Sensitivity Analysis of Soil Parameters in Crop Model Supported with High-Throughput Computing

Mikhail Gasanov[1]($\boxtimes$) , Anna Petrovskaia[1] , Artyom Nikitin[1] ,
Sergey Matveev[1,2] , Polina Tregubova[1] , Maria Pukalchik[1] ,
and Ivan Oseledets[1,2]

[1] Skolkovo Institute of Science and Technology,
Bolshoy Boulevard 30, bld. 1, Moscow 121205, Russia
`Mikhail.Gasanov@skoltech.ru`
[2] Marchuk Institute of Numerical Mathematics, RAS,
Gubkin st. 8, Moscow 119333, Russia
`http://www.skoltech.ru, http://www.inm.ras.ru`

Abstract. Uncertainty of input parameters in crop models and high costs of their experimental evaluation provide an exciting opportunity for sensitivity analysis, which allows identifying the most significant parameters for different crops. In this research, we perform a sensitivity analysis of soil parameters which play an essential role in plant growth for the MONICA agro-ecosystem model. We utilize Sobol' sensitivity indices to estimate the importance of main soil parameters for several crop cultures (soybeans, sugar beet and spring barley). High-throughput computing allows us to speed up the computations by more than thirty times and increase the number of sampling points significantly. We identify soil indicators that play an essential role in crop yield productivity and show that their influence is the highest in the topsoil layer.

Keywords: Crop model · Sobol' indices · Soil parameters UQ

1 Introduction

Numerical digital crop models are used for crop yield prediction worldwide nowadays [24] and have applications in decision-support systems for farmers [10,15]. These models require soil, environmental and agro-management input data to establish plant growth simulation. The most widespread crop models, such as CENTURY [13], APSIM [5], DNDC [2] and MONICA [11] include modules of soil processes, climate and crop properties and allow to improve model's forecast with the calibration of internal parameters. Unfortunately, measurements of soil parameters for spatial modeling might be expensive and time-consuming, especially in countries where agrochemical data is not freely available.

Various approaches to reduce the number of parameters in environmental models have been proposed [14]. One of the most promising tools is evaluation

of the performance of complex environmental models through sensitivity analysis (SA) [22]. Sensitivity analysis simplifies the process of modeling by identifying and removing unnecessary elements from the structure of the model. There is a series of recent publications regarding the assessment of soil and plant sensitivity indicators conducted by Krishnan [8], Zhang [26], Karki [7], Gunarathna [3].

In practice, sensitivity analysis involves a) sampling of the multidimensional input parameter space; and b) subsequent simulations of the model. To obtain reliable confidence intervals, it may require millions of simulation runs, which may be infeasible for general-purpose computers. In previous works, the number of input samples was limited to 2000–30 000 points [12, 23], which may be insufficient in other settings, where the amount of varied parameters is much larger. A natural way to speed up the simulations is to distribute them into independent blocks and perform parallel computations using a supercomputer [6].

In this work, we develop an effective and fast method for computing more than 500 000 agro-ecosystem MONICA model simulations per hour. It allows us to consider a much broader class of problems of practical interest. In particular, we increase the number of sampling points for sensitivity analysis of parameters up to 100 000, efficiently distribute calculations using a supercomputer and perform 2 000 000 model runs.

2   Materials and Methods

In this section, we describe the materials and methods that we use in our work.

2.1   MONICA Agro-Ecosystem Model

There is a variety of commercial and open-source models for crop growth simulations and yield prediction. In our research, we choose an open-source process-based agro-ecosystem model MONICA [11] that has been developed by ZALF institute during the last decades (Müncheberg, Germany). As input, MONICA requires soil parameters, crop rotation, fertilization schedule, and climate data.

