, the Loess Region ("Lößgebiet"), it is the most important agricultural area with the highest yield potential for demanding crops such as maize and wheat. The soil data discriminates the dominant WRB soil type SiL in region 3 after the German soil classification KA5 (Boden, 2006)) into two texture classes with almost equal share, namely Ut2 (clay/silt/sand: 12.5/70.83/16.67) and Us (clay/silt/sand: 4/77.5/18.5), while in region 2 it is almost exclusively classified as Ut2. These soils differ in their calculated initial SOC concentration; hence we selected a soil with higher clay content (Ut2) in region 2 and a soil with low clay content in region 3.

The Foreland of the Ore Mountains, region 4 ("Erzgebirgsvorland") is characterized by higher precipitation and a lower average temperature as the aforementioned three regions. It is again dominated by silty loams, resulting altogether in a lower BAT compared to regions 2 and 3. It has a higher share of more robust cereals such as triticale and rye as well as pastures, with which the amount of organic amendments and overall C_{rep} increases.

In region 5, the Ore Mountains Ridge ("Erzgebirgskamm"), silty loams and loamy soils are almost equally frequent. Being the coldest and rainiest region, it has a large share of pastures and less demanding cereals. Due to a combination of poor turnover conditions (the lowest BAT) with high amount of C input to soil (highest C_{rep}) we find here the highest SOM concentration of all regions.

Table 2: Average environmental conditions for the selected grid cells.

grid cell	BAT [d]	C_{rep} [kg/ha/a]	SOC_{ini} [M%]	C / U [M%]	WRB
G1	50.80	829	0.95	2.5 / 5	S
G2	17.2	944	1.66	12.5 / 70.8	SiL
G3	19.8	981	1.27	4 / 77.5	SiL
G4	15.8	1046	1.76	12.5 / 70.8	SiL
G5	12.7	1037	2.29	21 / 35.6	L

grid cell: selected grid cell of the respective region, BAT (Biologic Active Time) in [d], C_{rep} (Carbon reproduction flux) in [kg/ha/a], SOC_{ini} (initial SOC value) in [M%], C / U (Clay and Silt content, each in [M%], WRB (WRB soil texture class)

2.3 Sensitivity analysis

We used the values obtained from the data set described in section 2.2 as a baseline and assume an uniform distributed error of 10 per cent around this baseline for each input factor X_i under investigation (with i = 1,…, k. k = 5), as we had no better information on the distribution (Table 3). For the range, we oriented us on (Post et al., 2008) who show probability distributions for yield, org. amendments and initial SOC stock. But as we found no information for a possible error of tillage system or left by-product shares, we decided to simplify the error range overall by equalizing them.

Table 3: assumed error distributions of the input factors X_i.

input factor X_i		assumed error distribution
p_yield	**Yield of the main crops, from which the C-input is derived**	uniform ± 10%
p_oram	**Fresh matter amount of organic amendments from livestock**	uniform ± 10%
p_cons	**Tillage system (represented as share of conservational tillage)**	uniform ± 10%
p_bp	**Spatial share where by-products are left on the field**	uniform ± 10%
p_soc	**Initial SOC stock**	uniform ± 10%

We characterised the sensitivities by two different indices, namely the squared Standardized Regression Coefficient (sSRC) and the Total Sensitivity Index after Sobol` (T) (Saltelli et al., 2007). Both indices build upon random (or in our case quasi-random) samplings from the devised uncertainty ranges of the input factors and relate the variance for each input factor to the variance of the model output.

The sSRC builds upon a linear regression with no interaction terms included, Y being the respective model output (ΔC_{rel} or OM-N_{min}):

$$Y = \sum_{i=1}^{k} b_i \cdot X_i \quad (1)$$

$$sSRC_i = \left(b_i \cdot \frac{\sigma(X_i)}{\sigma(Y)}\right)^2 \quad (2)$$

The parameters b_i were calculated through the lm()-function in R (R Core Team, 2018) and then standardised by multiplying with $\sigma(X_i)/\sigma(Y)$, with σ being the standard deviation. Squaring the value, $sSRC(X_i)$ becomes directly correlated to the Variances (σ^2) of X_i and Y. As the sSRC is based on a linear regression its exploratory power decreases the more the model is non-linear. However, as the sum of the sSRCs equals R^2, it can give estimate for the fitness of this index (the closer the sum gets to 1, the better the fitness).

The Total order Sensitivity Index after Sobol` (T) has a more abstract, generalizing formulation, since it does not assume a linear relationship, but rather averages partial variances for N distinct values per factor. This `model-free` approach of sensitivity measurement comes however at a higher computational cost. For a thorough derivation of the formula we refer the reader to (Saltelli et al., 2007).

The sensitivity analysis was conducted within the R environment, using the packages *sensitivity* and *multisensi*, to assess the temporal behaviour of ΔC_{rel} and OM-N_{min} for each year (Bidot et al., 2018; Iooss et al., 2018; Lamboni et al., 2009; R Core Team, 2018). We used the arithmetic average of the calculated yearly indices as a measure of the total importance.

We divided one Sobol' sequence of 10 dimensions (2k) and N= 1024 observations as implemented in the *randtoolbox*-package (Dutang et al., 2019) into two parts to get the necessary "independent" samples as they are required for the sensitivity analysis after Sobol' (Lo Piano et al., 2019; Saltelli et al., 2010), totalling in 7168 computations (Nt = N*(k+2)) for each of the five pixels. During the computations, the sampled error was held constant. Afterwards, we computed the total-order sensitivity indices T_i after Jansen, for all $N = 2^m$ with m = 4…10, to be sure the indices became numerically stable (Jansen, 1999; Saltelli et al., 2010, 2007). The squared SRC values were only calculated for the first set of samples, again for Nt= N = 2^m computations with m = 4…10. They became stable after 128 runs. For the results shown, N of 1024 was used, as the calculation of T_i sometimes yielded negative numbers for small indices with N = 512.

