University of São Paulo
"Luiz de Queiroz" College of Agriculture

Sensitivity and uncertainty analysis of SAMUCA crop model across contrasting environments in Brazil

Rodolfo Armando de Almeida Pereira

Thesis presented to obtain the degree of Doctor in Science.
Area: Agricultural Systems Engineering

Piracicaba
2023

Rodolfo Armando de Almeida Pereira
Bachelor in Atmospheric Sciences

Sensitivity and uncertainty analysis of SAMUCA crop model across contrasting environments in Brazil

versão revisada de acordo com a Resolução CoPGr 6018 de 2011

Advisor:
Prof. Dr. **FÁBIO RICARDO MARIN**

Thesis presented to obtain the degree of Doctor in Science.
Area: Agricultural Systems Engineering

Piracicaba
2023

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Pereira, Rodolfo Armando de Almeida

Sensitivity and uncertainty analysis of SAMUCA crop model across contrasting environments in Brazil / Rodolfo Armando de Almeida Pereira. - - versão revisada de acordo com a Resolução CoPGr 6018 de 2011. - - Piracicaba, 2023.

107 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura "Luiz de Queiroz".

1. Cana de açúcar 2. Modelo de cultura baseado em processos 3. Análise de sensibilidade global I. Título

ACKNOWLEDGEMENTS

To my parents, Armando Eugênio and Maria Fátima, for supporting me and being role models, even in adverse situations;

To my brother, Dr. Guilherme Armando, for abetment, support, and help;

To my friends MSc. Cleverson, Dr. Flávio, Dr. Dário, Dr. Evandro, Dr. Luis, and MSc. Juliano for the happy and relaxed moments we spent on this arduous journey;

To Aline Munhoz, for their love, support, teaching, and much of patience;

To my advisor, Prof. Dr, Fábio Ricardo Marin, for his mentoring and assistance during the development of this research and since my master's degree (2016), as well as, the friendship, teaching, and motivational support;

To my co-advisor, Dr. Murilo dos Santos Vianna, for their mentoring, assistance, and teachings;

To ESALQ-CLIMA group for tremendous support on field experiments, especially by collecting and providing the 4th year dataset of the long-term experiment and also for workshops and practicing teaching and mentoring skills with undergraduate students;

To CAPES for the financial support;

To ESALQ-USP, for the support during the activities carried out during the course;

Thank you all so much for making this dream a reality!

EPIGRAPH

"We can only see a short distance ahead, but we can see plenty there that needs to be done."
Alan Mathison Turing.

CONTENTS

RESUMO

Análise de sensibilidade e incerteza do modelo SAMUCA em ambientes contrastantes no Brasil

A indústria canavieira brasileira está em constante desenvolvimento, testando e lançando novas cultivares e práticas de manejo para aumentar a produtividade. Devido às questões das mudanças climáticas e às limitações para expandir as áreas agrícolas, as fazendas brasileiras são constantemente pressionadas a aumentar a eficiência da produção. A utilização de modelos baseados em processos (PBCM) para testar cultivares e opções de manejo em diferentes ambientes de produção é uma realidade e vem sendo cada vez mais utilizada. Os PBCM são o estado da arte em modelagem agrícola e são cada vez mais complexos, requerendo diversos parâmetros para descrever os processos de cultivo e as condições de contorno. Em geral, os PBCM utilizam a abordagem determinística para simplificar a incerteza presente no ambiente usando um único conjunto de parâmetros. Na prática, essa incerteza é vista na variabilidade dos dados coletados em um experimento de campo, que são comumente representados por estatísticas de dispersão, como desvio padrão e variância. Uma maneira de explorar essa incerteza é usar a abordagem estocástica, inserindo uma faixa de variabilidade nos parâmetros e entradas da simulação. Este estudo teve como objetivo utilizar a abordagem estocástica para explorar a incerteza e determinar quais parâmetros do modelo SAMUCA são mais influentes no processo de simulação. Para isso foi utilizada a recente versão do modelo SAMUCA inserindo três cenários de incerteza: análise de incerteza apenas para parâmetros genéticos (UG), análise de incerteza apenas para parâmetros de solo (US) e análise de parâmetros de solo e genótipo (UGS). Nessa primeira etapa foram simulados esses três cenários para um experimento de campo de 4 anos, sendo a cultura cultivada sob efeito da palha (GCTB) e solo nu (Bare). A partir disso foi quantificado a variabilidade da simulação estocástica pela razão entre a média do desvio padrão das simulações e a média do desvio padrão dos dados observados. Posteriormente, para entender melhor quais os fatores que causam maior incerteza no processo de simulação foi relizada uma análise de sensibilidade global (GSA) pelo método *extented Fourier Amplitude Sensitivity Test* (eFAST) para o mesmo experimento de campo de 4 anos, visando a identificar quais parâmetros foram responsáveis por explicar a maior variância do modelo e verificar o impacto do intervalo dos parâmetros escolhidos, bem como o número de simulações necessárias para se ter uma GSA confiável. Por fim, sabendo que além do método, o ambiente pode influenciar o resultado da GSA, fez-se uma nova análise de sensibilidade com dois métodos , eFAST e *Partial Rank Correlation Coefficient* (PRCC) para as principais regiões produtoras de cana de açúcar no Brasil, considerando condições irrigadas e de sequeiro. Os resultados indicaram que a variabilidade observada no campo não é totalmente explicada pelos parâmetros do solo, possivelmente devido à irrigação e boa distribuição das chuvas na área experimental. A UG e a UGS tiveram a mesma capacidade de quantificar a variabilidade presente no campo experimental. Nesse caso, a sensibilidade aos parâmetros do solo poderia ser simplesmente ignorada e os parâmetros genéticos podem ser escolhidos como a única fonte de variabilidade para aplicações práticas. A maior parte da incerteza nesse experimento é atribuída ao parâmetro *plastochron*, porém identificou-se que o conjunto de intervalo dos parâmetros pode influenciar a ordem dos parâmetros mais importantes. Isso foi observado quando se realizou a análise para dois conjuntos de intervalos de parâmetros diferentes (o primeiro conjunto usou valores máximos e mínimos relatados na literatura; o segundo conjunto aplicou uma perturbação de 25% nos valores previamente calibrados). Por fim, dos 31 parâmetros, 24 genéticos e 7 de solo, apenas 13 parâmetros foram significativos, independentemente da variável de saída. Além disso, os resultados foram afetados pelo clima: em ambientes com boa distribuição pluviométrica o *plastochron* foi o principal parâmetro, enquanto em ambientes submetidos a maior estresse hídrico, o parâmetro *eff* foi o mais importante. Notou-se que qualquer parâmetro do solo foi indiferente para as condições irrigadas, enquanto que para as condições de sequeiro, a capacidade de campo e o ponto de murcha permanente foram relevantes em ambientes com baixa distribuição de chuvas e solos rasos. Locais chuvosos com solos profundos também não apresentaram sensibilidade aos parâmetros do solo.

Palavras-chave: Cana-de-açúcar, eFAST, PRCC, Sensibilidade, Incerteza

ABSTRACT

Sensitivity and uncertainty analysis of SAMUCA crop model across contrasting environments in Brazil

The Brazilian sugarcane industry is constantly developing, testing, and launching new cultivars and management practices to increase productivity. Due to climate change issues and limitations for expanding areas, farms are constantly pressured to increase agricultural efficiency. The use of process-based models (PBCM) to test cultivars and management options in different production environments is a reality and has been increasingly used. PBCMs are the state of the art in agricultural modeling and are increasingly complex, requiring several parameters to describe cultivation processes and boundary conditions. In general, PBCMs use the deterministic approach to simplify the uncertainty present in the environment using a single set of parameters. In practice, this uncertainty is seen in the variability of data collected in a field experiment, which is commonly represented by dispersion statistics such as standard deviation and variance. One way to explore this uncertainty is to use the stochastic approach, inserting a range of variability in the parameters and inputs of the simulation. This study aimed to use the stochastic approach to explore uncertainty and determine which parameters of the SAMUCA model are most influential in the simulation process. For this, the recent version of the SAMUCA model was used, inserting three uncertainty scenarios: uncertainty analysis only for genotype parameters (UG), uncertainty analysis only for soil parameters (US), and analysis of soil parameters and genotype (UGS). In this first stage, these three scenarios were simulated for a 4-year field experiment, with the crop cultivated under the effect of green cane trash blanket (GCTB) and bare soil (Bare). The variability of the stochastic simulation was quantified by the ratio between the mean standard deviation of the simulations and the mean standard deviation of the observed data. Subsequently, to better understand which factors caused greater uncertainty in the simulation process, a global sensitivity analysis (GSA) was performed using the extended Fourier Amplitude Sensitivity Test (eFAST) method for the same 4-year experiment, in order to identify which parameters were responsible for explaining the higher variance of the model and verifying the impact of the range of the chosen parameters, as well as the number of simulations necessary to have a reliable GSA. Finally, knowing that the environment can influence the GSA result, a new sensitivity analysis was carried out with two methods, eFAST and Partial Rank Correlation Coefficient (PRCC) for the main sugarcane-producing regions in Brazil, considering irrigated and rainfed conditions. The results indicated that the observed variability in the field is not fully explained by soil parameters, possibly due to irrigation and good rainfall distribution in the experimental area. The UG and the UGS had the same ability to quantify the variability present in the experimental field. In that case, sensitivity to soil parameters could simply be ignored and genotype parameters could be chosen as the sole source of variability for practical applications. Most of the uncertainty in this experiment is attributed to the *plastochron* parameter, however, it was identified that the parameter range set could influence the order of the most important parameters. This was observed when the analysis was carried out for two sets of different parameter intervals (the first set used maximum and minimum values reported in the literature; the second set applied a 25% perturbation to the previously calibrated values). Finally, out of 31 parameters, 24 genotype and 7 soil, only 13 parameters were significant, regardless of the output variable. In addition, the results were affected by climate: in environments with good rainfall distribution, *plastochron* was the main parameter, while in environments subjected to greater water stress, the *eff* parameter was the most important. It was noted that any soil parameter was indifferent to irrigated conditions. In contrast, for rainfed conditions, field capacity and permanent wilting point were relevant in environments with low rainfall distribution and shallow soils. Rainy sites with deep soils also showed no sensitivity to soil parameters.

Keywords: Sugarcane, eFAST, PRCC, Sensitivity, Uncertainty

1. INTRODUCTION

The use of modeling as a decision-making tool is a common practice in several areas of science. In agriculture, process-based crop models (PBCM) represent the state-of-art in this area of science (Jones et al., 2017). When properly calibrated, they are commonly used to simulate growth and development of crops in certain regions and test scenarios of management and adaptation strategies (Faivre et al., 2009). Scientist and decision makers have used crop modeling as a tool to address issues related to the climate change (Jones et al., 2015; Singels et al., 2013), plant breeding (Hoffman et al., 2018), risk analysis (Everingham et al., 2002) and yield forecasting (Everingham et al., 2016).

Most of the findings were achieved by using the deterministic approach, which meant that they considered a "best set" of parameters to characterize the simulated system and providing only one simulation path for the entire environment. This criterion implicitly means that such best value represents the state of the crop in the studied area, and that there are no sampling errors associated with the plant, microclimate, or soil variability (Petersen, 1994). However, agricultural experimental data usually shows great dispersion (variance and deviation) caused by the environment and management (Brogi et al., 2020; Usowicz and Lipiec, 2017; van Bussel et al., 2016). This dispersion in the measured data is common in a biological system, where the reality of processes that occur in nature are not deterministic, but rather stochastic (Wilkinson, 2006), as it considers situations influenced by random effects to be a stochastic process (e.g. light scattering). In this way, a stochastic process can show different possible pathways that a PBCM can take from varying a range of parameters (Wallach et al., 2018). This observed dispersion can be seen as uncertainty in the data collected and quantified in the PBCM simulation by the range of variation in the model's input parameters (He et al., 2009; Li et al., 2018).

Sugarcane is a crucial crop for world bioenergy and for food supply (Raza et al., 2019), and several authors have studied sugarcane crop modeling (Inman-Bamber and Smith, 2005; Jones and Singels, 2018; Keating et al., 1999; Marin and Jones, 2014; Singels and Bezuidenhout, 2002; Thorburn et al., 2005; Valade et al., 2014; Vianna et al., 2020). Singels (2013) presented a detailed review of the main sugarcane models in the literature, highlighting sugarcane as one of the crops with highest need to be represented in PBCM given its specific farming systems and logistic requirements. To represent its physiological complexity, sugarcane PBCM have many genotype parameters compared to other crops, such as maize and wheat. For instance, the sugarcane models DSSAT/CASUPRO and DSSAT/CANEGRO have 33 and 18 genotype cultivar parameters respectively, while DSSAT/CERES-MAIZE and DSSAT/CERES-WHEAT have only respectively 6 and 7 cultivar parameters to be calibrated.

According to Sinclair & Seligman (1996), the development of different PBCM by other research groups allows to improve the understanding of crop processes. Marin & Jones (2014) developed the SAMUCA focusing on the specific characteristics of sugarcane farming systems in Brazil. Recently, the SAMUCA model was improved by reducing the uncertainties around the soil water balance, heat flux and physiological mechanisms such as carbon partition, photosynthesis, tillering and root growth (Vianna et al., 2020). As any other PBCM, SAMUCA represents a simplification of the real system and requires several parameters whose determination is a problem for practical operational applications (Makowski et al., 2002). Most parameters are acquired through field observations, which are expensive and time-consuming, and the acquisition of certain parameters is difficult. Yet, many parameters vary depending on environmental conditions, cultivars, seasonal variation, among other factors (Wang et al., 2013).

In practice, it is well known that only part of the parameters is usually responsible for most of the model uncertainty, while most of them have only minor influence (LI et al., 2019; Varella et al., 2010; Zhang et al., 2020).

The parameter sensitivity analysis (SA) method can identify the most important parameters for a given model output variable, which allows users to focus on the most important model parameters during the calibration process. Furthermore, based on the SA, the balance and robustness of the model can be analyzed for future improvement, model development, and applications (Chu-Agor et al., 2011; Confalonieri, 2010; Fraedrich and Goldberg, 2000; Hirabayashi et al., 2011).

The SA can be divided into two groups: the local sensitivity analysis (LSA) and the global sensitivity analysis (GSA). The LSA consists of changing a single parameter at a time, while the other parameters are kept at their reference values; in other words, this method is based on the local derivatives of the model's output concerning the variation of a single parameter, which indicates how strong is the output changes around the reference parameter values (Saltelli et al., 1999). The GSA allows you to evaluate the entire uncertainty range of parameters, considering changes in all parameters along with their range, as well as the interactions among parameters (Saltelli et al., 1999).

The GSA has several aspects that can affect sensitivity indices and their uncertainty, regardless of the method adopted. In general, the most important uncertainty sources of GSA are: (i) sample size, (ii) range of parameters, and (iii) complexity of the model (Gan et al., 2014; Song et al., 2015; Xu and Gertner, 2011). To our knowledge, there are no studies in the literature investigating effects in GSA caused by sample size and parameter range on extended Fourier Amplitude Sensitivity test (eFAST) method in sugarcane models. In the case of sample size, the available studies used eFAST and were based on evidence provided by Wang et al. (2013), which has been replicated for different crops. In those studies that were not based on Wang et al. (2013), it was adopted a very large sample size without a clear definition criterion (Tan et al., 2016). However, by adopting the sample size suggested by Wang et al. (2013), the model characteristics are ignored, and when one uses a very large sample size, an unnecessary large computational time is required for the analysis.