Even though MONICA was developed for Western Europe soil conditions and climate, it can be optimized to other crop types by using model parameters for physiology and plant development. The MONICA model includes more than 120 parameters in soil hydrology processes, soil nutrients and organic matter turnover, plant physiology, and other blocks responsible for different processes that influence crop yield. MONICA receives soil data as different depth horizons (layers of soil with relatively uniform properties), which can be set up by a user in the format of a JSON-file.

2.2   Soil Parameters

The selection of parameters and their bounds play an essential role in sensitivity analysis [17]. In our research, we use agricultural data from a field experiment in the Russian Chernozem region. Among the great variety of climate conditions

and soil types in Russia, the Chernozem region has special significance because of its potential productivity due to the highest nutrition and carbon content. We examine the commercial field in Kursk, Russia (51°52′20″N, 37°50′52″E) with six years crop-rotation from 2011 to 2017. The crop rotation consists of three different crops, namely sugar beet (*Beta vulgaris*) for years 2011, 2014, 2017, spring-barley (*Hordeum vulgare*) in 2012, and soybean (*Glycine max*) in 2015. The soil profile consists of several layers (or horizons). The upper arable horizon is especially crucial for the growth and development of crops. Subsoil layers may take part in hydrology regime and affect plant growth as well.

MONICA model requires more than ten different parameters for simulation within each soil layer. We select six main soil parameters (see Table 1) for sensitivity analysis (Soil organic carbon, Soil pH, Clay content, Sand content, Carbon:Nitrogen ratio, Bulk density). These parameters represent significant soil properties and have a considerable impact on crop yield. The value boundaries for the parameters were taken from the Russian Soil database [1] and represent the actual values for chernozem soils. In our research, we concentrate on crop yield ($kgDryMatter * ha^{-1}$) as an output of the MONICA model for sensitivity analysis. Prediction of yield is a complicated task because the yield depends on almost all processes in an agricultural system.

Table 1. Soil parameters of MONICA model and their min/max values used in SA.

Parameter	Description	Unit	Min.	Max.
SOC	Soil organic carbon	%	2.58	6.20
Sand	Soil sand fraction	$kg * kg^{-1}$	0.01	0.30
Clay	Soil clay fraction	$kg * kg^{-1}$	0.01	0.30
pH	Soil pH value	–	4.6	6.9
CN	Soil carbon:nitrogen ratio	–	10.9	12.4
BD	Soil bulk density	$kg * m^{-3}$	900.0	1350.0

To identify the most critical horizons, we evaluate the sensitivity indices of soil parameters at various depths. MONICA model allows us to set up soil layers with various thickness and parameters. We set nine layers with different thickness typical for agricultural soils of the Chernozem region as follows: topsoil 30 cm, seven horizons with 10 cm depth and the subsoil layer with 100 cm depth. We iteratively select each parameter from the Table 1 and evaluate how changes of this parameter in each soil layer affect the model's predictions. After identifying the most influential (in terms of crop yield) horizon, we perform sensitivity analysis of all six soil parameters for this layer specifically.

2.3   The Sobol' Sensitivity Analysis

Sensitivity analysis is a methodology of qualitative investigation of a model and its parameters which helps to identify parameters affecting the output of the

model. It is possible to distinguish local and global sensitivity analyses. In our research, we choose the method developed by Sobol' [19] for global SA.

Consider the model output as

$$Y = f(X) = f(X_1, \ldots, X_p),$$

where f in our case depicts MONICA simulator, X are p varied input parameters and Y is the predicted crop yield. Following the techniques by Sobol' [21] we represent the multi-variate random function f using Hoeffding decomposition:

$$f(X_1, \ldots, X_p) = f_0 + \sum_i^p f_i + \sum_i^p \sum_{j>i}^p f_{ij} + \cdots + f_{1\ldots p}, \tag{1}$$

where f_0 is a constant term, $f_i = f_i(X_i)$ denotes main effects, $f_{ij} = f_{ij}(X_i, X_j)$ and others describe higher-order interactions. These terms can be written as