3 Results

3.1 Uncertainty of the model results

The development of the relative change of the SOC stock ΔC_{rel}, with reference to 1997, is almost steady with only little exceptions Figure 2. For the first years up to around 2003, the stock is declining at all selected sites. From then onwards, there are increases at G1, G3 and G5 resulting in total increases of 2.72‰, 3.4‰ and 2.07‰, respectively. The stocks are more or less stable at G2 and G4. The interquartile ranges decrease slightly over time, ranging from 0.36‰ at G1 to 0.80‰ at G3 on average. They are always very close to the average and smaller than the overall respective changes, indicating little influence of the changed setup for most of the model runs. The maximum result range decreases too and is again smallest for G1 with 1.21‰ and biggest for G3 with 3.30‰ on average. Except for G1, the induced uncertainty is only slightly smaller than the respective regional uncertainty, where we evaluated all grid cells with the same soil in each region, so climatic influences and different crop share developments are present as well.

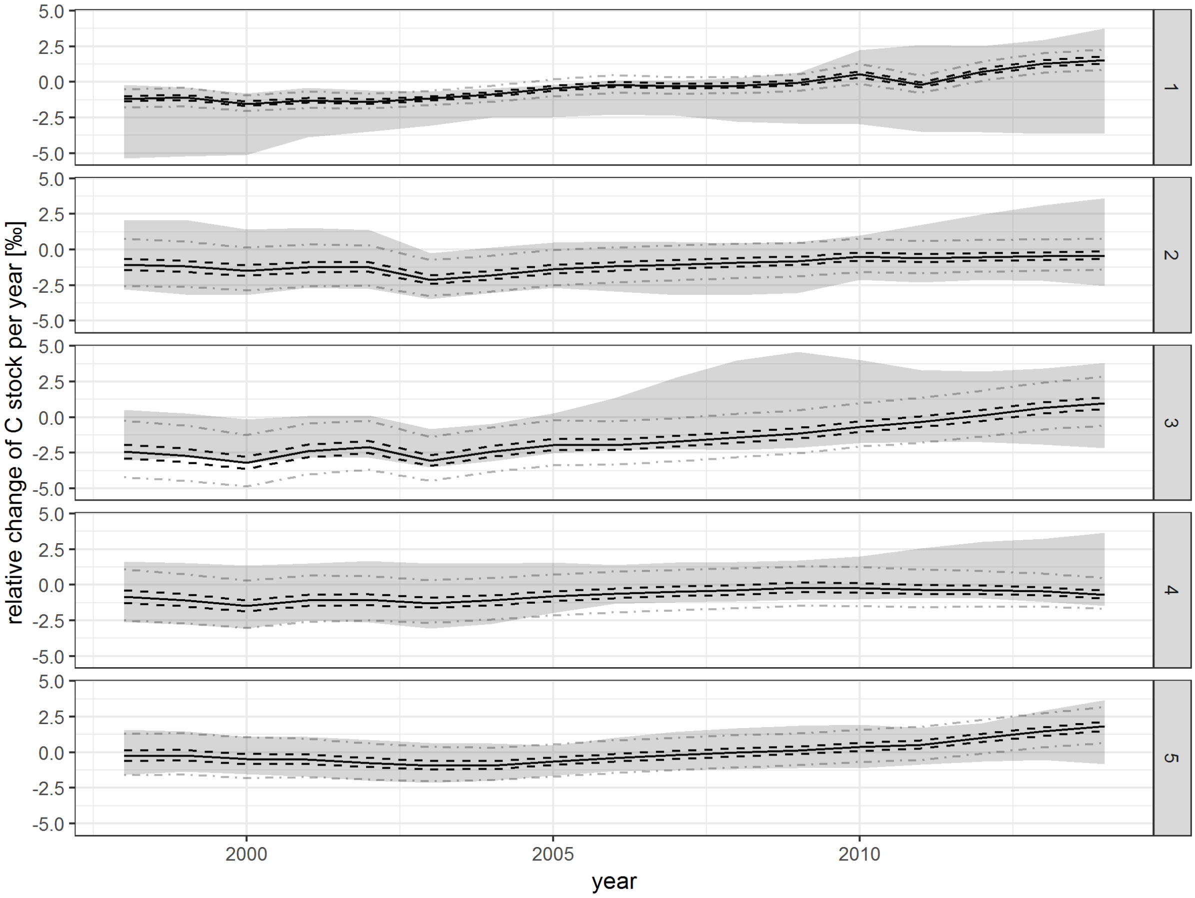

Figure 2: possible result range of relative SOC stock changes ΔC_{rel} to the year 1997 [‰]. Black solid: mean value, black dashed: q25 and q75, grey dashed: minimum and maximum, grey ribbon: possible result range of all grid cells with the same soil in that region

OM-N_{min} shows a different picture, as it is much more influenced by yearly climate conditions (Figure 3). This can be seen for the year 2003, where an extremely dry and warm weather resulted in lower yields and high mineralisation rates for G2 and G3, and from the year 2010 onwards, from where on the share of cultivated crops became more differentiated. This leads to a two to six fold increase of the overall regional variability and a differing mineralisation only on the first of our five grid cells.

For the most part, the uncertainty interval (dotted lines in Figure 3) is quite stable for every grid cell with an average interquartile range of 5.2 kg/ha with a maximum uncertainty of 27.1 kg/ha for G4 in 2014. Overall, the induced uncertainty is smaller than the already existing uncertainty through environmental or agricultural changes.