The SAMUCA crop model has been relatively undervalued compared to well-established models such as DSSAT/CANEGRO and APSIM-Sugar (Marin et al., 2015, 2014; Sexton et al., 2017; Thorburn et al., 2005), and some issues remained unclear for well understanding the model complexity. Studies have shown that a GSA can be dependent on the method used (Marino et al., 2008), the parameter range (Wang et al., 2013), the time series for which the output variable is analyzed (Xing et al., 2017), climate (Sexton et al., 2017), soil (Varella et al., 2010) and management (Zhang et al., 2020; Zhao et al., 2014). In Brazil there are different climates and soils in which sugarcane is produced under rainfed and irrigated conditions, and it is desired to understanding how the model would perform in terms of sensitivity in each condition. As far as we know, there are no studies evaluating the impact of soil and genotype parameters in different environments and what their real impacts are on the main output variables of SAMUCA crop model.

Based on the exposed rationale, this study tested the following hypotheses: the stochastic approach can represent the variability existing in an experimental field; (ii) the methodology applied in a global sensitivity analysis has a direct influence on the results; (iii) for different boundary conditions, such as soil, climate, and management, the model may present different parameters as the most influential for a given variable

1.1. Objectives

Based on the hypotheses presented above, the overall objective of this study was to test the capacity of PBCM to represent the variability in field experiments in different conditions of cultivation and identify the main parameters responsible for explaining the greater uncertainty of the model in different environments of sugarcane production in Brazil.

1.1.1. Specific objectives

This objective, in turn, can be subdivided into specific objectives, as follows:

(i) explore the uncertainty in the soil-hydraulic and textural parameters (US), genotype (UG) and both together (UGS); (iii) to model the variability present in the field considering the presence or absence of the green trash blanket (GCTB).

(ii) determine the optimal sample size for the eFAST method

(iii) Investigate whether there is a difference between the ranges of parameters used in GSA

(iv) Identify which parameters are responsible for the greatest uncertainty in the SAMUCA model.

(v) Determine which are the most important soil and genotype parameters using a robust set of experiments conducted across different producing environments in Brazil for irrigated and rainfed environments

(vi) Evaluate the importance of the soil parameters across different producing environments in Brazil.

References

Brogi, C., Huisman, J.A., Herbst, M., Weihermüller, L., Klosterhalfen, A., Montzka, C., Reichenau, T.G., Vereecken, H., 2020. Simulation of spatial variability in crop leaf area index and yield using agroecosystem modeling and geophysics-based quantitative soil information. Vadose Zo. J. 19. https://doi.org/10.1002/vzj2.20009

Chu-Agor, M.L., Muñoz-Carpena, R., Kiker, G., Emanuelsson, A., Linkov, I., 2011. Exploring vulnerability of coastal habitats to sea level rise through global sensitivity and uncertainty analyses. Environ. Model. Softw. 26, 593–604. https://doi.org/10.1016/j.envsoft.2010.12.003

Confalonieri, R., Bregaglio, S., Acutis, M., 2010. A proposal of an indicator for quantifying model robustness based on the relationship between variability of errors and of explored conditions. Ecol. Modell. 221, 960–964. https://doi.org/10.1016/j.ecolmodel.2009.12.003

Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agron. Sustain. Dev. 36. https://doi.org/10.1007/s13593-016-0364-z

Everingham, Y.L., Muchow, R.C., Stone, R.C., Inman-Bamber, N.G., Singels, A., Bezuidenhout, C.N., 2002. Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts. Agric. Syst. 74, 459–477. https://doi.org/10.1016/S0308-521X(02)00050-1

Faivre, R., Leenhardt, D., Voltz, M., Benoît, M., Papy, F., Dedieu, G., Wallach, D., 2009. Spatialising Crop Models BT - Sustainable Agriculture, in: Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S., Alberola, C. (Eds.), . Springer Netherlands, Dordrecht, pp. 687–705. https://doi.org/10.1007/978-90-481-2666-8_42

Fraedrich, D., Goldberg, A., 2000. Methodological framework for the validation of predictive simulations. Eur. J. Oper. Res. 124, 55–62. https://doi.org/10.1016/S0377-2217(99)00117-4

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. Environ. Model. Softw. 51, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

He, J., M. D. Dukes, M.D., J. W. Jones, J.W., W. D. Graham, W.D., J. Judge, J., 2009. Applying GLUE for Estimating CERES-Maize Genetic and Soil Parameters for Sweet Corn Production. Trans. ASABE 52, 1907–1921. https://doi.org/10.13031/2013.29218

Hirabayashi, S., Kroll, C.N., Nowak, D.J., 2011. Component-based development and sensitivity analyses of an air pollutant dry deposition model. Environ. Model. Softw. 26, 804–816. https://doi.org/10.1016/j.envsoft.2010.11.007

Hoffman, N., Singels, A., Patton, A., Ramburan, S., 2018. Predicting genotypic differences in irrigated sugarcane yield using the Canegro model and independent trait parameter estimates. Eur. J. Agron. 96, 13–21. https://doi.org/10.1016/j.eja.2018.01.005

Inman-Bamber, N.G., Smith, D.M., 2005. Water relations in sugarcane and response to water deficits. F. Crop. Res. 92, 185–202. https://doi.org/10.1016/J.FCR.2005.01.023

Jin, X., Li, Z., Nie, C., Xu, X., Feng, H., Guo, W., Wang, J., 2018. Parameter sensitivity analysis of the AquaCrop model based on extended fourier amplitude sensitivity under different agro-meteorological conditions and application. F. Crop. Res. 226, 1–15. https://doi.org/10.1016/j.fcr.2018.07.002

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. https://doi.org/10.1016/J.AGSY.2016.05.014

Jones, M.R., Singels, A., 2018. Refining the Canegro model for improved simulation of climate change impacts on sugarcane. Eur. J. Agron. 100, 76–86. https://doi.org/10.1016/j.eja.2017.12.009

Jones, M.R., Singels, A., Ruane, A.C., 2015. Simulated impacts of climate change on water use and yield of irrigated sugarcane in South Africa. Agric. Syst. 139, 260–270. https://doi.org/https://doi.org/10.1016/j.agsy.2015.07.007

Keating, B.A., Robertson, M.J., Muchow, R.C., Huth, N.I., 1999. Modelling sugarcane production systems I. Development and performance of the sugarcane module. F. Crop. Res. 61, 253–271. https://doi.org/10.1016/S0378-4290(98)00167-1

Li, Z. hai, Jin, X. liang, Liu, H. long, Xu, X. gang, Wang, J. hua, 2019. Global sensitivity analysis of wheat grain yield and quality and the related process variables from the DSSAT-CERES model based on the extended Fourier Amplitude Sensitivity Test method. J. Integr. Agric. 18, 1547–1561. https://doi.org/10.1016/S2095-3119(18)62046-5

Li, Zhenhai, He, J., Xu, X., Jin, X., Huang, W., Clark, B., Yang, G., Li, Zhenhong, 2018. Estimating genetic parameters of DSSAT-CERES model with the GLUE method for winter wheat (Triticum aestivum L.) production. Comput. Electron. Agric. 154, 213–221. https://doi.org/10.1016/j.compag.2018.09.009

Makowski, D., Wallach, D., Tremblay, M., 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. Agronomie 22, 191–203. https://doi.org/10.1051/agro:2002007

Marin, F.R., Jones, J.W., 2014. Process-based simple model for simulating sugarcane growth and production. Sci. Agric. 71, 1–16. https://doi.org/10.1590/S0103-90162014000100001

Marin, F.R., Ribeiro, R. V., Marchiori, P.E.R., 2014. How can crop modeling and plant physiology help to understand the plant responses to climate change? A case study with sugarcane. Theor. Exp. Plant Physiol. 26, 49–63. https://doi.org/10.1007/S40626-014-0006-2/FIGURES/5

Marin, F.R., Thorburn, P.J., Nassif, D.S.P., Costa, L.G., 2015. Sugarcane model intercomparison: Structural differences and uncertainties under current and potential future climates. Environ. Model. Softw. 72, 372–386. https://doi.org/10.1016/j.envsoft.2015.02.019

Pereira, R.A. de A., Vianna, M. dos S., Nassif, D.S.P., Carvalho, K. dos S., Marin, F.R., 2021. Global sensitivity and uncertainty analysis of a sugarcane model considering the trash blanket effect [WWW Document]. Eur. J. Agron. https://doi.org/https://doi.org/10.1016/j.eja.2021.126371

Petersen, R.G., 1994. Agricultural Field Experiments: Design and Analysis.

Raza, G., Ali, K., Hassan, M.A., Ashraf, M., Khan, M.T., Khan, I.A., 2019. Sugarcane as a Bioenergy Source. Sugarcane Biofuels 3–19. https://doi.org/10.1007/978-3-030-18597-8_1

Saltelli, A., Tarantola, S., Chan, P.-S., 1999. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. Technometrics 41, 39–56. https://doi.org/10.1080/00401706.1999.10485594

Sexton, J., Everingham, Y.L., Inman-Bamber, G., 2017. A global sensitivity analysis of cultivar trait parameters in a sugarcane growth model for contrasting production environments in Queensland, Australia. Eur. J. Agron. 88, 96–105. https://doi.org/10.1016/j.eja.2015.11.009

Silvestro, P.C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., Casa, R., 2017. Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. PLoS One 12. https://doi.org/10.1371/JOURNAL.PONE.0187485

Sinclair, T.R., Seligman, N.G., 1996. Crop Modeling: From Infancy to Maturity. Agron. J. 88, 698–704. https://doi.org/10.2134/agronj1996.00021962008800050004x

Singels, A., 2013. Crop models. Sugarcane Physiol. Biochem. Funct. Biol. 541–577. https://doi.org/https://doi.org/10.1002/9781118771280.ch20

Singels, A., Jones, M., Marin, F., Ruane, A., Thorburn, P., 2013. Predicting Climate Change Impacts on Sugarcane Production at Sites in Australia, Brazil and South Africa Using the Canegro Model. Sugar Tech 16, 347–355. https://doi.org/10.1007/s12355-013-0274-1

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. J. Hydrol. https://doi.org/10.1016/j.jhydrol.2015.02.013

Thorburn, P.J., Meier, E.A., Probert, M.E., 2005. Modelling nitrogen dynamics in sugarcane systems: Recent advances and applications, in: Field Crops Research. Elsevier, pp. 337–351. https://doi.org/10.1016/j.fcr.2005.01.016

Usowicz, B., Lipiec, J., 2017. Spatial variability of soil properties and cereal yield in a cultivated field on sandy soil. Soil Tillage Res. 174, 241–250. https://doi.org/10.1016/j.still.2017.07.015

Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., Martiné, J.-F., 2014. Modeling sugarcane yield with a process-based model from site to continental scale: uncertainties arising from model structure and parameter values. Geosci. Model Dev 7, 1225–1245. https://doi.org/10.5194/gmd-7-1225-2014

van Bussel, L.G.J., Ewert, F., Zhao, G., Hoffmann, H., Enders, A., Wallach, D., Asseng, S., Baigorria, G.A., Basso, B., Biernath, C., Cammarano, D., Chryssanthacopoulos, J., Constantin, J., Elliott, J., Glotter, M., Heinlein, F., Kersebaum, K.C., Klein, C., Nendel, C., Priesack, E., Raynal, H., Romero, C.C., Rötter, R.P., Specka, X., Tao, F., 2016. Spatial sampling of weather data for regional crop yield simulations. Agric. For. Meteorol. 220, 101–115. https://doi.org/10.1016/j.agrformet.2016.01.014

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environ. Model. Softw. 25, 310–319. https://doi.org/https://doi.org/10.1016/j.envsoft.2009.09.012

Vianna, M. dos S., Nassif, D.S.P., dos Santos Carvalho, K., Marin, F.R., 2020. Modelling the trash blanket effect on sugarcane growth and water use. Comput. Electron. Agric. 172, 105361. https://doi.org/10.1016/j.compag.2020.105361

Wallach, D., Thorburn, P., Asseng, S., Challinor, A.J., Ewert, F., Jones, J.W., Rotter, R., Ruane, A., 2016. Estimating model prediction error: Should you treat predictions as fixed or random? Environ. Model. Softw. 84, 529–539. https://doi.org/10.1016/j.envsoft.2016.07.010

Wang, J., Li, X., Lu, L., Fang, F., 2013. Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. Environ. Model. Softw. 48, 171–182. https://doi.org/10.1016/j.envsoft.2013.06.007

Wilkinson, D.J., 2006. Stochastic Modelling for Systems Biology. Chapman and Hall/CRC. https://doi.org/10.1201/9781420010664

Xing, H. min, Xu, X. gang, Li, Z. hai, Chen, Y. jin, Feng, H. kuan, Yang, G. jun, Chen, Z. xia, 2017. Global sensitivity analysis of the AquaCrop model for winter wheat under different water treatments based on the extended Fourier amplitude sensitivity test. J. Integr. Agric. 16, 2444–2458. https://doi.org/10.1016/S2095-3119(16)61626-X

Xu, C., Gertner, G., 2011. Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). Comput. Stat. Data Anal. 55, 184–198. https://doi.org/10.1016/j.csda.2010.06.028

Zhang, Y., Arabi, M., Paustian, K., 2020. Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. Environ. Model. Softw. 126. https://doi.org/10.1016/j.envsoft.2020.104642

Zhao, G., Bryan, B.A., Song, X., 2014. Sensitivity and uncertainty analysis of the APSIM-wheat model: Interactions between cultivar, environmental, and management parameters. Ecol. Modell. 279, 1–11. https://doi.org/10.1016/J.ECOLMODEL.2014.02.003

2. GLOBAL SENSITIVITY AND UNCERTAINTY ANALYSIS OF A SUGARCANE MODEL CONSIDERING THE TRASH BLANKET EFFECT

Abstract

The deterministic approach in crop modeling simplifies uncertainty present in the environment using a unique parameter set. In practice, this uncertainty is seen in the variability of data collected in a field experiment. One way to exploit this uncertainty is to use the stochastic approach, by inserting the range of plausible variability into the model parameters and inputs. This study aims to evaluate the ability of a process-based crop model to simulate the uncertainty of a sugarcane field. We employed the recently updated version of SAMUCA model to simulate the sugarcane growth and development in a 4-year field experiment, where the crop was grown under the effect of green cane trash blanket (GCTB) and bare soil (Bare). To analyze the effect of genotype and soil variability on output variables, a stochastic approach was applied to the corresponding parameters of the SAMUCA model. A global sensitivity analysis was utilized to prioritize and identify the most important parameters to explain the model uncertainty. Then, the uncertainty was analyzed in three different ways: uncertainty analysis only for genotype parameters (UG), uncertainty analysis only for soil parameters (US) and the analysis of both soil and genotype parameters (UGS). We quantified the variability of the stochastic simulation by the ratio between the average of the standard deviation of the simulations and the average of the standard deviation of the observed data. The variability observed in the field is not fully explained by the hydraulic parameters of the soil, possibly due to irrigation and good rainfall distribution in the area. Furthermore, the variability in US simulations were higher for GCTB than in Bare treatment, suggesting that the GCTB has a larger influence in SAMUCA's variability than for the hydraulic parameters in the conditions of this study. The UG and UGS had the same capacity to quantify the variability present in the environment for the treatments Bare and GCTB.Therefore, sensitivity to soil parameters can simply be ignored and genotype parameters can be chosen as the only source of variability for practical applications. Our suggestion for future work is to explore environments without irrigation, different amounts of GCTB and other soil parameters present in the model.

Keyword: Correlated parameters, GLUE, PRCC, stochastic

2.1. Introduction

The use of modeling as a decision-making tool is a common practice in several areas of science. In agriculture, process-based crop models (PBCM) represent the state-of-art in this area of science (Jones et al., 2017). When properly calibrated, they are commonly used to simulate the growth and development of crops in certain regions and test "what if " scenarios of managements and adaptation strategies (Faivre et al., 2009). Scientist and decision makers have used crop modeling as a tool to address issues related to the sugar and bioenergy sectors, including climate change (Jones et al., 2015; Singels et al., 2013), plant breeding (Hoffman et al., 2018), risk analysis (Everingham et al., 2002) and yield forecasting (Everingham et al., 2016).