$$f_0 = E(Y),$$
$$f_i = E_{X_{\sim i}}(Y|X_i) - E(Y),$$
$$f_{ij} = E_{X_{\sim ij}}(Y|X_i, X_j) - f_i - f_j - f_0,$$
$$\cdots$$

where E is mathematical expectation and $X_{\sim i}$ denotes all parameters except i^{th}. Under the assumption that the input parameters are independent, total variance $V(Y)$ of the crop yield can be decomposed as follows:

$$V(Y) = \sum_i^p V_i + \sum_i^p \sum_{j>i}^p V_{ij} + \cdots + V_{12\ldots p},$$

where partial variances are

$$V_i = V[f_i(X_i)] = V_{X_i}\left[E_{X_{\sim i}}(Y|X_i)\right],$$
$$V_{ij} = V[f_{ij}(X_i, X_j)] = V_{X_i X_j}\left[E_{X_{\sim ij}}(Y|X_i, X_j)\right] - V_i - V_j,$$
$$\cdots$$

This way, sensitivity indices (SI) can be introduced as

$$S_i = \frac{V_i}{V(Y)}, \ \ S_{ij} = \frac{V_{ij}}{V(Y)}, \ \ \cdots \tag{2}$$

One can note the total sum of the indices is normalized to 1. The value of the Sobol' index corresponds to the "order" of sensitivity of f to the change of the corresponding input parameter or their group (see the details in [19] or [21]). Analogously to Eq. 1, first-order indices denote variance induced by changes of a single parameter without any interactions; second-order indices consider second-order interactions between the parameters; etc. In order to incorporate all of the interactions for a particular parameter, one can compute the total effect index:

$$S_{T_i} = \frac{E_{X_{\sim i}}\left[V_{X_i}(Y|X_{\sim i})\right]}{V(Y)} = 1 - \frac{V_{X_{\sim i}}\left[E_{X_i}(Y|X_{\sim i})\right]}{V(Y)} \tag{3}$$
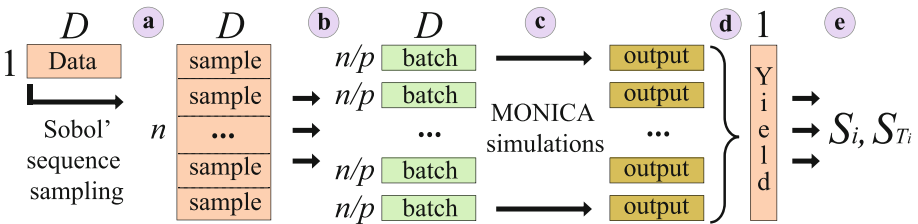
Fig. 1. Distributed computations comprising 5 steps: a) Sobol' sequence sampling from the initial input data, where D is the number of perturbed input parameters, $n = N \times (2 \times D + 2)$ is the total number of simulations and N is sampling size; b) mapping of acquired samples in batches to p HTC nodes; c) running n/p parallel MONICA simulations on each node; d) aggregating yield values from simulation results; and e) calculation of Sobol' sensitivity indices.

Evaluation of Sobol' indices requires us to perform a random sampling of the parameter hyperspace. In our work, we utilize quasi-random sampling approach. In general, such methods add new points into the sequence taking into account previously added points and may create a uniformly filled parameter space in the unit hypercube. In our work, we use the classical approach also proposed by Sobol' [18,20], which helps to achieve a convergence rate of confidence intervals almost $O(N^{-1})$, where N is the number of samples [9].