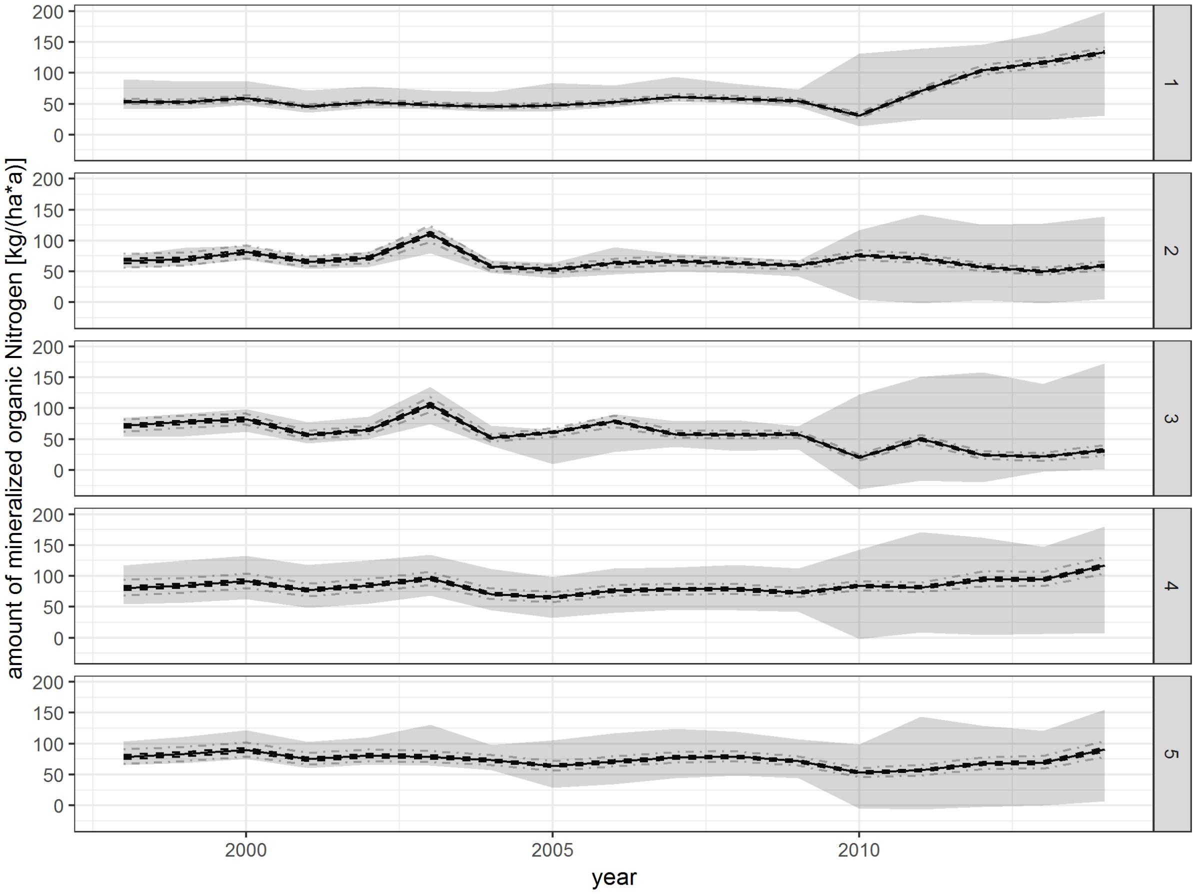

Figure 3: Possible result range of mineralised organic nitrogen ([kg/(ha*a)]). Black solid: mean value, black dashed: q25 and q75, grey dashed: minimum and maximum, grey ribbon: possible result range of all grid cells with the same soil in that region

3.2 Sensitivities

As stated above, OM-N_{min} is a highly dynamic variable. Therefor its sensitivity onto our input factors changes throughout the years. Figure 4 gives an example on how the nitrogen mineralisation responses on the given inputs. It shows in blue the slope of the standardized regression as indicator for sSRC that is the square of slope. The steepest regressions indicate the most sensitive input factors and the sign is not of importance. In the first year of the model run, the initial SOC content is by far the most dominant input parameter (little scattering, steepest regression). With time, its dominance is decreasing and other parameters may influence the result more (e.g. p_cons in 2014). Years with abrupt changes in the input constellations (like 2003) influence the slope of the regressions as well.