Most of the findings were achieved by using the deterministic approach, which meant that they considered a "best set" of parameters to characterize the simulated system and providing only one simulation path for the entire environment. This criterion implicitly means that such best value represents the state of the crop in the studied area, and that there are no sampling errors associated with the plant, microclimate, or soil variability (Petersen, 1994). However, agricultural experimental data usually shows great dispersion (variance and deviation) caused by the environment, management and by measurement errors (Brogi et al., 2020; Usowicz and Lipiec, 2017; van Bussel et al., 2016). This dispersion in the measured data is common in a biological system, where the reality of processes that occur in nature are not deterministic, but rather stochastic (Wilkison, 2006), as it considers situations influenced by random effects to be a stochastic process. In this way, a stochastic process can show the different possible pathways that a PBCM can take from varying a range of parameters (Wallach et al., 2018). This observed

dispersion can be seen as uncertainty in the data collected and quantified in the PBCM simulation by the range of variation in the model's input parameters (He et al., 2009; Li et al., 2018).

Four different approaches can be used to estimate uncertainty in PBCM simulations: (i) comparison of hindcasts with observations; (ii) multi-model ensemble studies; (iii) propagating input and/or parameter uncertainty through the model; (iv) using simulations with multiple model structures, multiple input and multiple parameter vectors for each model (Wallach et al., 2016). The first two approaches provide a unique answer or explore the uncertainty present in the structure of each PBCM. However, for daily-practical problems, we are often not interested in a model with average parameters, as a simulated area may have different genotypes and variability associated with soil properties and microclimate conditions (Wallach et al., 2016; Wallach and Thorburn, 2017).

One of the challenges in crop modeling stochastic simulation is to accurately choose parameters distributions respecting the correlation between them, which is often neglected (Jones et al., 2011). To preserve the correlation between the parameters, a normal multivariate distribution must be generated (He et al., 2009), and the Generalized Likelihood Uncertainty Estimator (GLUE) combined with the Cholesky decomposition of the variance-covariance matrix is a robust option for generating a set of correlated parameters (Baigorria and Jones, 2010; Marin et al., 2017). Yet, the sensitivity of the parameters is relevant when using the stochastic approach in PBCM, as it can aid on selecting the set of parameters with largest influence in the targeted process or output (Wallach et al., 2018; Zhang et al., 2020).

In a previous attempt to include uncertainty in the sugarcane model Marin et al., (2017) used a previous version of the SAMUCA model (Marin and Jones, 2014) under a stochastic approach. In that study, the uncertainty of 13 genotype parameters was considered, considering their correlation with parameters of two genotypes grown in several environments of Brazil. However, those authors listed some important limitations in that study: (i) the structural uncertainty of the model, (ii) uncertainty in the experimental data, (iii) uncertainty present in the environment, mainly in relation to the soil parameters. Yet, Marin et al., (2017) only used data from plant cane, they neglected the sensitivity of model parameters, and they did not evaluate the model simulation ability to capture the effect of green cane trash blanket (GCTB) on the growth and development of sugarcane, an important component of Brazilian sugarcane cropping systems.

Being the sugarcane the main source of sugar and the second largest feedstock for bioenergy in the world (Goldemberg et al., 2008; Jaiswal et al., 2017; Marin et al., 2019) and to overcome the model and experimental limitations reported in Marin et al. (2017), we use a detailed 4-year experiment to evaluate model uncertainty under a stochastic approach together with a new version of SAMUCA (Vianna et al., 2020), which would allowed us to evaluate aspects related to soil variability, different crop stages (plant cane and ratoons) and the effect of GCTB on the crop growth and development. Thus, in this paper we aimed to evaluate a sugarcane crop model used under a stochastic approach to represent the existing variability in an experimental plot. Our specific objectives were: (i) to perform a global sensitivity analysis of genotype parameters to determine which are significant and use them in the stochastic simulation with correlated parameters; (ii) to explore the uncertainty in the soil-hydraulic and textural parameters (US), genotype (UG) and both of them together (UGS); (iii) to model the variability present in the field considering the presence or absence of the GCTB.

2.2. Material and Methods

2.2.1. Brief history of the SAMUCA model

The SAMUCA model was created due to the argument of Sinclair and Seligman (1996), where they highlight the importance of developing the proper models for knowledge groups, allowing to deepen the mechanisms involved in the simulation process and the uncertainties inherent to the used models. In addition, the SAMUCA model also had the objective of exploring the uncertainty in genotype parameters, incorporating a calibration procedure based on the Generalized Likelihood Uncertainty Estimator (GLUE) to ensure a consistent and reliable adaptation of the model for applications in Brazil (Marin and Jones, 2014). The SAMUCA model was built with a database of different locations in Brazil, comprising of different climates, soils and managements which is also used to evaluate other widely used sugarcane dynamic models (Marin et al., 2015). Even with good results to simulate the growth and development of sugarcane, it was a first version with several limitations. Such limitations were primarily related to the oversimplified soil water balance and the non-inclusion of GCTB effect into the model routines as it is extremely important to represent the Brazilian sugarcane cropping systems.

Because of this, a new version of SAMUCA model was built by Vianna et al. (2020) to reduce the uncertainties around model structure, soil moisture and heat flow in comparison with its previous version. Soil moisture is simulated by the widely tested "tipping bucket" method, whereas heat flow is solved numerically according to Kroes et al. (2009). Both processes can also be simulated under the effect of GCTB, which has recently emerged as an important operational practice for Brazilian farmers (Carvalho et al., 2017). Further improvements were also made to the subroutines dedicated to the simulation of carbon partitioning at phytomer level, layered-canopy photosynthesis, tillering and root growth (Bezuidenhout et al., 2003; Laclau and Laclau, 2009; O'Leary, 2000). This new version of SAMUCA model is also included in the DSSAT platform v4.8.

2.2.2. Field Experiment

We conducted a field experiment of approximately 2.5 hectares of sugarcane at the experimental fields of the College of Agriculture "Luiz de Queiroz", Piracicaba, São Paulo (Lat: 22º41'55"S, Lon: 47º38'34"W, Alt: 540 m). The sugarcane cultivar was RB86-7515, a widely used genotype in Brazil (ca. 30% of Brazil's planted area). It was planted on October 16, 2012, with a row spacing of 1.4 m and depth of 0.2 m. A bare soil treatment (Bare) was conducted during the four sequential years, whereas the GCTB treatment started in the first ratoon (Oct-2013) and was carried out for 3 years. Agricultural practices were adopted to represent high yield farming systems and to ensure the crop was free from pests, diseases, and nutritional stress. The climate is characterized by hot and humid summer with dry winter (Cwa – Köppen classification), and the soil classified as Typic Hapludox. The experiment was irrigated by a center-pivot, based on monitoring the soil moisture by Frequency Domain Reflectometry (FDR) and the evapotranspiration by Bowen Ratio Method (BRM) in both treatments (Nassif et al., 2014)

Table.1 Description of seasons, planting and harvesting dates, duration in days, treatments, and measurements variables of the field experiment in Piracicaba, Brazil.

Season	Planting	Haverst	Duration	Variables	Treatments
Plant Cane	10/16/2012	10/15/2013	364	SDM,SFM,TIL,LAI *and* POL	Bare
1st Ratoon	10/15/2013	07/15/2014	273	SDM,SFM,TIL,LAI *and* POL	Bare and GCTB
2nd Ratoon	07/15/2014	06/08/2015	328	SDM,SFM,TIL,LAI *and* POL	Bare and GCTB
3rd Ratoon	06/08/2015	06/08/2016	365	SDM,SFM,TIL	Bare and GCTB

Green cane trash blanket (GCTB), stalk dry mass (SDM) and stalk fresh mass (SFM) , leaf area index (LAI), sucrose concentration in fresh matter (POL) and tillering (TIL).

Crop growth was monitored by regular destructive sampling of biomass (stalk fresh and dry mass; SFM and SDM) throughout the sugarcane growing cycles. A total of 30 plants per treatment were collected every month at random locations and immediately transported to weigh fresh biomass. Biomass was then dried at 60ºC in an air circulation oven (TE-394/5-MP, Tecnal®, Piracicaba, São Paulo, Brazil) for four days before weighing as dry biomass parts with a precision balance (Toledo, model 2098). Crop development was monitored with non-destructive sampling in four sub-plots of 35 m² randomly positioned at the beginning of each season (total of 8 plots). The tiller population (TIL) was regularly counted in the non-destructive plots and scaled to 1.0 m². The Leaf Area Index (LAI) was regularly measured with a plant canopy analyser (LAI-2000, LI-COR, Inc, Lincoln, Nebraska, USA) with eight repetitions for each treatment. During crop maturation, fifteen culms per treatment were randomly cut and immediately transported for milling to determine the fraction of fiber and sugars using a digital saccharimeter (SDA5900, Acatec, São Paulo, São Paulo Brazil) and precision balance (Prix 110, Mettler Toledo, Mississauga, Ontario, Canada), this allowed to determine the sucrose concentration in fresh matter (POL.

2.2.3. Genotype parameters and global sensitivity analysis

The choice of genotype parameters for uncertainty analysis was based on a global sensitivity analysis (GSA) using the partial rank correlation coefficient (PRCC) method (Wallach et al., 2019). We employed this method as the arbitrary selection of parameters could not generate variations in the output of the model that would explain the variability in the real environment (Varella et al., 2010). The method consists of massive sampling of parameters using the Monte Carlo method, to assess the correlation between each parameter and model output. To do so, we obtained the linear relationship between the genotype parameters and the model output with the PRCC method, where the positive PRCC values a direct linear relationship while the negative PRCC values an inverse linear relationship. The difference between the PRCC and its advantage over Person correlation coefficient and the partial correlation coefficient is that it can explore the non-linear relationships between inputs and outputs. The PRCC values range from -1 to 1, as does the Pearson correlation, taking a measure of the strength of a linear association between an input and an output. Mukaka (2012) presented different classes of interpretation for the PRCC correlation (Table 2). In the following analysis, we only considered the genotype parameters that were statistically significant at 1% for the output model components of sugarcane: SDM, SFM, TIL, LAI and POL.

Table 2. Rules for interpreting the size of a correlation coefficient (Mukaka, 2012)

Size of Correlation	Interpretation
0.90 to 1.00 (-0.90 to -1.00)	Very high positive (negative) correlation
0.70 to 0.90 (-0.70 to -0.90)	High positive (negative) correlation
0.50 to 0.70 (-0.50 to -0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (-0.30 to -0.50)	Low positive (negative) correlation
0.00 to 0.30 (-0.00 to -0.30)	Negligible correlation

2.2.4. Soil parameters

The hydraulic soil parameters (HSP) were obtained from samples taken at four random locations within the experimental area. At each location, three repetitions of undisturbed soil samples were taken at the depths of 5, 15, 30, 60 and 100 cm. The 60 undisturbed samples were used to obtain water retention curves (at the potentials of 10, 20, 60, 100, 330, 1,000, 3,000, and 15,000 kPa) for each depth, used to derive the permanent wilting point (*WPp*), field capacity (*FCp*), saturation point (*STp*) and saturated hydraulic conductivity (*Ksat*) required by the SAMUCA model. Thus, a retention curve was obtained for each depth and location, where maximum, minimum and average values of parameters were obtained for each depth (Table 3). We chose to work with the maximum and minimum values to generate a uniform distribution, regardless of the spatial position of the sample; that is, within the study area we considered that the soil parameters varied berween these maximum and minimum values.

The soil texture parameters (TSP) used were clay (*Pclay*), sand (*Psand*) and silt (*Psilt*) for the same depths as HSP. The TSP interval (Table 3) was obtained in the literature from two studies conducted in the same experimental area at different periods, at a depth of 60 cm and in this case, we considered for the depth of 100 cm the same interval measured at a depth of 60 cm.

To determine which parameters would be inserted in the uncertainty and stochastic simulations, we performed a GSA by applying the same method as the parameters of the genotype, considering the parameters presented in Table 3 and their respective depths. If at least one of depths was significant, we assumed the other depths would have the uncertainty inserted, maintaining the correlation between them.

Table 3. Average (Avg), maximum (Max) and minimum (Min) for: depth (DP), wilting point (WPp), field capacity (FCp), saturation point (STp), saturated hydraulic conductivity (Ksat), content clay (Pclay), content silt (Psilt), content sand (Psand)[y].

Hydraulic parameters												
DP	FCp ($cm^3.cm^{-3}$)			K_{sat} ($cm.h^{-1}$)			STp ($cm^3.cm^{-3}$)			WPp ($cm^3.cm^{-3}$)		
(cm)	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min
5	0.285	0.305	0.255	1.7	2.51	1.03	0.38	0.413	0.34	0.216	0.23	0.191
15	0.303	0.325	0.287	1.01	1.2	0.85	0.352	0.396	0.334	0.24	0.245	0.224
30	0.347	0.414	0.305	0.95	1.02	0.14	0.39	0.448	0.36	0.278	0.357	0.231
60	0.394	0.406	0.346	0.62	1.02	0.14	0.428	0.474	0.392	0.307	0.35	0.273
100	0.393	0.434	0.357	0.21	0.4	0.1	0.456	0.486	0.422	0.253	0.304	0.198
Texture parameters												
DP	Pclay ($g.g^{-1}$)			Psilt ($g.g^{-1}$)			Psand ($g.g^{-1}$)					
(cm)	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min			
5	0.544	0.624	0.464	0.234	0.296	0.172	0.222	0.240	0.204			
15	0.544	0.624	0.464	0.234	0.296	0.172	0.222	0.240	0.204			
30	0.596	0.694	0.498	0.215	0.292	0.138	0.189	0.210	0.167			
60	0.633	0.689	0.576	0.200	0.264	0.136	0.168	0.160	0.175			
100	0.633	0.689	0.576	0.200	0.264	0.136	0.168	0.160	0.175			

2.2.5. Generalized likelihood uncertainty estimation method

The generalized likelihood uncertainty estimation method (GLUE) was used to select the parameter set with the highest likelihood to reproduce the end-of-season observation; hereafter called the best parameter set. Yet, GLUE was used to create the variance-covariance correlation matrix of model parameters, which in turn was used for generate the correlated parameter sets for stochastic simulations. It is a parameter estimation method that deals with problems associated with parameter interactions and non-linearity in the model's response (Beven and Binley, 1992). Present in platforms such as DSSAT, it is widely used to estimate genotype parameters (He et al., 2010; Jones et al., 2011), especially those that cannot be measured directly in typical experiments; instead, they should be estimated based on data measured in experiments (Marin and Jones, 2014). The method is an approach based on the Monte Carlo application, which uses a set of parameters in massive simulation process to select a set of parameters in a uniform distribution within the sample space (Sreelash et al., 2012).