2.4   Crop Simulation and High-Throughput Computing

Sensitivity analysis requires the results of a significant number of simulations with various parameters. The number of simulations necessary for the convergence of sensitivity indices can be computationally expensive. In our work, we use "Zhores" supercomputer to tackle this problem [25]. Figure 1 outlines our approach. First, the initial values of D parameters are used to generate the $n \times D$ matrix of samples using Sobol' quasi-random sequence, more particularly, its extension proposed by Saltelli [16], where $n = N \times (2 \times D + 2)$ and N is the sample size. Second, these samples are grouped into batches of size $n/p \times D$ and each batch is then mapped to one of p HTC nodes. Third, each node performs n/p MONICA simulations in parallel. Then, yield predictions are extracted from output results and aggregated back to a vector of size n. Finally, these yield values are used to calculate Sobol' sensitivity indices using SALib Python library [4]. The most computationally expensive step is running the simulations, whereas the generation of samples and sensitivity analysis are negligible (several minutes in our experiments compared to hours of simulations).

To obtain the convergence of confidence intervals for sensitivity indices we use a different number of input sample points (from 10 to 100 000) to find the optimal amount needed for SA. To evaluate the acceleration, we compare the time spent for 2 000 000 simulations on a single core and $p = 96$ *Intel Xeon C6140* cores. This set of simulations is the main time-consuming part
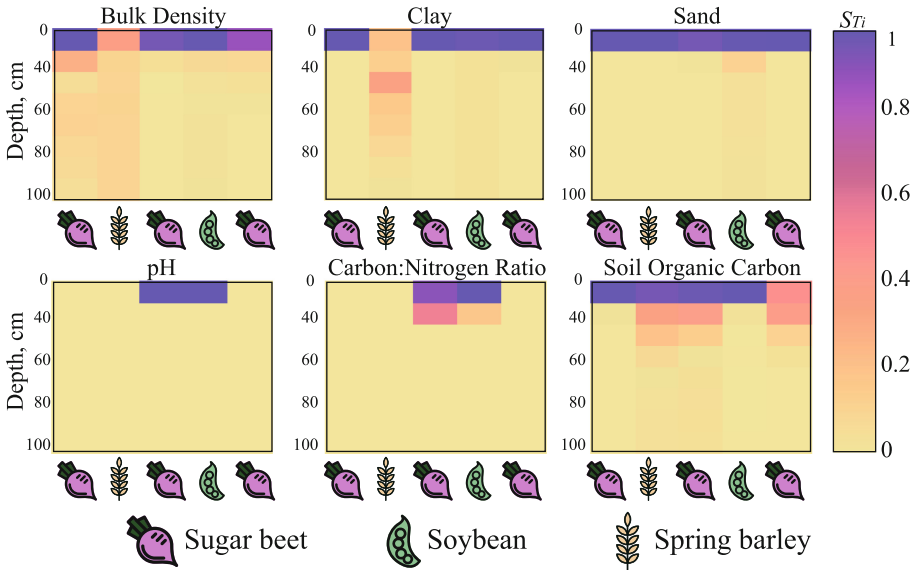
Fig. 2. Heatmap of S_{T_i} index for six soil parameters across different horizons in the five-year crop rotation. For the majority of crops, the most considerable role in yield variation is played by the change of parameters in the upper horizon. (Color figure online)

of computational work. It takes almost $\sim$112 h to calculate this on one core, and 3.5 h on 96 cores. The acceleration is 32.5 times and limited mostly by the performance of the file system. The speedup is defined by $S = t_s/t_p$, where t_s and t_p are the time for serial and parallel model simulations. We could have achieved additional acceleration of computations via the storage of simulation input files and technical outputs in RAM instead of direct creation and removal of files on hard-drives. We plan to do it in our future work.

3   Results and Discussion

In this section, we investigate the effect of input soil data on crop yields and provide our experimental results. For this purpose, we select six principal soil parameters important for plant growth, develop and evaluate the sensitivity of the model for each indicator at different soil depths. To demonstrate which soil horizons have the most significant impact on crop productivity, first, we divide the soil profile into nine horizons of different thickness and, second, perform separate sensitivity analyses for each of the parameters from the Table 1.