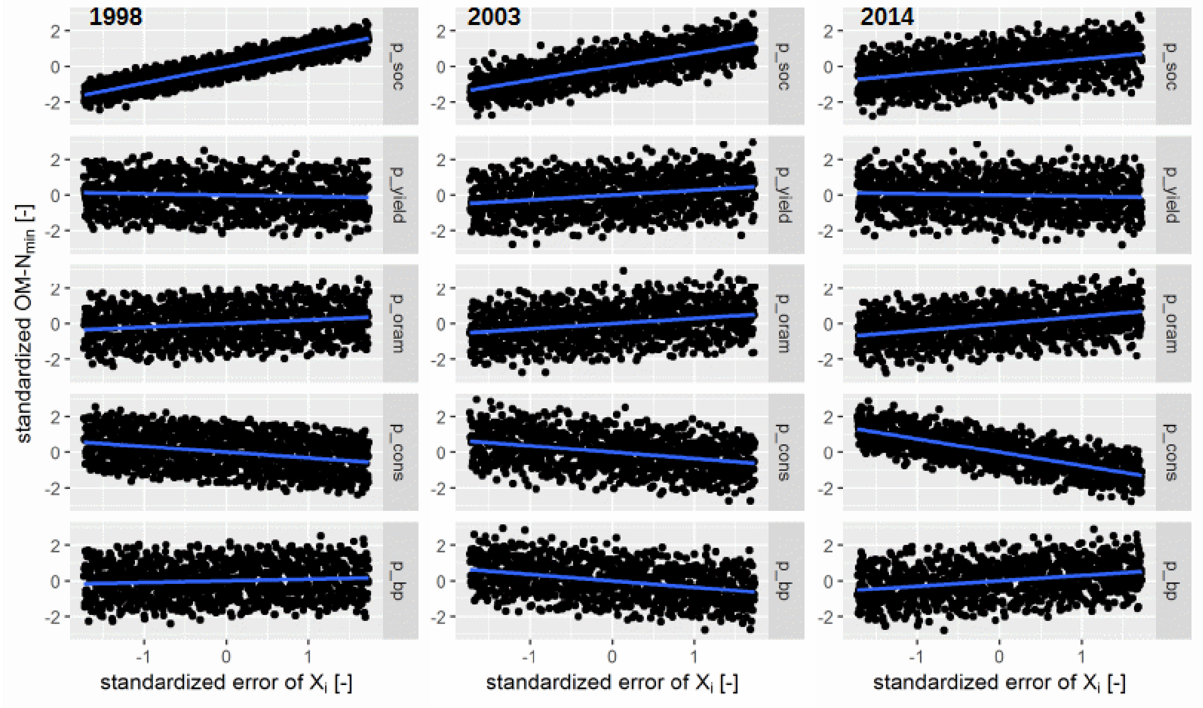

Figure 4: Scatterplots with linear regression of standardized OM-N_{min} (mineralized organic nitrogen, [-]) at G3 against the assumed standardized errors for all X_i [-] and the time steps 1998, 2003, 2014. p_soc: error of initial SOC stock, p_oram: error of amount of organic amendments, p_cons: error of area share of conservartional tillage, p_bp: error of by-product incorporation, p_yield: error of yield

Figure 5 shows the stacked sSRC values for both our target variables to get an overview over the changing importance of our input factors with time. As the values always sum up to (approximately) 1, our model can indeed be considered linear. A further indication is that for T_i, we get almost identical values, as the interaction terms were negligible (see Figure 6 for averaged values over the whole model period).

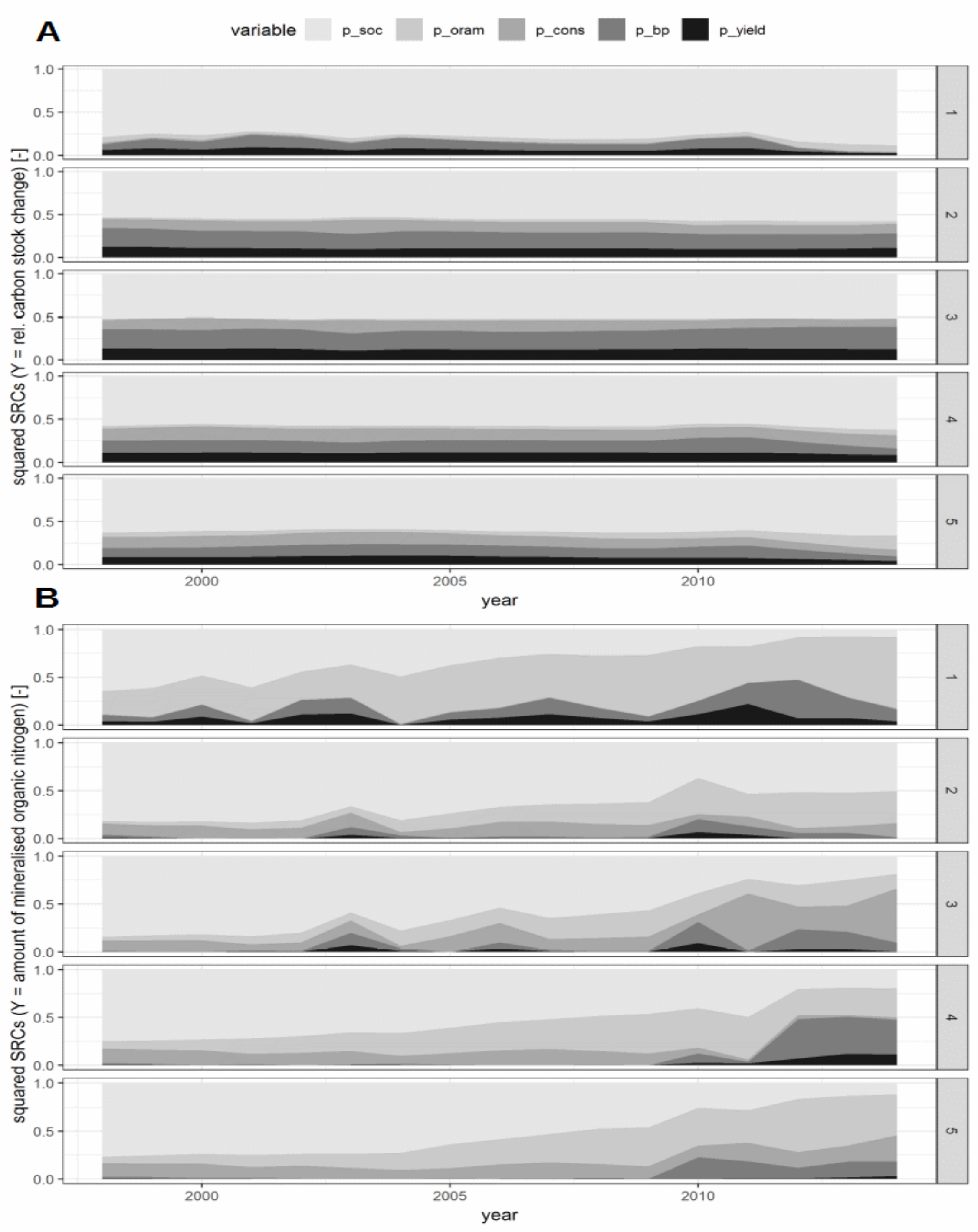

Figure 5: squared SRC values for all time steps and all Xi at each of the five sites.