The GLUE procedure consists of five stages: (i) Develop prior parameter distributions, in this case, we assume uniform distributions from predefined range of variation for soil and genotype parameters (Marin et al., 2017); (ii) Generate random parameters sets from prior parameter distributions based on the Monte Carlo method, where the largest the number of simulations leads to more stable results. However, only a limited number of parameter sets had significant likelihood values that could be used to derive posterior distributions, even though 10,000 sets of parameters were generated in this study, considered a large sample (He et al., 2010); (iii) Run the model with the random parameters sets, where the model was run for each parameter set using developed R-scripts. The input files for the parameters were changed to simulate each random parameter set in sequence and for each parameter set the model outputs (SDM, SFM,TIL, LAI and POL) were tabulated for use in the GLUE likelihood calculations; (iv) Calculate the likelihood values to generated observations (O, three replicates each for each variables) were used along with the corresponding simulated outputs to compute the likelihood value, $L(\theta_i|O)$, for each of the N generated parameter vectors θ_i. Then, the probability p_i of each parameter set was computed with the following equation (1) and likelihood function was:

$$L(\theta_i|O) = \prod_{j=1}^{M} \frac{1}{\sqrt{2.\pi.\sigma_o^2}}.exp\left(-\frac{(O_j - f(\theta_i))^2}{2\sigma_o^2}\right) \quad (1)$$

$$p(\theta_i) = \frac{L(\theta_i|O)}{\sum_j^N L(\theta_i|O)} \quad (2)$$

$$\mathrm{L_{comb}}[\theta_i] = \prod_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{L_k}(\theta_\mathrm{i}|\mathrm{O_k}) \quad (3)$$

where $p(\theta_i)$ is probability or likelihood weight of the *i*th parameter set θ_i; $L(\theta_i|O)$ is the likelihood value of parameter set θ_i; given observations O_j the *j*-th observation of O. The *M* is the number of observation replicates; $f(\theta_i)$ is the model output referring to θ_i; *K* is the number of observation types; $\mathrm{L_{comb}}[\theta_i]$ is the combined likelihood value of *i*th parameter set θ_i; σ_o^2 the variance model errors, assumed to be the variances of observations for this study.

(v) Construct posterior distribution and statistics. The pairs of parameter sets and probabilities, (θ_i, p_i), i=1 , …N, were used to construct empirical posterior distributions and to compute the means and variance of selected parameters using the following equations:

$$\hat{\mu} = \sum_{i=1}^{N} p(\theta_i).\theta_i \quad (4)$$

$$\hat{\sigma}^2 = \sum_{i=1}^{N} p(\theta_i).(\theta_i - \hat{\mu})^2 \quad (5)$$

where $\hat{\mu}$, $\hat{\sigma}^2$ they are the mean and variance of the posterior distribution of the set parameters; N (10,000) is the number of random parameters set.

To apply and evaluate the performance of the GLUE method, this study used the following measured data: dry (SDM) and fresh stalk mass (SFM), leaf area index (LAI), tillers population (TIL) and sucrose concentration on fresh sugarcane basis (POL). For the GLUE method, only the measured data of SDM, SFM, and TIL were used to estimate the optimal parameters (genotype and soil), since these were the only variables sampled continuously over the four years of the experiment. To evaluate the model performance and the stochastic simulations, we used SDM, SFM, TIL, LAI, and POL.

2.2.6. Simulation of correlated parameters

We applied the Toeplitz-Cholesky decomposition (Baigorria, 2014) from a correlation matrix obtained from the 10,000 sets of parameters generated by the GLUE method. From this correlation matrix, we then generate a new set of parameters, comprised of 10,000 combinations. This new set was used to run the stochastic simulations, respecting the correlation among parameters.

$$\mathbf{R} = \begin{Bmatrix} r_1^{\psi} = r_1 C_{1,1} + \cdots + r_n C_{1,n} \\ \vdots \\ r_n^{\psi} = r_n C_{n,1} + \cdots + r_n C_{n,n} \end{Bmatrix} \quad (6)$$

The **R** matrix is multiplied by a square matrix containing the weighting values C_{ij} (i is the parameter and jth simulation), which are calculated based on the pairwise correlation values that form the correlation matrix **P** (Baigorria and Jones, 2010). As mentioned, the factorization matrix used here was the Toeplitz-Cholesky factorization matrix **C**:

$$\mathbf{C} = \mathbf{U}\mathrm{diag}(\mathbf{U}^{\frac{1}{2}}) \quad (7)$$

where **U** is an upper triangular matrix with positive diagonal entries generated from a special case of the symmetric **LU** decomposition of the correlation matrix, with $\mathbf{L}=\mathbf{U}^{\mathbf{T}}$.

2.2.7. Parameter set analysis and model evaluation.

From the 10,000 simulations with the correlated parameters sets we extracted the standard deviation of the simulations outputs to evaluate the model performance in replicating the variability observed in the field. This analysis was divided into three different sets to isolate the uncertainty of: (i) genotype parameters (UG); (ii) soil parameters (US); and (iii) both genotype and soil parameters (UGS). This means that only genotype parameters were considered for the GLUE method in the UG analysis, only soil parameters in the US, and both sets of parameters were considered for GLUE processing for the UGS analysis. When the GLUE method is not used for uncertainty analysis (e.g. soil parameters in UG), we assume the genotype and soil parameter values as reported by Vianna et al. (2020) (Table 4 and Table 3).

To evaluate the model performance in replicating the average condition of the experiment and its uncertainty, the statistical analysis was done in two different ways. Firstly, we compared the best set parameters obtained by GLUE with the average of the observed data. In this way, it was used the statistical indices root mean squared error (RMSE), determination index (R2), Nash-Sutcliff modeling efficiency (EF) (Nashand Sutcliffe, 1970) bias index (Bias) and Wilmont accuracy index (d) (Willmott et al., 2012). Secondly, we compared the variability observed in the field experiment, by using the standard deviation $\sigma(_{\sigma obs})$, with those from stochastic simulations using the standard deviation of the stochastic simulations $\sigma(_{\sigma sim})$. We also calculated the ratio (ξ in %) between σ_{sim} and σ_{obs} in order to verify the model skill in representing the observed variability.

Table 4. Cultivar-specific parameters, descriptions, units, and range used for uniform distribution sampling and standard values assumed for initial simulations. In bold are the parameters used in GLUE.

Parameter	Description	Min	θ	Max	Reference
amax	Assimilation rate at light saturation point *(μmol.m^{-2}.s^{-1})*	41.3	44.9	60.7	Sage et al., (2013)
chudec	Heat units for start of tiller abortion *(°C.d)*	1200	1600	1800	Liu et al., (1998)
chumat	Heat units for population establishment *(ºC.d)*	1500	1600	2850	Zhou and Shoko, (2011)/Marin and Jones, (2014)
chupeak	Heat units for population peak *(°C.d)*	400	1400	1950	Coelho et al., (2020); Marin et al., (2017) Nassif et
chustk	Heat units for start culm elongation *(°C.d)*	404	650	1050	Marin et al., (2017); /Singels and Bezuidenhout,
eff	Carboxylation efficiency *(μmol.m^{-2}.s^{-1} / μmol.m^{-2}.s^{-1})*	0.040	0.069	0.080	Sage et al., (2013)
end_tt_it_gro	Thermal time for completion of internode growth *(°C.d)*	600	1200	1400	Lingle, (1999)
end_tt_lf_gro	Thermal time for completion of leaf growth *(°C.d)*	1100	1300	1500	Smit and Singels, (2006)
init_lf_area	Initial leaf area of first appeared leaf *(cm²)*	15	10	30	Zhou et al., (2003)
max_ini_la	Initial leaf area of leaves appeared after top parts formation *(cm²)*	80	120	180	Zhou et al., (2003)
max_it_dw	Maximum dry biomass of internodes *(g)*	18	28	35	Lingle, (1999)
maxdgl	Maximum number of developed green leaf a tiller can hold *(#/tiller)*	6	6	12	Vianna et al., (2020)
maxgl	Maximum number of green leaf a tiller can hold (#/tiller)	10	12.0	12	Marin et al., (2015)
mid_tt_it_gro	Thermal time where internodes can achieve half of its maximum biomass	380	400	600	Lingle, (1999)
mid_tt_lf_gro	Thermal time where leaves can achieve half of its maximum biomass	400	700	800	Smit and Singels, (2006)
mla	**Maximum leaf area *(cm²)***	**450**	**600**	**800**	**Marin et al. (2014)**
n_lf_it_from	**Number of leaves appeared before internode formation *(#/tiller)***	**2**	**3**	**6**	**Vianna et al., (2020)**
n_lf_stk_em	**Number of leaves appeared before stalks emerges at soil surface *(#/tiller)***	**3**	**4**	**8**	**Vianna et al., (2020)**
phyllochron	Phyllochron interval for leaf appearance (*°C.d)*	107	132	169	Marin et al., (2015)/Inman-Bamber, 1994
plastochron	**Thermal time required for the appearance of phytometer *(ºC.d)***	**107**	**132**	**169**	**Marin et al., (2015)/Inman-Bamber, 1994**
popmat	Number of tillers on maturation (*tiller/m²)*	8.0	9.5	12.0	Marin and Jones, (2014)
poppeak	Maximum number of tillers (*tiller/m²)*	17.0	22.0	30.0	Marin et al., (2015)
sla	Specific leaf area *(cm^2.g^{-1})*	100.0	120.0	121.0	Ehara et al., (1994)
tillochron	**Thermal time required for emergence of new tiller *(ºC.d)***	**48.1**	**69.0**	**134.8**	**Bezuidenhout, (2000); Zhou and Shoko, (2011)**

θ is the value calibrated by Vianna et al. (2020) for cultivar RB867515; Max and Min value are range used for random parameters uniform distribution.

2.3. Results

2.3.1. Global sensitivity analysis for soil and genotype parameter

The GSA was performed for 24 genotype parameters (Table 5), and among those only five were statistically significant and had a monotonic response to the model outputs: *n_lf_stk_emerg*, *n_lf_it_form*, *tillochron*, *mla*, *plastochron*. Thus, these five genotype parameters were used to perform the stochastic simulations with correlated parameters in the UG and UGS analysis. The GSA analysis was performed considering the two treatments used in the experiment (GCTB and Bare), and no difference was found in terms of the correlation among parameters. For GCTB treatment, all output variables have at least one significant (0.01) parameter, being: two parameters for SDM (*plastochron*; *n_lf_it_from*) and SFM (*plastochron*; *n_lf_it_from*) and only one for TIL (*tillochron*), POL (*n_lf_it_from*) and LAI (*mla*) (Table 5). For the Bare treatment, only POL variable does not have any significant parameter, while SDM (*plastochron*; *n_lf_it_from*; *n_lf_stk_eme*) and SFM (*plastochron*; *n_lf_it_from*; *n_lf_stk_eme*) showed three significant parameters, and LAI (*mla*) and TIL (*tillochron*) had only one significant parameter (Table 5). The correlation levels obtained from all parameters were classified as high or very high correlation levels, as described in Table 3. We observed that among the significant parameters analyzed, only mla showed positive correlation (PRCC = 0.92) for LAI. The remaining parameters have a strong negative correlation with other variables, such as *tillochron* for TIL (PRCC = -0.92), *plastochron* for SDM and SFM (PRCC = -0.85 and PRCC = -0.83, respectively), and *n_lf_it_from* for POL (PRCC = -0.81) (Table 5). We perfomerd the GSA considering the parameters for the different layers of the soil and found its significance depending on the layer. Unlike the GSA for genotype parameters, there was no parameter with a strong correlation with the model output variables. The highest correlations were *WPp* for POL (PRCC = -0.70 and -0.65) at a depth of 100 cm (Table 6). TIL was the only variable that did not present any significant soil parameterts in both treatments. The texture parameters in both treatments. The texture parameters *Psand*, *Psilt* and *Pclay* were not evaluted, as well as the hydraulic parameters *Ksat* and *STp*. Finally, only the *FCp* and *WPp* parameters were the significant soil parameters, so in the stochastic simulation we inserted the uncertainty in the five layers (5,15,30,60, and 100 cm) to maintain the correlation between them.

Table 5. Value of the partial rank correlation coefficient (PRCC) for genotype parameters (PAR) for output variables stalk dry mass (SDM), stalk fresh mass (SFM), tillering (TIL), sucrose concentration (POL) and leaf area index (LAI). Parameters marked with * were statiscally significant at 1%.

PAR	GCTB treatment - PRCC					Bare treatment -PRCC				
	SDM	SFM	TIL	POL	LAI	SDM	SFM	TIL	POL	LAI
amax	0.00	0.01	-0.05	0.00	0.03	0.02	0.01	-0.21	0.02	0.06
chudec	-0.03	-0.03	0.05	0.01	-0.01	-0.03	-0.04	0.05	0.00	-0.03
chumat	0.03	0.04	-0.03	0.00	-0.05	0.03	0.05	-0.02	-0.04	-0.03
chupeak	-0.06	-0.07	0.03	0.00	-0.02	-0.07	-0.07	0.03	0.00	-0.02
chustk	0.05	0.04	-0.02	0.03	0.02	0.04	0.03	-0.07	0.03	0.02
eff	0.00	-0.01	0.04	0.07	0.08	0.10	0.08	0.00	0.04	0.17
end_tt_it_gro	-0.07	0.01	-0.02	-0.20	-0.01	-0.33	-0.19	-0.02	-0.34	0.07
end_tt_lf_gro	-0.02	-0.01	0.11	-0.10	-0.44	-0.02	-0.02	0.12	0.01	-0.40
init_lf_area	-0.06	-0.06	-0.11	0.00	-0.03	-0.07	-0.05	-0.26	-0.02	-0.11
max_ini_la	-0.03	-0.04	-0.12	0.02	-0.10	-0.08	-0.07	-0.41	0.01	0.21
max_it_dw	0.35	0.35	0.00	0.06	0.25	0.53	0.52	-0.27	-0.05	-0.26
maxdgl	0.01	0.02	-0.01	0.00	0.05	0.01	0.02	-0.02	0.00	0.03
maxgl	0.03	0.04	0.00	0.00	0.46	0.00	0.02	-0.02	0.02	0.50
mid_tt_it_gro	-0.54	-0.62	-0.01	0.34	0.01	-0.59	-0.73	0.00	0.64	0.07
mid_tt_lf_gro	0.04	0.06	0.69	0.00	-0.31	0.06	0.08	0.50	-0.02	0.01
mla	-0.11	-0.11	-0.65	0.02	0.92*	-0.15	-0.14	-0.50	0.02	0.91*
n_lf_it_from	-0.85*	-0.82*	-0.01	-0.50	0.11	-0.88*	-0.81*	0.00	-0.81*	0.25
n_lf_stk_em	-0.85*	-0.83*	-0.14	-0.50	0.3	-0.79	-0.70	-0.40	-0.64	0.49
phyllochron	-0.07	-0.06	-0.03	-0.10	0.01	-0.07	-0.07	-0.06	-0.02	-0.01
plastochron	-0.85*	-0.83*	0.63	-0.40	-0.55	-0.87*	-0.82*	0.69	-0.68	-0.13
popmat	0.17	0.19	0.71	0.00	0.73	0.23	0.25	0.63	-0.02	0.65
poppeak	-0.01	0.00	0.04	0.00	-0.01	0.01	0.01	-0.01	0.00	-0.02
sla	0.02	0.02	0.04	0.06	-0.03	0.02	0.01	-0.02	0.01	-0.01
tillochron	-0.15	-0.16	-0.92*	0.00	-0.71	-0.29	-0.30	-0.91*	0.01	-0.77

Table 6. Absolute value of the partial rank correlation coefficient (PRCC) for soil parameters (PAR) for output variables stalk dry mass (SDM), stalk fresh mass (SFM), tillering (TIL), sucrose concentration (POL) and leaf area index (LAI). Parameters marked with * were statistically significant at 1%.