We present the obtained results in Fig. 2 in the form of heatmaps displaying the soil profile. Crop rotation and the depth of soil horizons are represented with X and Y axes, respectively. We use a sample size N equal to 100 000 and conduct
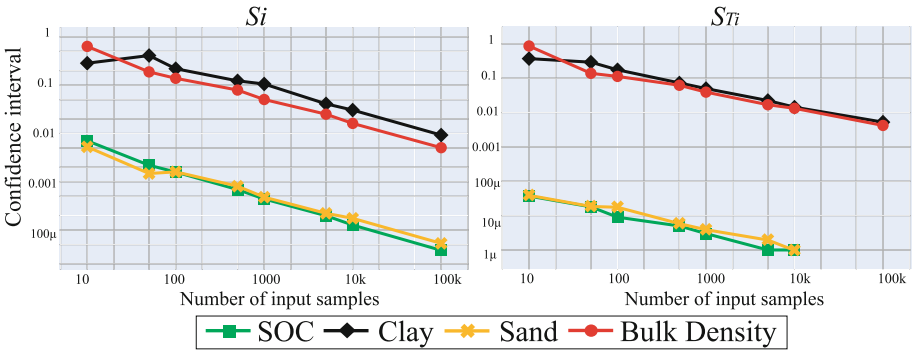
Fig. 3. The convergence of S_i (left) and S_{T_i} (right) confidence intervals for the topsoil layer of Sugar beet crop (2017 year) with different input sample sizes N. It can be seen that the convergence rate achieved is equivalent to $O(N^{-1})$.

2 000 000 simulations for each soil parameter, which allows us to obtain suitable confidence intervals. Color depicts the values of the total-order Sobol' SI, where light yellow color indicates no impact of parameter variation on the crop yield, and purple color indicates significant influence. We conclude that for most of the considered crops the main influence of parameters on final yield concentrates in the top horizon. However, the clay fraction and soil organic carbon affect the yield of barley in the entire soil profile. The effect of soil organic matter content on sugar beet yields changes during crop rotation. The content of organic matter in the upper horizon affects the yield only at the beginning of crop rotation. The transformation of organic matter in the soil leads to the distribution of organic compounds along the profile, and the subsurface horizons begin to affect the yield of sugar beet. For further analysis of soil parameters, we concentrate only on the upper horizon of 0–30 cm.

To identify the parameters in the topsoil layer that have the most significant impact on crop yield, we analyze first-order S_i and total effect S_{T_i} indices. One of the necessary conditions for successful SA is the convergence of the obtained SI values. As noted above, we use quasi-random sampling method proposed by Sobol' to increase the convergence rate of sensitivity indices values with a sample size N varying from 10 to 100 000. Figure 3 demonstrates convergence of the confidence intervals for S_i and S_{T_i} indices of the main six parameters, which achieves the rate of $O(N^{-1})$ for Sugar beet crop (2017). For other crops, we obtain the same qualitative results.