A) Y = relative carbon stock change. B) Y = amount of mineralised organic Nitrogen. p_soc: error of initial SOC stock, p_oram: error of amount of organic amendments, p_cons: error of area share of conservational tillage, p_bp: error of by-product incorporation, p_yield: error of yield

As ΔC_{rel} develops quite steadily, the sensitivity indices of the input factors are quite steady as well. The largest fluctuations are caused by p_bp. ΔC_{rel} is always dominated at least by 50% through p_soc. For G2 and G3, p_bp has higher influence as well, which is almost always more important than p_yield. The influence of p_oram is always very low, though it gains some importance after 2012 for G1 and G5. p_cons has everywhere more or less the same importance, except for the sandy soil of G1, where it has virtually no influence.

The dynamic of the sensitivity indices for the fluxes of OM-N_{min} is, as the fluxes itself, higher. The sensitivity onto p_soc decreases steadily over time, mostly to be replaced through an increasing sensitivity onto p_oram. Again, the sensitivity against p_yield is always below the sensitivity against p_bp. After 2010, the respective sensitivities change very differently throughout the sites. While for G1 and G4, p_bp becomes more important, and G2 and G5 show the main increase at p_oram, G3 becomes more sensitive to p_cons.

The averaged sensitivity indices over the whole model period show in condensed form the dominant role of the initial carbon stock, especially for ΔC_{rel}. But also for OM-N_{min}, only on the sandy soil with low initial carbon (G1) the p_oram is more important, which again ranks second on all other sites but G3, where it is close behind p_cons. The only agricultural factor that can have some notable influence on ΔC_{rel} is p_bp.

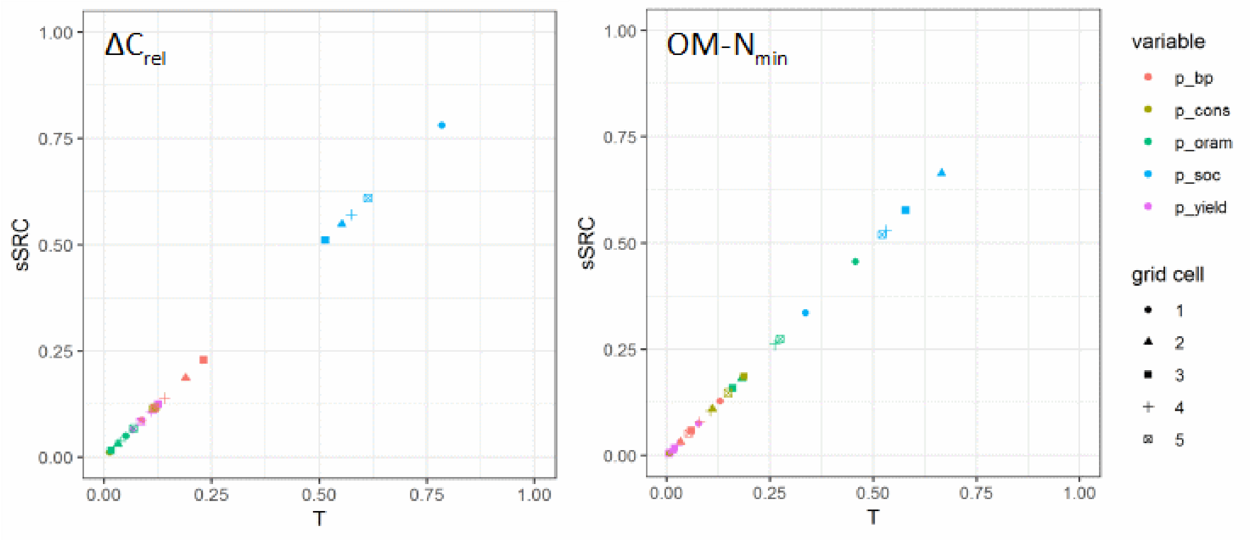

Figure 6: Averaged sensitivity indices over time. ΔC_{rel}: relative Carbon stock change, OM-N_{min}: mineralised organic Nitrogen. sSRC: squared Standardized Regression Coefficient, T: Total-order sensitivity index. p_soc: error of initial SOC stock, p_oram: error of amount of organic amendments, p_cons: error of area share of conservartional tillage, p_bp: error of by-product incorporation, p_yield: error of yield

4 Discussion

The initial SOC content, our most sensitive input factor, was also identified by other studies as one of the most influencing parameters (Post et al., 2008; Qin et al., 2017). Furthermore, it is probably one of the most uncertain factors as well, as it is almost always derived from the assumption to be in steady-state with a certain management that is extrapolated several years into the past. The question whether this steady-state, both in terms of quantity and quality (i.e. the distribution between the conceptual pools), is always reached, is an ongoing dilemma for SOM modelling (Luo et al., 2014; Wutzler and Reichstein, 2007). Here, we only assumed the quantity to be uncertain, but the presented approach could also account for the uncertainty of the quality. This would most certainly increase the uncertainty of the results and its importance even further.

(Qin et al., 2017) further lists the amount of organic amendments and the share of incorporated by-products within the four most influential parameters, which corresponds to our observations, despite the fact that a different model was used. We therefore consider these three as the variables on which future data collection should focus the most.