PAR	Depth	GCTB treatment - PRCC					Bare treatment -PRCC				
	(cm)	SDM	SFM	LAI	POL	TIL	SDM	SFM	LAI	POL	TIL
FCp	5	0.03	0.05	0.01	-0.07	0.00	-0.15	-0.11	-0.17	-0.06	-0.11
	15	0.01	0.02	-0.04	-0.03	-0.01	-0.06	-0.05	-0.06*	-0.02	0.06
	30	0.17	0.25	-0.19	-0.34	0.00	-0.17	-0.08	-0.29*	-0.14	0.03
	60	0.06	0.09*	-0.04	-0.12*	0.00	0.08	0.12*	-0.03	-0.13*	0.05
	100	0.35	0.43*	0.07	-0.49*	-0.04	0.31	0.47*	0.01	-0.59*	0.03
Ksat	5	-0.01	-0.05	0.05	0.09	0.00	0.02	0.01	0.03	0.03	-0.09
	15	0.02	0.03	-0.02	-0.02	-0.04	-0.01	-0.03	0.00	0.03	-0.08
	30	-0.05	-0.09	0.10	0.17	-0.14	0.06	0.03	0.09	0.05	0.02
	60	0.07	0.06	0.12	-0.03	0.01	0.01	0.03	0.07	-0.01	-0.01
	100	0.06	0.07	0.05	-0.01	0.07	0.03	0.04	0.00	-0.03	-0.05
Pclay	5	0.00	0.02	0.03	-0.06	-0.03	-0.05	-0.03	0.03	-0.01	0.00
	15	0.04	0.03	0.08	0.00	0.02	0.06	0.05	0.09	-0.01	0.02
	30	0.02	0.02	0.05	0.00	0.00	0.06	0.04	0.08	0.01	-0.01
	60	-0.01	-0.01	-0.08	-0.02	-0.10	0.07	0.07	-0.03	-0.01	0.00
	100	-0.04	-0.03	-0.07	0.01	-0.06	-0.05	-0.03	-0.07	-0.02	-0.04
Psand	5	-0.01	0.01	0.03	-0.04	-0.10	-0.04	0.00	0.02	-0.06	-0.02
	15	-0.01	-0.03	0.04	0.02	-0.04	0.00	-0.01	0.04	0.07	0.05
	30	-0.02	-0.04	0.05	0.07	-0.10	0.03	0.03	0.05	-0.01	0.02
	60	-0.11	-0.06	-0.06	-0.07	-0.05	-0.06	-0.09	-0.01	0.09	-0.02
	100	0.07	0.07	0.00	-0.04	0.02	0.07	0.06	0.02	-0.03	0.04
Psilt	5	0.04	0.04	0.06	0.00	0.05	0.02	0.03	0.08	-0.04	0.04
	15	0.04	0.06	0.09	-0.09	-0.03	0.07	0.04	0.17	0.01	-0.06
	30	-0.06	-0.08	0.00	0.07	0.02	-0.02	-0.03	0.02	0.06	0.01
	60	-0.03	-0.05	-0.03	0.08	0.06	0.02	0.01	0.00	0.05	0.01
	100	0.00	0.04	0.00	-0.09	0.08	0.00	0.00	0.00	-0.06	0.00
STp	5	-0.16	-0.14	-0.35	0.04	0.10	-0.08	-0.07	-0.25	0.00	0.03
	15	-0.08	-0.08	-0.35	0.01	0.15	-0.05	-0.03	-0.23	-0.05	0.06
	30	-0.15	-0.17	-0.25	0.16	-0.02	-0.13	-0.11	-0.22	-0.02	-0.04
	60	0.04	0.04	0.17	0.01	-0.10	-0.03	-0.03	0.09	0.04	-0.05
	100	0.13	0.15	0.14	-0.14	-0.01	0.13	0.12	0.15	-0.03	0.09
WPp	5	-0.05	-0.11	0.07	0.18	-0.02	0.14	0.07	0.12	0.13	0.05
	15	-0.07	-0.08	0.07	0.05	-0.02	0.04	0.02	0.09	0.05	0.00
	30	-0.17	-0.24*	0.28*	0.32*	-0.02	0.24	0.13	0.44*	0.20	0.00
	60	-0.20*	-0.25*	-0.09	0.26*	0.00	0.00	-0.07	0.02	0.18*	0.02
	100	-0.60*	-0.67*	-0.30*	0.70*	0.02	-0.44*	-0.59*	-0.17	0.65*	0.00

2.3.2. Best parameters set obtained with GLUE

The values obtained from GLUE were compared with the values reported by Vianna et al., (2020), that used the BFGS technique (Broyden-Fletcher-Goldfarb-Shanno) to calibrate the genotype parameters of SAMUCA model for the same dataset. We noticed that there was a difference in Tilochron and mla (Table 7). These two genotype parameters increased by 18% for TILL and 4% for mla compared to the prior calibration. In the soil parameters the biggest difference was WPp at a depth of 30 cm, reaching 6%. The 5 and 15 cm layers did not exceed 2%. In the 60 and 100 cm layer, the variation was up to 5% for WPp compared with Vianna et al. (2020), that considered the average soil values for each soil layer.

To confirm there were differences in soil water storage from the parameters estimated by GLUE, we calculated the available water (AW) for sugarcane in each layer. Figure 1 shows that the correlation of soil parameters resulted in almost constant available water (AW) to the crop within the 0-15 cm soil depth, whereas the AW significantly varied for deeper layers (30-to-100 cm). The total available water (TAW) was 90 mm from the GLUE and the TAW obtained from the data by Vianna et al. (2020) was 102 mm.

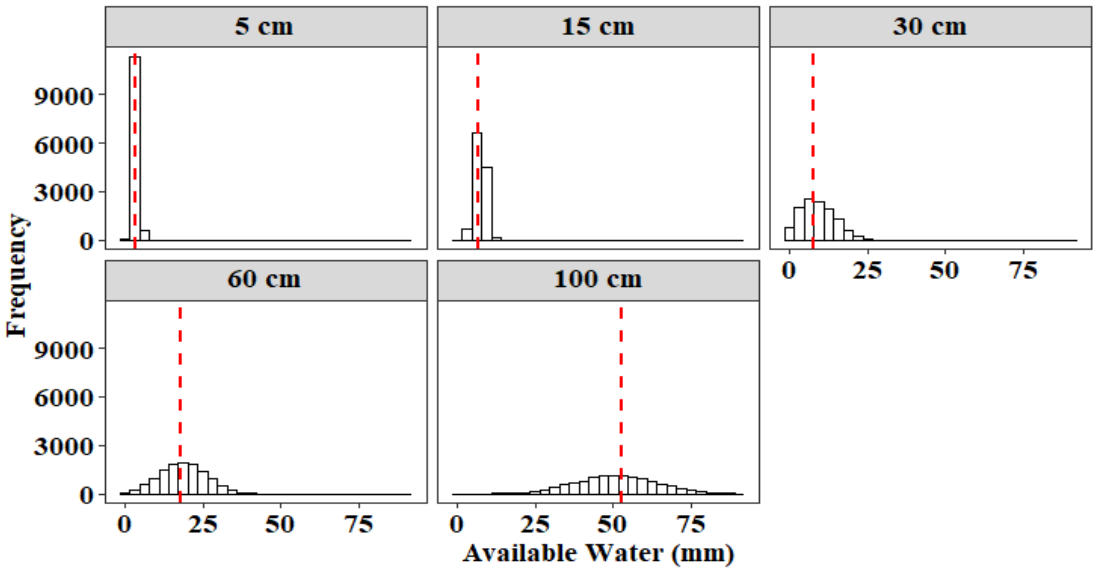


Figure 1. Histogram of total available water (AW) for 5 layers. The red dashed lines are the AW averages.

Table 7. Best set of parameter values considering four crop seasons (1 plant cane and 3 ratoons) for cultivar RB867515 based on the generalized likelihood uncertainty estimation method (GLUE) analyzing the uncertainty due to genotype parameters (UG), uncertainty due to soil parameters (US) and uncertainty due both genotype and soil parameters together (UGS).

	§ Parameters	UG	US	UGS	Calibration by
		(μ ± σ)	(μ ± σ)	(μ ± σ)	Vianna et al. (2020)
Genotype	n_lf_stk_eme	5 ± 1	#	5 ± 1	4
	n_lf_it_form	3 ± 1	#	3 ± 1	3
	tillochron	82 ± 20	#	82 ± 20	69
	mla	625 ± 82	#	627± 84	600
	plastochron	134 ± 15	#	135 ± 15	132
Soil	FCp (5 cm)	#	0.2800 ± 0.0144	0.2800 ± 0.0144	0.2850
	FCp (15 cm)	#	0.3060 ± 0.0055	0.3060 ± 0.0055	0.3030
	FCp (30 cm)	#	0.3590 ± 0.0316	0.3590 ± 0.0316	0.3470
	FCp (60 cm)	#	0.3760 ± 0.0175	0.3760 ± 0.0175	0.3940
	FCp (100 cm)	#	0.3900 ± 0.0209	0.3900 ± 0.0209	0.3930
	WPp(5 cm)	#	0.2110 ± 0.0112	0.2110 ± 0.0112	0.2160
	WPp(15 cm)	#	0.2340 ± 0.0061	0.2340 ± 0.0061	0.2400
	WPp(30 cm)	#	0.2950 ± 0.0361	0.2950 ± 0.0361	0.2780
	WPp(60 cm)	#	0.3110 ± 0.0222	0.3110 ± 0.0222	0.3070
	WPp(100 cm)	#	0.2661 ± 0.0283	0.2660 ± 0.0283	0.2530

§ For parameter definitions, see Table 2 and Table 3; # value used by calibration Vianna et al. (2020)
μ average calculated by Eq. 4 and σ standard deviation by Eq. 5.

2.3.3. Uncertainty analysis considering the genotype parameters (UG)

In the GCTB treatment, variables that had better model efficiency were SDM (EF = 0.83), SFM (EF = 0.75), POL (EF = 0.56), TIL (EF = 0.32) and LAI (EF = 0.70), respectively. The σ_{sim} over time (gray area in Figure 2A) was less or equal than the σ_{obs} for SFM (Figure 2A), representing 106% of the observed field variability (Table 8). The simulations for SFM (blue line) underestimated the observed data (Bias = -14.69 Mg ha^{-1}; Table 8), with an RMSE = 23.79 Mg ha-1 (Table 8). For SDM, the variability of the stochastic simulation was able to explain 64% of the observed data (Table 8). The simulations were also underestimated (Bias = -2.80 Mg ha-1; Table 8), and with an RMSE = 4.30 Mg ha-1. The POL, TIL, and LAI variables were also underestimated in comparison with observed data, showing Bias = -0.50, -2.48, and -0.01, respectively. Unlike SFM, SDM, and TIL, for POL and LAI, the variability was overestimated by 23% and 13%, respectively. The simulated variability for TIL = 52% of the observed one (Table 8).

For the Bare treatment, the SDM and SFM variables had EF = 0.83 and 0.87 (Table 8), followed by POL (EF = 0.58), TIL (EF = 0.53) and LAI (EF = 0.44) (Table 8). The variability of the stochastic simulation (gray area in Figure 2B) was less or equal, over time than the standard deviation of the observed SFM data (Figure 2B), explaining 85% of the variability seen in the field (Table 8). The simulation for SFM (blue line) was underestimated about the observed data (Bias = -12.66 Mg ha^{-1}; Table 8), with an RMSE of 20.94 Mg ha-1 (Table 8). For SDM, the variability of the stochastic simulation was able to explain 56% of the observed data (Table 8). The average of the simulations was also underestimated (Bias = -2.21 Mg ha^{-1}; Table 8), with an RMSE of 4.21 Mg ha-1. The variables POL and TIL were also underestimated about the observed data, with Bias = -0.50, -2.48, respectively. Unlike SFM, SDM, and TIL, for the POL and LAI variables, the variability was overestimated by 21% and 24% (Table 8), respectively. The simulated variability in TIL was 68% the observed variability (Table 8).

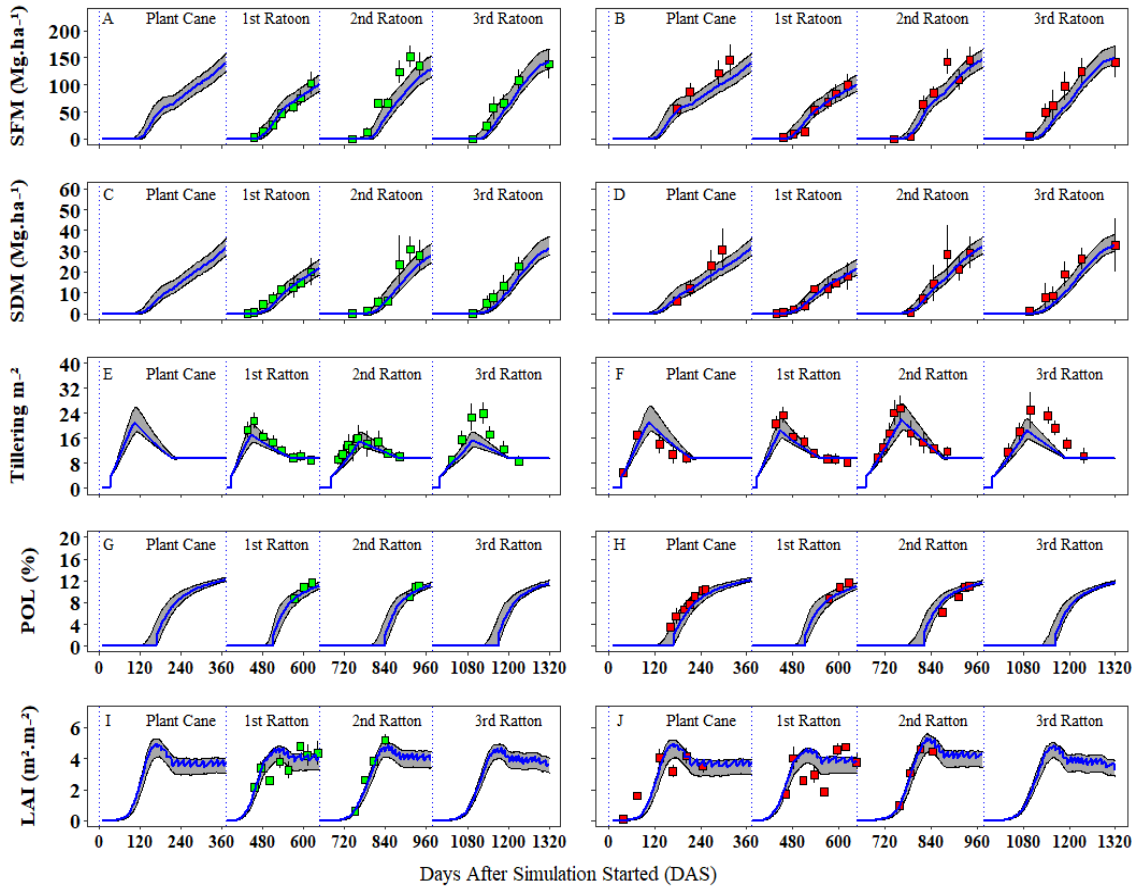


Figure 2. Representation of the uncertainty due to genotype parameters (UG) in stalk fresh (SFM) and dry (SDM) mass, tillering, sucrose concentration of fresh matter (POL) and leaf area index (LAI), considering parameters statistically significant in the global sensitivity analysis. Blue line simulation with best set parameters (Table 7); gray area is the standard deviation of the stochastic simulation; green and red square are the observed data with their respective error bar for treatments GCTB and Bare, respectively.

2.3.4. Uncertainty analysis considering the soil parameters (US)

Considering the data collected for the Bare treatment, the variability of the stochastic simulation did not well represent the variability observed in the observed data, considering all the variables analyzed (Figure 3). The variance in US was almost zero (SDM = 0.001%; SFM = 0.004%; POL = 0.001%; TIL = 0.0002% and LAI = 0.002%) for all variables (Table 8). The variables referring to mass, SFM and SDM, were well characterized over time by the best set parameters (blue line in Figures 4 A and C), with EF = 0.87 for both SDM and SFM (Table 8). TIL simulations well agreed with observed data (Table 8), with major discrepancies only for 3rd ratoon (Figure 3 E), for in which simulated TIL did not reach the observed peak of TIL and decreased faster than other ratooning cycles. For POL, EF = 0.82 and R^2 = 0.84, showing that model well simulated this output variable. For LAI, EF = 0.48 and R^2 = 0.58, and such weak results might be related to the great dispersion observed in this variable, mainly for 1st ratoon.