Figure 4 shows S_i and S_{T_i} values and additionally their confidence intervals for different cultures: spring barley (2012), soybean (2015) and sugar beet (2017). We exclude the plots for two other years of Sugar beat because they are qualitatively the same as in 2017. Soil parameters with sensitivity indices close to zero achieved stable values faster than the parameters with higher indices. Significant model parameters (which have strongly nonzero sensitivity indices)
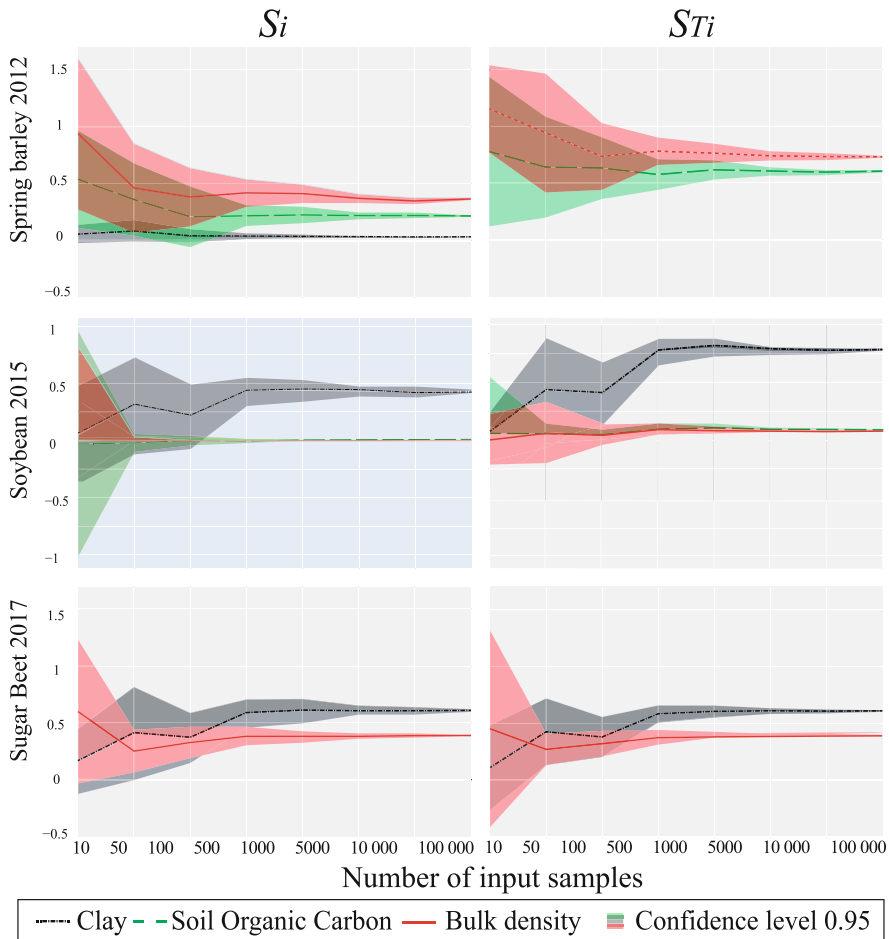
Fig. 4. Values of S_i (left) and S_{T_i} (right) indices for different crops, sample sizes N and soil parameters. Filled regions depict confidence intervals of the respective indices. Some of the parameters have rapid convergence of their confidence intervals because their Sobol' indices are very close to zero.

required input samples size from 1000 to 5000. It can be seen that different soil parameters have different importance for crop yield. Soil organic carbon content plays an essential role in all crops. The change in bulk density and soil organic carbon has a more significant impact on spring barley yields than on other indicators. Almost all the difference in soybean yield is due to the change in the soil clay fraction. The yield of sugar beet depends on the content of soil clay and bulk density, which coincides with real data, since beets are demanding to water nutrition, and clay content and bulk density can affect water regime. The values S_i and S_{T_i} for soil pH and carbon-nitrogen ratio in soil organic matter were almost equal to zero. It seems strange that they did not affect the crop yield.
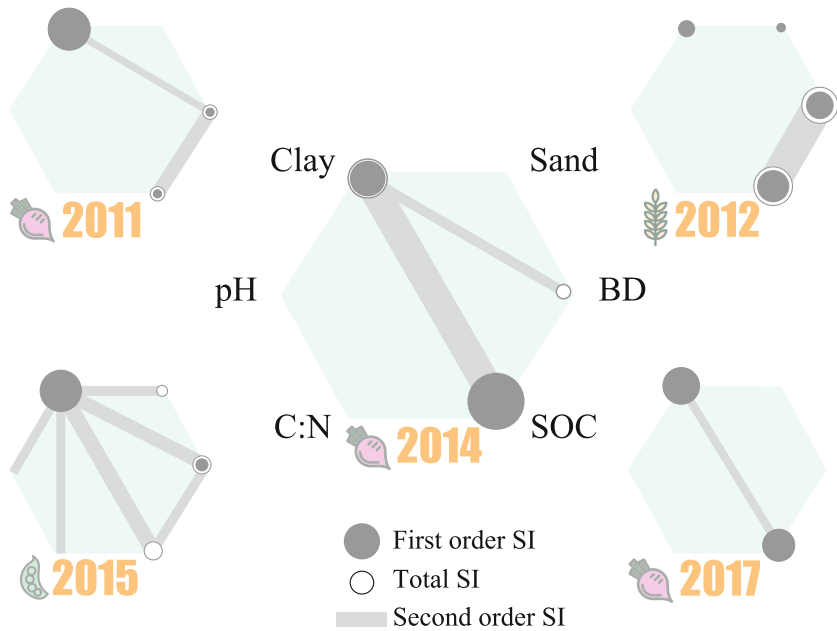
Fig. 5. First-order, second-order and total Sobol' sensitivity indices' for the six soil parameters and three crops: sugar beet (2011, 2014, 2017), spring barley (2012) and soybean (2015).