The spatial variability and differentiability of the model results within the regions is to be questioned, especially regarding the carbon stock changes. Only for the sandy site of the Heathlands, with the lowest initial SOC content and hence lowest “absolute error”, environmental and agricultural influences account for remarkably more variability than our assumed error. To use an uniform error of 10% might be a very general, but if the error distributions are known better, they can become helpful in setting up well-differentiable regional data. If they are not known however, we think that such a simple approach can already help to better assess the results in their reliability.

Already before 2010, the regional variability of OM-N_{min} is higher than for ΔC_{rel}, compared to our assumed uncertainty at the respective grid cells. This shows, that for OM-N_{min}, different drivers that were not accounted for in this study, have a substantial influence as well. These are for one, different cropping spectra with different amounts of organic amendments, for another, varying weather conditions. After 2010, changing survey methods lead to drastically changing data about crop shares and hence drastically changing mineralisation rates, which is certainly a problem of many long-term census data, as this is a kind of variability that is hard to cope with.

With 1997 as reference year, no grid cell would reach the 4‰ goal within the time span, no matter of the sampled input uncertainty. Only with 2000 as reference, G3 would reach it for all sampled combinations. That raises the question on the respective timespan under consideration. But it is also a question, whether the selected soils are already close to their SOM content optimum (White et al., 2018), if even the most benevolent conditions (lowest p_soc and highest p_cons, p_yield, p_bp and p_oram) are sometimes far from 4‰. Also for OM-N_{min}, the factor ranking is very dependent on the respective time frame, as the initial SOC content loses its priority after 8, 13 or more than the 17 years that were modelled, depending on the respective environmental and agricultural conditions.

(Luo et al., 2016; Ogle et al., 2010; Post et al., 2008) showed, that the biggest source of uncertainty lies within the model parametrization itself, so the shown trends for the SOC stock may not be significant at all. However, in this study we were not interested in the overall uncertainty, but on a factor ranking to offer guidance for future data collection, given that CCB is a validated model that should not be too far from reality.

For linear models, the most common type for soil carbon modelling, a simple and computationally cheap sensitivity index is completely sufficient, especially with an economic sampling as provided by the Sobol'-sequence. After already 128 model runs, the $sSRC_i$ became quite stable, whereas the model-free approach of T_i required 7168 runs to not compute some negative indices at certain years without adding any new information. With this, existing models could implement an uncertainty and sensitivity framework with little effort, as this is partly done already in the Yasso07 model (Liski et al., 2009).

5 Conclusions

Our selected approach was able to show how much uncertainty of selected model input factors was transferred to the model results and how their influences may change over time. For short time spans, the initial SOC content is the most influencing factor for both target variables. Regarding the relative SOC stock change, all sensitivities are quite steady and the agricultural practices do not gain influence over the whole model period of 17 years. Regarding the nitrogen mineralisation, it takes more than 10 years for the agricultural practices to have more influence than the initial SOC content, except for sandy soils with the lowest SOC content, where they dominate after already 7 years. For this highly dynamic variable a sensitivity analysis, as it was conducted here, can help to disentangle the various overlapping and contradicting influences.

For our model CCB, the linear $sSRC_i$ and the "model free" T_i both yielded the same result, however at totally different computational costs. So also for other SOM models that usually are linear, the sSRC should be preferred.

6 Declaration of interest / Funding

Declarations of interest: none

The research was partly funded by the LfULG and partly by the Bundesanstalt für Landwirtschaft und Ernährung, grant number: 281B302516

7 Acknowledgements

We would like to thank the LfULG and GALF (Gesellschaft für Angewandte Landschaftsforschung bR) for the provision and pre-processing of data within the framework of the research project "Dynamische Bilanzierung von Humushaushalt und Nährstoffaustrag im regionalen Maβstab im Kontext von Landnutzungs- und Klimawandel – Phase 2: Bilanzierung von Stickstoffeinträgen in sächsische Gewässer"

8 Data availability

The agricultural and soil data used is owned by the Landesamt für Umwelt, Landwirtschaft und Geologie (LfULG), Sachsen, Germany and can be made available with prior consent.

9 References

Bidot, C., Lamboni, M., Monod, H., 2018. multisensi: Multivariate Sensitivity Analysis.

Boden, H.: W.E.R.: H.S.W.G.K.-J.H.R.H.P.J.H.J.D.K.K.-J.S.R.T.H.: A.-A., 2006. Bodenkundliche Kartieranleitung. KA5. Schweizerbart Science Publishers, Stuttgart, Germany.

Dutang, C., Savicky, P., Chalabi, Y., Wuertz, D., 2019. randtoolbox: Generating and Testing Random Numbers, R-Package. https://CRAN.R-project.org/package=randtoolbox.

Franko, U., Kolbe, H., Thiel, E., Ließ, E., 2011. Multi-site validation of a soil organic matter model for arable fields based on generally available input data. Geoderma 166, 119–134. https://doi.org/10.1016/j.geoderma.2011.07.019

Franko, U., Merbach, I., 2017. Modelling soil organic matter dynamics on a bare fallow Chernozem soil in Central Germany. Geoderma 303, 93–98. https://doi.org/10.1016/j.geoderma.2017.05.013

Franko, U., Oelschlägel, B., 1995. Einfluss von klima und textur auf die biologische aktivität beim umsatz der organischen bodensubstanz. Arch. Agron. Soil Sci. 39, 155–163. https://doi.org/10.1080/03650349509365898

Franko, U., Spiegel, H., 2016. Modeling soil organic carbon dynamics in an Austrian long-term tillage field experiment. Soil Tillage Res. 156, 83–90. https://doi.org/10.1016/j.still.2015.10.003