For GCTB treatment, the stochastic simulated variability for US was generally lower than that observed in the observed data (Table 8 and Figure 3), representing only 3% for SDM, 6% for SFM, 23% for POL, 3% for TIL and 15% for LAI (Table 8). Still, the variables referring to crop mass, such as SDM and SFM, were well characterized over time by the best set parameters (blue line in Figs 3A and B), with EF = 0.90 and 0.84 for SDM and SFM, respectively. For TIL, the simulations obtained good statistical indexes (Table 8), being negatively affected per 3rd

ratoon (Figure 3 E), for which maximum value of TIL was not well simulated. The LAI data observed had less dispersion in the 1st ratoon in the Bare treatment, which resulted in better statistical indices in relation to the simulation with GCTB

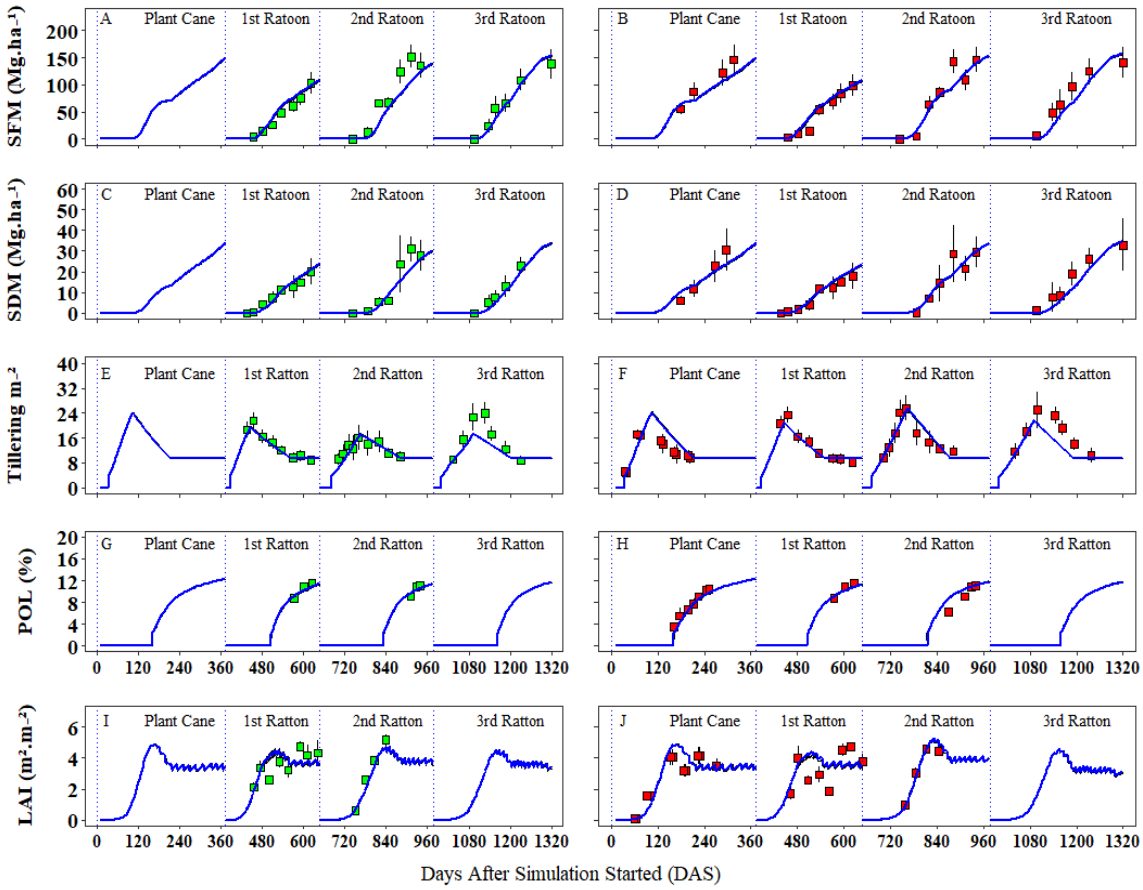


Figure 3. Representation of the uncertainty due to soil parameters (US) in stalk fresh (SFM) and dry (SDM) mass and tillering, sucrose concentration of fresh matter (POL) and leaf area index (LAI), considering parameters statistically significant in the global sensitivity analysis. Blue line simulation with best set parameters (Table 7); gray area is the standard deviation of the stochastic simulation; green and red square are the observed data with their respective error bar for treatments GCTB and Bare, respectively.

2.3.5. Uncertainty analysis considering the combined effect genotype and soil parameters (UGS)

For the Bare treatment, the UGS analysis had a similar performance than the UG in explaining both the variability and average of the observed data. The variables that had the best performance based on the coefficient of modeling efficiency (EF), were SDM (EF = 0.83), SFM (EF = 0.81), POL (EF = 0.55), TIL (EF = 0.49) and LAI (EF = 0.48). The variability of the stochastic simulation (gray area in Figure 4B) was less or equal, over time than the standard deviation of the observed SFM data (Figure 4B), explaining 84% of the variability seen in the field (Table 8). The simulations for SFM (blue line) underestimated observed data (Bias = -13.33 Mg ha^{-1}; Table 8), with an RMSE = 21.24 Mg ha^{-1} (Table 8). For SDM, the variability of the stochastic simulation was able to explain 56% of the observed data (Table 8). The simulation was also underestimated (Bias = -2.37 Mg ha^{-1}; Table 8), and with an RMSE = 4.24 Mg ha^{-1}. The TIL and POL variables were also underestimated about the observed data, with Bias = -

1.97 and -0.88, respectively. Unlike SFM, SDM, and TIL, POL and LAI variables, the variability was overestimated by 115% and 24% (Table 8), respectively. The simulated variability for TIL was 66% of the observed variability (Table 8).

The variability of the stochastic simulation was greater than the observed over time for GCTB treatment for all variables (Figure 4). It was overestimatied by 6% for SFM, 120% for POL, and 13% for LAI (Table 8). In the GCTB treatment the output variables with better performance were SDM (EF = 0.79), SFM (EF = 0.71), TIL (EF = 0.61), LAI (EF = 0.60) and POL (EF = 0.33, and the simulations for SFM (blue line) underestimated the observed data (Bias = -16.63 Mg ha^{-1}; Table 8), with an RMSE = 25.61 Mg ha^{-1} (Table 8). The simulations also underestimated SDM (Bias = -3.20 Mg ha^{-1}; Table 8), and with and RMSE = 4.69 Mg ha^{-1}, as well as POL and TIL (Bias = -0.72 and -1.49, respectively).

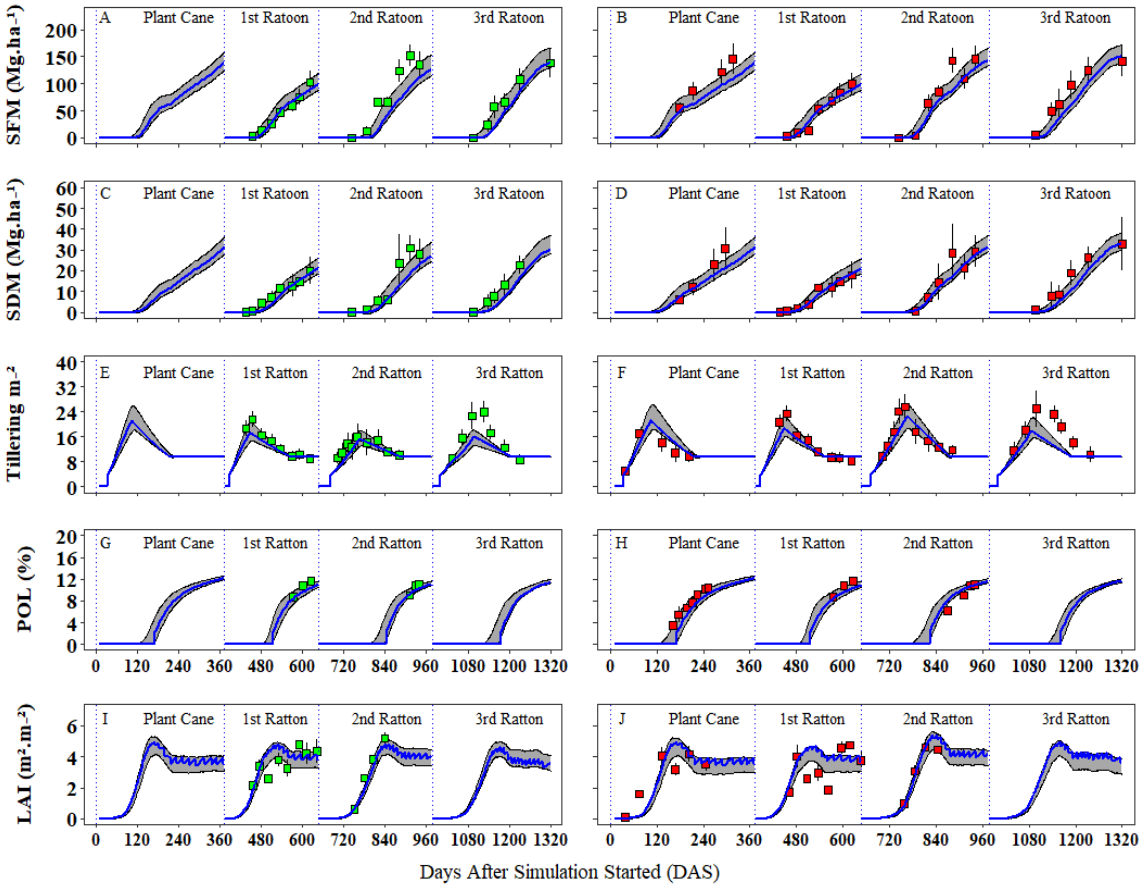


Figure 4. Representation of the uncertainty due to genotype and soil parameters (UGS) in stalk fresh (SFM) and dry (SDM) mass and tillering, sucrose concentration of fresh matter (POL) and leaf area index (LAI), considering parameters statistically significant in the global sensitivity analysis. Blue line simulation with best set parameters (Table 7); gray area is the standard deviation of the stochastic simulation; green and red square are the observed data with their respective error bar for GCTB and Bare treatments, respectively.

Table 8. Statistical indexes of performance of the SAMUCA model applied with best set of parameters. UG: Uncertainty analysis considering only genotype parameters; US: Uncertainty analysis considering only soil parameters; UGS: Uncertainty analysis considering both genotype and soil parameters.

Bare Treatment										
Variables	Uncertainty Analysis	Bias	RMSE	EF	R^2	d	σ_{obs}	σ_{sim}	Sample Size	ξ
Stalk	UG	-2.21	4.21	0.83	0.88	0.84		2.72		56%
Dry Mass	US	-0.87	3.69	0.87	0.88	0.84	4.83	$5.0e^{-05}$	25	0.001%
(Mg ha^{-1})	UGS	-2.37	4.24	0.83	0.89	0.84		2.71		56%
Stalk	UG	-12.66	20.94	0.87	0.88	0.84		12.92		85%
Fresh Mass	US	-5.24	17.41	0.87	0.89	0.84	15.18	$6.0e^{-04}$	24	0.004%
(Mg ha^{-1})	UGS	-13.33	21.24	0.81	0.89	0.80		12.80		84%
	UG	-0.77	1.52	0.58	0.83	0.68		1.15		221%
POL	US	-0.14	1.00	0.82	0.84	0.80	0.52	$3.0e^{-06}$	14	0.001%
(%[fresh])	UGS	-0.88	1.57	0.55	0.83	0.67		1.12		215%
Tillering	UG	-2.02	3.76	0.53	0.68	0.69		1.71		72%
(# m^{-2})	US	-0.73	3.07	0.68	0.70	0.76	2.53	$4.5e^{-06}$	34	0.0002%
	UGS	-1.97	3.89	0.49	0.63	0.68		1.67		70%
LAI	UG	0.19	0.99	0.44	0.58	0.65		0.42		124%
($m^2.m^{-2}$)	US	0.03	0.95	0.48	0.58	0.66	0.34	$7.1e^{-06}$	23	0.002%
	UGS	0.18	0.95	0.48	0.10	0.67		0.42		124%

*continues on the next page

GCTB Treatment										
Stalk	UG	-2.80	4.30	0.83	0.92	0.87		2.59		64%
Dry Mass	US	-1.32	3.27	0.90	0.92	0.87	4.02	0.11	21	3%
(Mg ha^{-1})	UGS	-3.20	4.69	0.79	0.90	0.81		2.59		64%
Stalk	UG	-14.69	23.79	0.75	0.87	0.83		13.26		106%
Fresh Mass	US	-6.83	18.87	0.84	0.87	0.83	12.53	0.77	20	6%
(Mg ha^{-1})	UGS	-16.63	25.61	0.71	0.83	0.77		13.28		106%
POL	UG	-0.50	0.70	0.56	0.87	0.66		0.78		223%
(%[fresh])	US	-0.23	0.57	0.71	0.94	0.71	0.35	0.08	6	23%
	UGS	-0.72	0.87	0.33	0.88	0.60		0.77		220%
Tillering	UG	-2.48	3.28	0.32	0.70	0.63		1.30		52%
(# m^{-2})	US	-1.49	2.73	0.61	0.74	0.74	2.52	0.07	24	3%
	UGS	-1.49	2.71	0.61	0.74	0.74		1.32		52%
LAI	UG	-0.01	0.65	0.70	0.72	0.73		0.44		113%
($m^2.m^{-2}$)	US	-0.09	0.70	0.66	0.68	0.69	0.39	0.06	12	15%
	UGS	0.04	0.76	0.60	0.66	0.69		0.44		113%

Bias: model bias index; RMSE: Root mean squared error; EF: Modeling efficiency; R^2: Determination index;
d: accuracy index of Wilmot; σ_{obs} is the average of the standard deviation of the observed data;
σ_{sim} is the average of the standard deviation of the simulated data; ξ is the ratio bet $\sigma_{sim}/\sigma_{obs}$ in percentage

2.4. Discussion

We observed that Bare and GCTB treatments influenced the GSA results, here used to select which genotype and soil parameters to use in the uncertainty analysis (Table 5 and 6). Therefore, using both treatment was complemental for choosing the parameters sets that best represent the field variability. Results from the GSA shown that the most influential parameters of SAMUCA to the main sugarcane growth components were the genotype parameters of *plastochron*, *n_lf_when_eme*, *n_lf_it_from*, *tillochron*, and *mla*; and the soil parameters of field capacity (*FCp*) and wilting point (*WPp*). We note the inclusion of the *n_lf_when_eme* parameter only for the GCTB treatment (SFM and SDM), whereas we didn't find any statistically significant parameter for the POL output. In total, the new version of SAMUCA has 101 parameters that were divided to represent the species, ecotype and genotype characteristics of sugarcane, accordingly with the DSSAT framework (Vianna et al., 2020; Jones et al., 2003). In our study, we considered that only the genotype parameters would have an influence on the simulation's uncertainty (Table 4), assuming that the species and ecotype parameters were well defined. Further, finding plausible ranges for all the species and ecotype parameters is challenging, and considering the full list of parameters would dramatically increase the computation requirements of this study (GSA, GLUE and stochastic).

The calibration obtained from GLUE to UG had a lower performance for all variables (Table 8), when compared to the simulation performed by Vianna et al., (2020). However, we must emphasize that we do not estimate all genotype parameters, only those significant that were obtained from the GSA. In future studies it would be interesting to evaluate the different calibration methods, such as GLUE and BFGS to the operational cost for the simulation and performance, while there's still no consensus on the choices of methods and decisions made by modelers during crop models calibrations (Wallach et al., 2020). Nevertheless, we observed that the application of GLUE to soil parameters (US) generated a performance like the results of Vianna et al., (2020) (Table 8).