Considering that pH condition of soil determines the availability of nutrients for plants, and carbon to nitrogen ratio shows the quality of organic matter, it could influence the activity of soil microbial community. We plan to provide a more detailed analysis of corresponding MONICA submodels in our future research.

To represent second-order effects we employ diagrams in Fig. 5. It shows S_i, S_{ij} and S_{T_i} for different crop rotations, where gray lines describe interactions between soil parameters, black and white circles denote first-order and total sensitivity indices, respectively. There are total SI values for sand content and organic matter for soybean, which indicates the importance of the interactions of these parameters for yield. The figure demonstrates that clay, soil organic carbon, and bulk density are the parameters that have the highest value of first-order SI for the majority of years. As the first-order SI measures the fractional contribution of a single parameter to the output variance, we conclude that these parameters have the most significant impact on yield in our case. Second-order sensitivity indices in the figure show which parameter interactions play the biggest role in yield prediction. The soybean yield is affected by second-order interactions between almost all indicators. In contrast, spring barley yield is affected only by the coupled effect of soil density and organic matter content. From Fig. 5 we also see that the importance of SOC parameter and values of the second-order Sobol' indices change in time due to the transformation of soil organic matter during the crop rotation.

4   Conclusions

In this research, we investigated the importance of different soil parameters in MONICA crop simulation model. For this purpose, we used a variance-based sensitivity analysis method developed by Sobol'. For successful convergence of the algorithm it requires numerous runs of simulations, and to tackle this problem, we applied high throughput computing. Our results indicate that for each studied crop a different set of soil parameters affects the yield. The transformation of organic matter in the soil during the crop rotation modifies the importance of this parameter for sugar beet productivity. The results show the presence of collective influence of model input parameters on crop productivity. Moreover, for the selected region of study the C:N ratio and pH parameters could be excluded from MONICA, or corresponding submodels should be updated accordingly. The source code and the results are freely available at our GitHub repository[1].

References

1. Russian Soil Database, Soil Science Institute named by V.V. Dokuchaev (2010). http://egrpr.esoil.ru/content/norm.html
2. Giltrap, D.L., Li, C., Saggar, S.: DNDC: a process-based model of greenhouse gas fluxes from agricultural soils. Agr. Ecosyst. Environ. **136**(3–4), 292–300 (2010)
3. Gunarathna, M., Sakai, K., Nakandakari, T., Momii, K., Kumari, M.: Sensitivity analysis of plant-and cultivar-specific parameters of APSIM-sugar model: variation between climates and management conditions. Agronomy **9**(5), 242 (2019)
4. Herman, J.D., Usher, W.: SALib: an open-source python library for sensitivity analysis. J. Open Source Softw. **2**(9), 97 (2017)
5. Holzworth, D.P., et al.: APSIM-evolution towards a new generation of agricultural systems simulation. Environ. Model Softw. **62**, 327–350 (2014)
6. Huang, X., et al.: A dynamic agricultural prediction system for large-scale drought assessment on the Sunway Taihulight supercomputer. Comput. Electron. Agric. **154**, 400–410 (2018)
7. Karki, R., Srivastava, P., Bosch, D.D., Kalin, L., Lamba, J., Strickland, T.C.: Multi-variable sensitivity analysis, calibration, and validation of afield-scale SWAT model: building stakeholder trust in hydrologic/water quality modeling. Trans. ASABE **63**, 523–539 (2020)
8. Krishnan, P., Aggarwal, P.: Global sensitivity and uncertainty analyses of a web based crop simulation model (web InfoCrop wheat) for soil parameters. Plant Soil **423**(1–2), 443–463 (2018). https://doi.org/10.1007/s11104-017-3498-0
9. Kucherenko, S., Rodriguez-Fernandez, M., Pantelides, C., Shah, N.: Monte Carlo evaluation of derivative-based global sensitivity measures. Reliab. Eng. Syst. Safe. **94**(7), 1135–1148 (2009)