Grosz, B., Dechow, R., Gebbert, S., Hoffmann, H., Zhao, G., Constantin, J., Raynal, H., Wallach, D., Coucheney, E., Lewan, E., Eckersten, H., Specka, X., Kersebaum, K.-C., Nendel, C., Kuhnert, M., Yeluripati, J., Haas, E., Teixeira, E., Bindi, M., Trombi, G., Moriondo, M., Doro, L., Roggero, P.P., Zhao, Z., Wang, E., Tao, F., Rötter, R., Kassie, B., Cammarano, D., Asseng, S., Weihermüller, L., Siebert, S., Gaiser, T., Ewert, F., 2017. The implication of input data aggregation on up-scaling soil organic carbon changes. Environ. Model. Softw. 96, 361–377. https://doi.org/10.1016/j.envsoft.2017.06.046

Helton, J.C., Johnson, J.D., Sallaberry, C.J., Storlie, C.B., 2006. Survey of sampling-based methods for uncertainty and sensitivity analysis. Reliab. Eng. Syst. Saf. 91, 1175–1209. https://doi.org/10.1016/j.ress.2005.11.017

Hoffmann, H., Zhao, G., Asseng, S., Bindi, M., Biernath, C., Constantin, J., Coucheney, E., Dechow, R., Doro, L., Eckersten, H., Gaiser, T., Grosz, B., Heinlein, F., Kassie, B.T., Kersebaum, K.-C., Klein, C., Kuhnert, M., Lewan, E., Moriondo, M., Nendel, C., Priesack, E., Raynal, H., Roggero, P.P., Rötter, R.P., Siebert, S., Specka, X., Tao, F., Teixeira, E., Trombi, G., Wallach, D., Weihermüller, L., Yeluripati, J., Ewert, F., 2016. Impact of Spatial Soil and Climate Input Data Aggregation on Regional Yield Simulations. PLOS ONE 11, e0151782. https://doi.org/10.1371/journal.pone.0151782

Iooss, B., Janon, A., Pujol, G., Boumhaout, with contributions from K., Veiga, S.D., Delage, T., Fruth, J., Gilquin, L., Guillaume, J., Gratiet, L.L., Lemaitre, P., Nelson, B.L., Monari, F., Oomen, R., Rakovec, O., Ramos, B., Roustant, O., Song, E., Staum, J., Sueur, R., Touati, T., Weber, F., 2018. sensitivity: Global Sensitivity Analysis of Model Outputs.

Iooss, B., Lemaître, P., 2014. A review on global sensitivity analysis methods. ArXiv14042405 Math Stat.

Iooss, B., Saltelli, A., 2016. Introduction to Sensitivity Analysis, in: Ghanem, R., Higdon, D., Owhadi, H. (Eds.), Handbook of Uncertainty Quantification. Springer International Publishing, Cham, pp. 1–20. https://doi.org/10.1007/978-3-319-11259-6_31-1

Jansen, M.J.W., 1999. Analysis of variance designs for model output. Comput. Phys. Commun. 117, 35–43. https://doi.org/10.1016/S0010-4655(98)00154-4

Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., Monod, H., 2009. Multivariate global sensitivity analysis for dynamic crop models. Field Crops Res. 113, 312–320. https://doi.org/10.1016/j.fcr.2009.06.007

Liski, J., Tuomi, M., Rasinmäki, J., 2009. Yasso07 user-interface manual 14.

Lo Piano, S., Ferretti, F., Puy, A., Albrecht, D., Tarantola, S., Saltelli, A., 2019. Is it possible to improve existing sample-based algorithm to compute the total sensitivity index? 26. https://doi.org/arXiv:1703.05799v2 [stat.AP]

Luo, Z., Wang, E., Bryan, B.A., King, D., Zhao, G., Pan, X., Bende-Michl, U., 2013. Meta-modeling soil organic carbon sequestration potential and its application at regional scale. Ecol. Appl. 23, 408–420.

Luo, Z., Wang, E., Fillery, I.R.P., Macdonald, L.M., Huth, N., Baldock, J., 2014. Modelling soil carbon and nitrogen dynamics using measurable and conceptual soil organic matter pools in APSIM. Agric. Ecosyst. Environ. 186, 94–104. https://doi.org/10.1016/j.agee.2014.01.019

Luo, Z., Wang, E., Shao, Q., Conyers, M.K., Liu, D.L., 2016. Confidence in soil carbon predictions undermined by the uncertainties in observations and model parameterisation. Environ. Model. Softw. 80, 26–32. https://doi.org/10.1016/j.envsoft.2016.02.013

Nguyen, A.-T., Reiter, S., 2015. A performance comparison of sensitivity analysis methods for building energy models. Build. Simul. 8, 651–664. https://doi.org/10.1007/s12273-015-0245-4

Ogle, S.M., Breidt, F.J., Easter, M., Williams, S., Killian, K., Paustian, K., 2010. Scale and uncertainty in modeled soil organic carbon stock changes for US croplands using a process-based model. Glob. Change Biol. 16, 810–822. https://doi.org/10.1111/j.1365-2486.2009.01951.x

Post, J., Hattermann, F.F., Krysanova, V., Suckow, F., 2008. Parameter and input data uncertainty estimation for the assessment of long-term soil organic carbon dynamics. Environ. Model. Softw. 23, 125–138. https://doi.org/10.1016/j.envsoft.2007.05.010

Qin, F., zhao, Y., Shi, X., Xu, S., Yu, D., 2017. Uncertainty and Sensitivity Analyses for Modeling Long-Term Soil Organic Carbon Dynamics of Paddy Soils Under Different Climate-Soil-Management Combinations. Pedosphere 27, 912–925. https://doi.org/10.1016/S1002-0160(17)60436-3

R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. 181, 259–270. https://doi.org/10.1016/j.cpc.2009.09.018

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2007. Global Sensitivity Analysis. The Primer. John Wiley & Sons, Ltd, Chichester, UK.