The PRCC method provides answers to questions about how the result is affected if we increase (or decrease) a specific parameter (linearly discounting the effects on the other parameters) (Marino et al., 2008). Thus, the PRCC can be informative about which parameters to target if we are to achieve specific objectives. For example, one can identify the set of parameters that most likely can be used to determine how to increase biomass (SDM or SFM) with the PRCC results. The main limitation of this method is that it does not answer which parameters are responsible for the greatest variance in the model's output (Marino et al., 2008). Different simulation conditions such as the biophysical environment (Sexton et al., 2017), management (Zhang et al., 2020), and even GSA methods (Drouet et al., 2011; Marino et al., 2008) can generate divergent results obtained from GSA. Thus, for a more robust overview of the model's sensitivity to parameters, more than one GSA method and other experimental sets could be considered in future studies to confirm our findings.

Nevertheless, when one wants to explore the variability in the environment through stochastic simulation, not necessarily the use of statistically significant parameters would produce the best results. In the present study, we show that by considering only the statistically significant parameters we were able to well simulate the mean of field observations. However, such procedure overestimated the variability of some model's outputs in comparison with that observed in the experimental field, as it was found for POL and LAI variables in Bare and SFM, POL, and LAI in GCTB (Table 8). We used the whole set of observed data in GLUE procedure, but a possible solution would be on implementing a filter in GLUE methodology to constrain the generated parameters within the observed variability.

The greatest variability in simulations for UG and UGS was due to the greater model sensitivity to genotype parameters. For the US, we found that the variability was less than the other two (UG and UGS). This result agreed with previous studies showing soil parameters with less influence on model behavior likely as a result of irrigation in the experiment, which further reduced the model sensitivity to soil parameters (Attia et al., 2021; Dejonge et al., 2012; Zhang et al., 2020). We have not included the soil textural parameters because they did not have any significance in the GSA, nor were the physical characteristics of the mulch layer in our analysis of US and UGS. In addition, our field conditions were not limited (adequate inputs in clay soil) and the distribution of soil parameters showed small variation TAW, which can help explain why the soil parameters did not have a greater influence in the field variability.

The model performance under GCTB conditions was slightly better for the SDM, TIL and LAI in comparison with the Bare treatment, whereas the Bare simulations performed better in the SDM and POL simulation. The GCTB is interpreted in the SAMUCA model as an additional soil layer, with its respective saturation point and water content (Vianna et al., 2020). According to Ritchie (1998), the number of layers and their depth is an important factor to simulate the water balance more precisely. This is specifically important to guarantee water and heat fluxes in the soil medium (Harper et al., 2020). Furthermore, the SAMUCA is still not capable of capturing all the belowground processes affecting crop growth, such as soil compaction, nutrient uptake, and microbiological processes (Vianna et al., 2020). These model limitations may explain the low capacity to simulate the variability seen in US. Finally, it leads to two possible causes for the low model responses to soil parameters found in the present study: (i) the low influence of the soil hydraulic parameters in a irrigated experiment; and (ii) that the observed variability in the field is not fully explained by the soil hydraulic process and parameters represented in the model.

2.5. Conclusion

The GSA was a useful tool for choosing parameters for stochastically simulating crop growth and development aiming to explore the genotype variability existing in the environment. The UG and UGS had the same capacity to quantify the variability present in the environment for the treatments Bare and GCTB, and we did not find any influence of soil parameters in model variability, probably because our data were collected in fully irrigated experiments and with no nutritional limitation. In our case, because the water stress is the main reducing factor linked with soil that is accounted for in the SAMUCA model, the sensitivity to the soil parameters may be simply ignored and the genotype parameters can be chosen as the only source of variability for practical applications. Indeed, the simulated variability found in the US was caused by GCTB and not due to soil hydraulic parameters. Our suggestion for future work is to explore rainfed environments, different amounts of GCTB and other soil types.

References

Attia, A., El-Hendawy, S., Al-Suhaibani, N., Tahir, M.U., Mubushar, M., Vianna, M. dos S., Ullah, H., Mansour, E., Datta, A., 2021. Sensitivity of the DSSAT model in simulating maize yield and soil carbon dynamics in arid Mediterranean climate: Effect of soil, genotype and crop management. F. Crop. Res. 260, 107981. https://doi.org/10.1016/j.fcr.2020.107981

Baigorria, G.A., 2014. Stochastic Models to Generate Geospatial-, Temporal-, and Cross-Correlated Daily Maximum and Minimum Temperatures. https://doi.org/10.1155/2014/365362

Baigorria, G.A., Jones, J.W., 2010. GiST: A Stochastic Model for Generating Spatially and Temporally Correlated Daily Rainfall Data. J. Clim. 23, 5990–6008. https://doi.org/10.1175/2010JCLI3537.1

Bezuidenhout, C., 2000. A model review and proposed mechanistic tiller model for the canegro sugarcane crop model. undefined. Faculty of Engineering at Technikon Natal, Durban.

Bezuidenhout, C.N., O'Leary, G.J., Singels, A., Bajic, V.B., 2003. A process-based model to simulate changes in tiller density and light interception of sugarcane crops. Agric. Syst. 76, 589–599. https://doi.org/10.1016/S0308-521X(02)00076-8

Brogi, C., Huisman, J.A., Herbst, M., Weihermüller, L., Klosterhalfen, A., Montzka, C., Reichenau, T.G., Vereecken, H., 2020. Simulation of spatial variability in crop leaf area index and yield using agroecosystem modeling and geophysics-based quantitative soil information. Vadose Zo. J. 19. https://doi.org/10.1002/vzj2.20009

Carvalho, J.L.N., Nogueirol, R.C., Menandro, L.M.S., Bordonal, R. de O., Borges, C.D., Cantarella, H., Franco, H.C.J., 2017. Agronomic and environmental implications of sugarcane straw removal: a major review. GCB Bioenergy 9, 1181–1195. https://doi.org/10.1111/gcbb.12410

Coelho, A.P., Dalri, A.B., Fischer Filho, J.A., de Faria, R.T., Silva, L.S., Gomes, R.P., 2020. Calibration and evaluation of the DSSAT/Canegro model for sugarcane cultivars under irrigation managements. Rev. Bras. Eng. Agric. e Ambient. 24, 52–58. https://doi.org/10.1590/1807-1929/agriambi.v24n1p52-58

Dejonge, K.C., Ascough Ii, J.C., Ahmadi, M., Andales, A.A., Arabi, M., 2012. Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. Ecol. Modell. 231, 113–125. https://doi.org/10.1016/j.ecolmodel.2012.01.024

Ehara, H., Takamura, T., Tsuchiya, M., 1994. Growth and dry matter production of sugar cane in warm temperate zone of Japan. Japanese J. Trop. Agric. 38, 335–342.

Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agron. Sustain. Dev. 36. https://doi.org/10.1007/s13593-016-0364-z

Everingham, Y.L., Muchow, R.C., Stone, R.C., Inman-Bamber, N.G., Singels, A., Bezuidenhout, C.N., 2002. Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts. Agric. Syst. 74, 459–477. https://doi.org/10.1016/S0308-521X(02)00050-1

Faivre, R., Leenhardt, D., Voltz, M., Benoît, M., Papy, F., Dedieu, G., Wallach, D., 2009. Spatialising Crop Models BT - Sustainable Agriculture, in: Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S., Alberola, C. (Eds.), . Springer Netherlands, Dordrecht, pp. 687–705. https://doi.org/10.1007/978-90-481-2666-8_42

Goldemberg, J., Coelho, S.T., Guardabassi, P., 2008. The sustainability of ethanol production from sugarcane. Energy Policy 36, 2086–2097. https://doi.org/https://doi.org/10.1016/j.enpol.2008.02.028

Harper, A., Williams, K., McGuire, P., Duran Rojas, M.C., Hemming, D., Verhoef, A., Huntingford, C., Rowland, L., Marthews, T., Breder Eller, C., Mathison, C., Nobrega, R., Gedney, N., Vidale, P.L., Otu-Larbi, F., Pandey, D., Garrigues, S., Wright, A., Slevin, D., De Kauwe, M., Blyth, E., Ärdo, J., Black, A., Bonal, D., Buchmann, N., Burban, B., Fuchs, K., de Grandcourt, A., Mammarella, I., Merbold, L., Montagnani, L., Nouvellon, Y., Restrepo-Coupe, N., Wohlfahrt, G., 2020. Improvement of modelling plant responses to low soil moisture in JULESvn4.9 and evaluation against flux tower measurements. Geosci. Model Dev. Discuss. 1–42. https://doi.org/10.5194/gmd-2020-273

He, J., Jones, J.W., Graham, W.D., Dukes, M.D., 2010. Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method. Agric. Syst. 103, 256–264. https://doi.org/10.1016/j.agsy.2010.01.006

He, J., M. D. Dukes, M.D., J. W. Jones, J.W., W. D. Graham, W.D., J. Judge, J., 2009. Applying GLUE for Estimating CERES-Maize Genetic and Soil Parameters for Sweet Corn Production. Trans. ASABE 52, 1907–1921. https://doi.org/10.13031/2013.29218

Hoffman, N., Singels, A., Patton, A., Ramburan, S., 2018. Predicting genotypic differences in irrigated sugarcane yield using the Canegro model and independent trait parameter estimates. Eur. J. Agron. 96, 13–21. https://doi.org/10.1016/j.eja.2018.01.005

Inman-Bamber, N.G., 1994. Temperature and seasonal effects on canopy development and light interception of sugarcane. F. Crop. Res. 36, 41–51. https://doi.org/10.1016/0378-4290(94)90051-5

Jaiswal, D., De Souza, A.P., Larsen, S., Lebauer, D.S., Miguez, F.E., Sparovek, G., Bollero, G., Buckeridge, M.S., Long, S.P., 2017. Brazilian sugarcane ethanol as an expandable green alternative to crude oil use. Nat. Clim. Chang. 7, 788–792. https://doi.org/10.1038/nclimate3410

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Brief history of agricultural systems modeling. Agric. Syst. 155, 240–254. https://doi.org/10.1016/J.AGSY.2016.05.014

Jones, J.W., He, J., Boote, K.J., Wilkens, P., Porter, C.H., Hu, Z., 2011. Estimating DSSAT cropping system cultivar-specific parameters using Bayesian techniques. Methods Introd. Syst. Model. into Agric. Res. 365–394. https://doi.org/https://doi.org/10.2134/advagricsystmodel2.c13

Jones, M.R., Singels, A., Ruane, A.C., 2015. Simulated impacts of climate change on water use and yield of irrigated sugarcane in South Africa. Agric. Syst. 139, 260–270. https://doi.org/https://doi.org/10.1016/j.agsy.2015.07.007

Laclau, P.B., Laclau, J.P., 2009. Growth of the whole root system for a plant crop of sugarcane under rainfed and irrigated environments in Brazil. F. Crop. Res. 114, 351–360. https://doi.org/10.1016/j.fcr.2009.09.004

Li, Zhenhai, He, J., Xu, X., Jin, X., Huang, W., Clark, B., Yang, G., Li, Zhenhong, 2018. Estimating genetic parameters of DSSAT-CERES model with the GLUE method for winter wheat (Triticum aestivum L.) production. Comput. Electron. Agric. 154, 213–221. https://doi.org/10.1016/j.compag.2018.09.009

Lingle, S.E., 1999. Sugar Metabolism during Growth and Development in Sugarcane Internodes. Crop Sci. 39, 480–486. https://doi.org/10.2135/cropsci1999.0011183X0039000200030x

Liu, D.L., Kingston, G., Bull, T.A., 1998. A new technique for determining the thermal parameters of phenological development in sugarcane, including suboptimum and supra-optimum temperature regimes. Agric. For. Meteorol. 90, 119–139. https://doi.org/10.1016/S0168-1923(97)00087-7

Marin, F., Jones, J.W., Boote, K.J., 2017. A stochastic method for crop models: including uncertainty in a sugarcane model. Agron. J. 109, 483–495. https://doi.org/https://doi.org/10.2134/agronj2016.02.0103

Marin, F.R., Jones, J.W., 2014. Process-based simple model for simulating sugarcane growth and production. Sci. Agric. 71, 1–16. https://doi.org/10.1590/S0103-90162014000100001

Marin, F.R., Rattalino Edreira, J.I., Andrade, J., Grassini, P., 2019. On-farm sugarcane yield and yield components as influenced by number of harvests. F. Crop. Res. 240, 134–142. https://doi.org/10.1016/j.fcr.2019.06.011

Marin, F.R., Thorburn, P.J., Nassif, D.S.P., Costa, L.G., 2015. Sugarcane model intercomparison: Structural differences and uncertainties under current and potential future climates. Environ. Model. Softw. 72, 372–386. https://doi.org/10.1016/j.envsoft.2015.02.019

Mukaka, M.M., 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24, 69–71.

Nash, J.E., Sutcliffe, J.V., 1970. River flow forecasting through conceptual models part I — A discussion of principles. J. Hydrol. 10, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nassif, D.S.P., Marin, F.R., Costa, L.G., 2014. Evapotranspiration and Transpiration Coupling to the Atmosphere of Sugarcane in Southern Brazil: Scaling Up from Leaf to Field. Sugar Tech 16, 250–254. https://doi.org/10.1007/s12355-013-0267-0

O'Leary, G.J., 2000. A review of three sugarcane simulation models with respect to their prediction of sucrose yield. F. Crop. Res. 68, 97–111. https://doi.org/10.1016/S0378-4290(00)00112-X

Petersen, R.G., 1994. Agricultural Field Experiments: Design and Analysis.

Sage, R.F., Peixoto, M.M., Sage, T.L., 2013. Photosynthesis in Sugarcane, Sugarcane: Physiology, Biochemistry, and Functional Biology. https://doi.org/10.1002/9781118771280.ch6

Silva, M. de A., Jifon, J.L., dos Santos, C.M., Jadoski, C.J., da Silva, J.A.G., 2013. Photosynthetic capacity and water use efficiency in sugarcane genotypes subject to water deficit during early growth phase. Brazilian Arch. Biol. Technol. 56, 735–748. https://doi.org/10.1590/S1516-89132013000500004

Silveira Pinto Nassif, D., Ricardo Marin, F., José Pallone Filho, W., Souza Resende, R., Queiroz Pellegrino, G., Antonio, S., 2012. Parametrização e avaliação do modelo DSSAT/Canegro para variedades brasileiras de cana-de-açúcar Parameterization and evaluation of the DSSAT/Canegro model for Brazilian sugarcane varieties, Pesq. agropec. bras. https://doi.org/https://doi.org/10.1590/S0100-204X2012000300001

Sinclair, T.R., Seligman, N.G., 1996. Crop Modeling: From Infancy to Maturity. Agron. J. 88, 698–704. https://doi.org/10.2134/agronj1996.00021962008800050004x

Singels, A., Bezuidenhout, C.N., 2002. A new method of simulating dry matter partitioning in the Canegro sugarcane model. F. Crop. Res. 78, 151–164. https://doi.org/https://doi.org/10.1016/S0378-4290(02)00118-1

Singels, A., Jones, M., Marin, F., Ruane, A., Thorburn, P., 2013. Predicting Climate Change Impacts on Sugarcane Production at Sites in Australia, Brazil and South Africa Using the Canegro Model. Sugar Tech 16, 347–355. https://doi.org/10.1007/s12355-013-0274-1

Smit, M.A., Singels, A., 2006. The response of sugarcane canopy development to water stress. F. Crop. Res. 98, 91–97. https://doi.org/10.1016/j.fcr.2005.12.009

Sreelash, K., Sekhar, M., Ruiz, L., Tomer, S.K., Guérif, M., Buis, S., Durand, P., Gascuel-Odoux, C., 2012. Parameter estimation of a two-horizon soil profile by combining crop canopy and surface soil moisture observations using GLUE. J. Hydrol. 456–457, 57–67. https://doi.org/10.1016/j.jhydrol.2012.06.012

Usowicz, B., Lipiec, J., 2017. Spatial variability of soil properties and cereal yield in a cultivated field on sandy soil. Soil Tillage Res. 174, 241–250. https://doi.org/10.1016/j.still.2017.07.015

van Bussel, L.G.J., Ewert, F., Zhao, G., Hoffmann, H., Enders, A., Wallach, D., Asseng, S., Baigorria, G.A., Basso, B., Biernath, C., Cammarano, D., Chryssanthacopoulos, J., Constantin, J., Elliott, J., Glotter, M., Heinlein, F., Kersebaum, K.C., Klein, C., Nendel, C., Priesack, E., Raynal, H., Romero, C.C., Rötter, R.P., Specka, X., Tao, F., 2016. Spatial sampling of weather data for regional crop yield simulations. Agric. For. Meteorol. 220, 101–115. https://doi.org/10.1016/j.agrformet.2016.01.014

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environ. Model. Softw. 25, 310–319. https://doi.org/https://doi.org/10.1016/j.envsoft.2009.09.012

Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2018. Working with dynamic crop models: methods, tools and examples for agriculture and environment. Academic Press.