---

[1] https://github.com/mishagrol/SA_agro_model.

10. Lavik, M.S., Hardaker, J.B., Lien, G., Berge, T.W.: A multi-attribute decision analysis of pest management strategies for Norwegian crop farmers. Agric. Syst. **178**, 102741 (2020)
11. Nendel, C., et al.: The MONICA model: testing predictability for crop growth, soil moisture and nitrogen dynamics. Ecol. Model. **222**(9), 1614–1625 (2011)
12. Nossent, J., Elsen, P., Bauwens, W.: Sobol' sensitivity analysis of a complex environmental model. Environ. Model Softw. **26**(12), 1515–1525 (2011)
13. Parton, W.J., Stewart, J.W., Cole, C.V.: Dynamics of C, N, P and S in grassland soils: a model. Biogeochemistry **5**(1), 109–131 (1988). https://doi.org/10.1007/BF02180320
14. Razavi, S., Gupta, H.V.: What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and environmental systems models. Water Resour. Res. **51**(5), 3070–3092 (2015)
15. Rurinda, J., et al.: Science-based decision support for formulating crop fertilizer recommendations in sub-Saharan Africa. Agric. Syst. **180**, 102790 (2020)
16. Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S.: Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput. Phys. Commun. **181**(2), 259–270 (2010)
17. Saltelli, A., Tarantola, S., Campolongo, F., Ratto, M.: Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, vol. 1. Wiley, New York (2004)
18. Sobol, I.M.: Uniformly distributed sequences with an additional uniform property. USSR Comput. Math. Math. Phys. **16**(5), 236–242 (1976)
19. Sobol, I.M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. **55**(1–3), 271–280 (2001)
20. Sobol', I.M.: On the distribution of points in a cube and the approximate evaluation of integrals. Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki **7**(4), 784–802 (1967)
21. Sobol', I.M.: On sensitivity estimation for nonlinear mathematical models. Matematicheskoe modelirovanie **2**(1), 112–118 (1990)
22. Varella, H., Guérif, M., Buis, S.: Global sensitivity analysis measures the quality of parameter estimation: the case of soil parameters and a crop model. Environ. Model Softw. **25**(3), 310–319 (2010)
23. Vazquez-Cruz, M., Guzman-Cruz, R., Lopez-Cruz, I., Cornejo-Perez, O., Torres-Pacheco, I., Guevara-Gonzalez, R.: Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. Comput. Electron. Agric. **100**, 1–12 (2014)
24. Webber, H., Hoffmann, M., Rezaei, E.E.: Crop models as tools for agroclimatology. Agroclimatol. Link. Agric. Clim. **60**, 519–546 (2020)
25. Zacharov, I., et al.: "Zhores"–petaflops supercomputer for data-driven modeling, machine learning and artificial intelligence installed in Skolkovo Institute of Science and Technology. Open Eng. **9**(1), 512–520 (2019)
26. Zhang, Y., Arabi, M., Paustian, K.: Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. Environ. Model Softw. **126**, 104642 (2020)