White, R.E., Davidson, B., Lam, S.K., Chen, D., 2018. A critique of the paper ‘Soil carbon 4 per mille’ by Minasny et al. (2017). Geoderma 309, 115–117. https://doi.org/10.1016/j.geoderma.2017.05.025

Witing, F., Gebel, M., Kurzer, H.-J., Friese, H., Franko, U., 2019. Large-scale integrated assessment of soil carbon and organic matter-related nitrogen fluxes in Saxony (Germany). J. Environ. Manage. 237, 272–280. https://doi.org/10.1016/j.jenvman.2019.02.036

Wutzler, T., Reichstein, M., 2007. Soils apart from equilibrium – consequences for soil carbon balance modelling. Biogeosciences 4, 125–136. https://doi.org/10.5194/bg-4-125-2007

Appendix

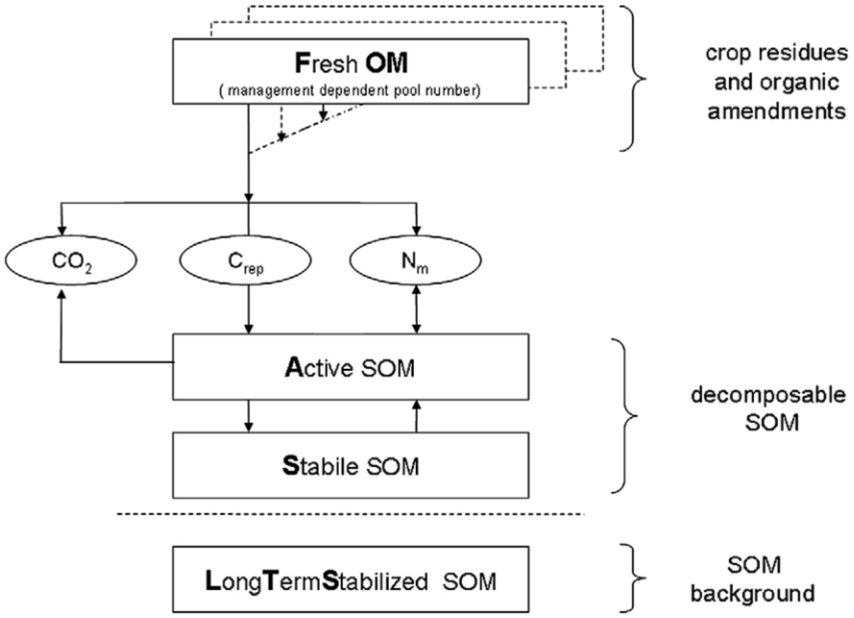

Figure A-1: Block scheme of the CCB model with pools (rectangles) and fluxes (ovals) (from Franko et al. 2011). C_{rep}: carbon reproduction flux from fresh organic matter (FOM) to soil organic matter (SOM). N_m: nitrogen flux changing the external pool of mineral nitrogen, which is not included in the model. CO_2: release of carbon dioxide. LTS-SOM: long-term stabilized soil organic matter with no turnover during simulation time

The turnover of the pools is based on first order kinetics. The time step is represented through Biologic Active Time (BAT) concept, that is further explained in (Franko and Oelschlägel, 1995 and Franko and Spiegel, 2016) or the CCB manual. Each FOM is characterised by a specific turnover rate k_i (A1) and a synthesis coefficient η_i (A2). Also the SOM pools have specific k_i values, with k_m being the loss of CO_2 into the atmosphere (A3). For the initial distribution between $C_{A\text{-}SOM}$ and $C_{S\text{-}SOM}$, please refer to (Franko et al., 2011). In our study, p_yield, p_bp and p_oram had influence on the amount of C_{FOM}, p_cons influenced the BAT and p_soc influenced the amount of the initial values of A-SOM and S-SOM.

$$\dot{C}_{FOM} = \frac{dC_{FOM}}{dt} = -k_{FOM} \cdot C_{FOM} \tag{A1}$$

$$\dot{C}_{rep} = \dot{C}_{fom} \cdot \eta_{fom} \tag{A2}$$

$$\dot{C}_{A-SOM} = \dot{C}_{rep} + k_a \cdot C_{S-SOM} - k_s \cdot C_{A-SOM} - k_m \cdot C_{A-SOM} \tag{A3}$$

$$\dot{C}_{S-SOM} = k_s \cdot C_{A-SOM} - k_a \cdot C_{S-SOM} \tag{A4}$$

The carbon and Nitrogen fluxes are coupled via pool-specific C/N ratios, indicated as γ_i. Each FOM can have a different γ_i and the needed amount of N ($N_{rep,}$ Eq. A6) for the reproduction of A-SOM might vary from year to year, so there can be a surplus (positive values) or a demand (negative values) of mineralized N (OM-N_{min}).

$$\dot{N}_{FOM} = \dot{C}_{FOM} \cdot \frac{1}{\gamma_{FOM}} \tag{A5}$$

$$\dot{N}_{rep} = \dot{C}_{fom} \cdot \eta_{fom} \cdot \frac{1}{\gamma_{A-SOM}} \tag{A6}$$

$$\dot{N}_{A-SOM} = k_m \cdot C_{A-SOM} \cdot \frac{1}{\gamma_{A-SOM}} \tag{A7}$$

$$OM\dot{} - N_{min} = \dot{N}_{rep} + \dot{N}_{A-SOM} - \dot{N}_{FOM} \tag{A8}$$