Wallach, D., Makowski, D., Jones, J.W., Brun, F., Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2019. Data Assimilation for Dynamic Models. Work. with Dyn. Crop Model. 487–518. https://doi.org/10.1016/B978-0-12-811756-9.00014-9

Wallach, D., Palosuo, T., Thorburn, P., Hochman, Z., Gourdain, E., Andrianasolo, F., Asseng, S., Basso, B., Buis, S., Crout, N., Dibari, C., Dumont, B., Ferrise, R., Gaiser, T., Garcia, C., Gayler, S., Ghahramani, A., Hiremath, S., Hoek, S., Horan, H., Hoogenboom, G., Huang, M., Jabloun, M., Jansson, P.E., Jing, Q., Justes, E., Kersebaum, K.C., Klosterhalfen, A., Launay, M., Lewan, E., Luo, Q., Maestrini, B., Mielenz, H., Moriondo, M., Nariman Zadeh, H., Padovan, G., Olesen, J.E., Poyda, A., Priesack, E., Pullens, J.W.M., Qian, B., Schütze, N., Shelia, V., Souissi, A., Specka, X., Srivastava, A.K., Stella, T., Streck, T., Trombi, G., Wallor, E., Wang, J., Weber, T.K.D., Weihermüller, L., de Wit, A., Wöhling, T., Xiao, L., Zhao, C., Zhu, Y., Seidel, S.J., 2020. The chaos in calibrating crop models. bioRxiv. https://doi.org/10.1101/2020.09.12.294744

Wallach, D., Thorburn, P., Asseng, S., Challinor, A.J., Ewert, F., Jones, J.W., Rotter, R., Ruane, A., 2016. Estimating model prediction error: Should you treat predictions as fixed or random? Environ. Model. Softw. 84, 529–539. https://doi.org/10.1016/j.envsoft.2016.07.010

Wallach, D., Thorburn, P.J., 2017. Estimating uncertainty in crop model predictions: Current situation and future prospects. Eur. J. Agron. 88, A1–A7. https://doi.org/10.1016/j.eja.2017.06.001

Willmott, C.J., Robeson, S.M., Matsuura, K., 2012. A refined index of model performance. Int. J. Climatol. 32, 2088–2094. https://doi.org/10.1002/joc.2419

Zhang, Y., Arabi, M., Paustian, K., 2020. Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. Environ. Model. Softw. 126. https://doi.org/10.1016/j.envsoft.2020.104642

Zhou, M., Singels, A., Savage, M.J., 2003. Physiological parameters for modelling differences in canopy development between sugarcane cultivars. Proc. South African Sugar Technol. Assoc. 77, 610–621.

Zhou, M.M., Shoko, M.D., 2011. Seasonal and varietal effects on tiller population development of sugarcane (Saccharum Officinarum L.). South African J. Plant Soil 28, 11–16. https://doi.org/10.1080/02571862.2011.10640007

3. SAMPLING SCHEME, RANGE OF PARAMETERS, AND TIME-DEPENDENT EFFECTS ON GLOBAL SENSITIVITY ANALYSIS IN SUGARCANE MODELLING

Abstract:

Process-based crop models (PBCM) are the state of the art in agricultural modeling. Some PBCMs are increasingly complex and require many parameters to describe crop processes and boundary conditions. Sugarcane is a key crop for food and fuel security in many countries around the world. PBCMs for sugarcane present a high number of genotype parameters compared to other crops, which make it harder to calibrate. Global sensitivity analysis (GSA) has thus become an important tool for understanding, calibrating and further developing PBCMs. The GSA methods are based on Bayesian approach and rely on sampling techniques and parameters variation ranges. In this paper we used a recently updated sugarcane model (SAMUCA) to simulate the crop growth and development along a 4-year field experiment, conducted with two treatments: with green cane trash blanket (GCTB) and under bare soil (Bare). Using the extended Fourier Amplitude Sensitivity (eFAST) algorithm, GSA was performed on the 24 genotype parameters of the SAMUCA. The objective of this study was the determine the sample size, the influence of parameters range, and to quantify the genotype parameters responsible for the greater uncertainty in simulation of the SAMUCA model with bare soil and GCTB. The results showed that sample size highly affected the convergence of the sensitivity indices, and it differed as a function of the output variable. In our case, the required sample size must be greater than 2049 for the analysis to cover all variables. Two sets of parameter ranges were used for analysis (the first set uses maximum and integer values of each parameter reported in the literature; the second set applied a 25% perturbation to the previously calibrated values), and the results indicated that the parameters range affected the order of importance of the parameters. Furthermore, we identified that at different phenological stages during the sugarcane, distinct parameters were responsible for explaining the most variance of the output. However, there was no difference among ratoons or interference in the results of bare soil or GCTB.

Keywords: SAMUCA, Uncertainty, Sample size, Crop modelling

3.1. Introduction

In agriculture, process-based crop models (PBCM) represent the state of the art for simulating crop growth and development (Jones et al., 2017; Marin et al., 2017). When properly calibrated, they are commonly used to simulate crop growth and development under different conditions, thus being able to test hypothetical management, climate, and soil scenarios (Faivre et al., 2009). Scientists and decision-makers have used modelling as a tool to address issues related to the sugar and bioenergy sectors, including climate change (Jones et al., 2015; Marin et al., 2013; Singels et al., 2013), plant breeding (Hoffman et al., 2018), risk analysis (Everingham et al., 2002) and crop forecast (Everingham et al., 2016).

Sugarcane is a crucial crop for world bioenergy (Raza et al., 2019), and several authors have studied sugarcane crop modeling and sugarcane crop modelling (Inman-Bamber and Smith, 2005; Jones and Singels, 2018; Keating et al., 1999; Marin and Jones, 2014; Singels and Bezuidenhout, 2002; Thorburn et al., 2005; Valade et al., 2014; Vianna et al., 2020). Singels (2013) presented a detailed review of the main sugarcane models in the literature, highlighting sugarcane as one of the crops with a high need to be represented in PBCM given its specific farming systems and logistic requirements. To represent its physiological complexity, sugarcane PBCM have many genotype parameters compared to other crops, such as maize and wheat. For instance, the sugarcane models DSSAT/CASUPRO and DSSAT/CANEGRO make use of 33 and 18 genotype cultivar parameters respectively, while DSSAT/CERES-MAIZE and DSSAT/CERES-WHEAT use only respectively 6 and 7 cultivar parameters to be calibrated.

According to Sinclair & Seligman (1996), the development of different PBCM by more research groups allows improve the understanding of processes. In this context, Marin & Jones (2014) developed the SAMUCA focusing on the specific characteristics of sugarcane farming systems in Brazil. Recently, the SAMUCA model was improved by reducing the uncertainties around the soil water balance, heat flux and physiological mechanisms such as carbon partition, photosynthesis, tillering and root growth (Vianna et al., 2020).

As any other PBCM, SAMUCA represents a simplification of the real system and requires several parameters whose determination is a problem for practical operational applications (Makowski et al., 2002). Most parameters are acquired through field observations, which are expensive and time-consuming, and the acquisition of certain parameters is difficult. Yet, many parameters vary depending on environmental conditions, cultivars, seasonal variation, among other factors (Wang et al., 2013).

Furthermore, for reliable simulations, accurate parameter estimation is required (Guérif and Duke, 2000; Wallach et al., 2019), and so several parameter estimation algorithms have been developed, part of which based on Bayesian approaches (He et al., 2010, 2009; Marin et al., 2017; Sheng et al., 2019; Sreelash et al., 2012; Zhang et al., 2020). To some extent, these methods solved the problem of difficult-to-acquire parameters, and they are quite efficient but applicable to a small number of parameters (Varella et al., 2010). Still, the inclusion of many parameters in a PBCM raises a dilemma related to the difficulty to simultaneously estimate all unknown parameters and ensuring at the same time they keep their biophysical meaning coherent.

In practice, it is well known that only part of the parameters is usually responsible for most of the model uncertainty, while most of them have only minor influence (LI et al., 2019; Varella et al., 2010; Zhang et al., 2020). The parameter sensitivity analysis (SA) method can identify the most important parameters for a given model output variable, which allows users to focus on the most important model parameters during the calibration process. Furthermore, based on the SA, the balance and robustness of the model can be analyzed for future improvement, model development, and applications (Chu-Agor et al., 2011; Confalonieri, 2010; Fraedrich and Goldberg, 2000; Hirabayashi et al., 2011).

The SA can be divided into two groups: the local sensitivity analysis (LSA) and the global sensitivity analysis (GSA). The LSA consists of changing a single parameter at a time, while the other parameters are kept at their reference values; in other words, this method is based on the local derivatives of the model's output concerning the variation of a single parameter, which indicates how strong is the output changes around the reference parameter values (A Saltelli et al., 1999). The GSA allows you to evaluate the entire uncertainty range of parameters, considering changes in all parameters along with their range, as well as the interactions among parameters (A Saltelli et al., 1999).

The GSA methods can also be classified into three groups: screening, regression, and variance; all of these following the Bayesian sampling principle. The most used screening method is the Morris method, which permits to define the most important model parameters and it is often considered a qualitative method (Dejonge et al., 2012; Morris, 1991). Regression methods, such as the Partial Rank Correlation Coefficient, provide the correlation between the model's output and the selected parameters and they are mandatory when the parameter and the model outputs have a monotonic relationship (Krishnan and Aggarwal, 2018; Marino et al., 2008). The methods based on the variance are most common, and the three mains are: Sobol (2001), Fourier Amplitude Sensitivity Test (FAST) (Cukier et al., 1978), and the extended Fourier Amplitude Sensitivity Test (eFAST) (A Saltelli et al., 1999); they provide information about the parameters causing the highest variability in the model output and they usually demand a high computational cost. To define which method to be applied is the most suitable, some properties of the model must be known (linearity, prior distribution of parameters, and monotonicity), furthermore considering

the number of parameters to be evaluated and the computational cost (Iooss and Lemaître, 2015). The Sobol and eFAST are the most applicable to any type of PBCM, but while Sobol is very computationally expensive, eFAST integrates the merits of FAST and Sobol's algorithms, representing a method with high efficiency and precision, and ability to adequately compute interaction effects among parameters (Iooss and Lemaître, 2015).

The GSA has several aspects that can affect sensitivity indices and their uncertainty, regardless of the method adopted. In general, the most important uncertainty sources of GSA are: (i) sample size, (ii) range of parameters, and (iii) complexity of the model (Gan et al., 2014; Song et al., 2015; Xu and Gertner, 2011). To our knowledge, there are no studies in the literature investigating effects in GSA caused by sample size and parameter range on eFAST method in sugarcane models. In the case of sample size, the available studies used eFAST and were based on evidence provided by Wang et al. (2012), which has been replicated for different crops. In those studies that were not based on Wang et al. (2012), very large sample size without a clear definition criterion was adopted (Tan et al., 2016). However, by adopting the sample size suggested by Wang et al. (2012), the model characteristics are ignored, and when using a very large sample size, an unnecessary large computational time required for the analysis.

The parameter range is another source of uncertainty in PBCM when using Bayesian approaches for parameter estimation (Makowski et al., 2006). Depending on the range of parameters, it is possible to generate calibrations that do not represent the desired genotype (He et al., 2010, 2009; Marin et al., 2017; Sexton et al., 2016) or correctly quantify the uncertainty (Dzotsi et al., 2013; Gan et al., 2014; Pereira et al., 2021; Soetaert and Petzoldt, 2010; Zhang et al., 2020b). Wang et al. (2012) compared a range of parameters measured for maize and a relative range of 10% in relation to a reference calibration and found important differences in the GSA results. Li et al., (2019) evaluated different relative ranges, from 10 to 50%, in relation to a reference value and concluded that the most important parameters for the 10% range diverged from those obtained using the 50% range. Many recent PBCM studies (Jin et al., 2018; LI et al., 2019; Tan et al., 2016; Vazquez-Cruz et al., 2014) have adopted relative parameter ranges to apply a GSA, which can result in serious methodological errors. According to (Homma and Saltelli, 1996), the GSA principle is to identify the parameters that cause the greatest uncertainty in the model, and this is not possible when all parameters are disturbed to create relative parameters range. Finally, there is still the model complexity, which is a source of considerable uncertainty and a complicated issue to be considered in the GSA (Razavi and Gupta, 2015).

The SAMUCA crop model has been relatively little evaluated compared to well-established models such as DSSAT/CANEGRO and APSIM-Sugar (Marin et al., 2014; Marin et al., 2015; Thorburn et al., 2005, Sexton et al., 2017). Pereira et al. (2021) used the PRCC to perform a GSA to identify the most important parameters for SAMUCA and then to explore the model uncertainty. However, some issues remained unclear due to the following limitations: (i) the GSA was performed only for the end-of-season output values; (ii) the PRCC method is limited when the parameter relation is not monotonic, and only 4 out of the 24 genotype parameters are monotonic in SAMUCA, which means that some parameters responses might be neglected during the GSA; (iii) authors did not consider the effect of the range of parameters on GSA results. The first two limitations may result in the omission of important parameters because they affect certain variables at a different time of simulation (Dejonge et al., 2012; Lamboni et al., 2009), and the third has never been investigated for sugarcane crop models, and even for other crops the range of parameters effect on GSA was little studied (Wang et al., 2012).

Considering the aspects mentioned above, this paper aimed to: (i) determine the optimal sample size for the eFAST method; (ii) Investigate whether the range of parameters used in GSA affects the; (iii) identify which parameters are responsible for the greatest uncertainty in the SAMUCA model.

3.2. Material and Methods

3.2.1. SAMUCA model

The SAMUCA model is a mechanistic crop growth model. The motivation for the development of this model came from the argument of Sinclair and Seligman (Sinclair and Seligman, 1996), that research groups must develop their models, as this way it is possible to deepen the process in the simulation process and as uncertainties inherent to the models used. Thus, the SAMUCA model was mainly developed by Marin and Jones (Marin and Jones, 2014) and Marin et al. (2017), using a large database for different Brazilian production conditions. Subsequently, Vianna et al. (2020) improved the model structure by decreasing the uncertainty in the soil water balance and including the effect of straw cover on sugarcane growth and development, modifying routines of soil moisture and the flow of water and heat from soil, compared to the previous versions. The most recent version is also included in the DSSAT platform v.4.8.

3.2.2. Data and management

The crop growth simulation scenario was based on a field experiment conducted in the College of Agriculture "Luiz de Queiroz", Piracicaba, São Paulo (Lat: 22º41'55"S, Lon: 47º38'34"W, Alt: 540 m). The sugarcane cultivar was the RB86-7515, a widely used genotype in Brazil (ca. 30% of Brazil's planted area). It was planted on October 16, 2012, with a row spacing of 1.4 m and depth of 0.2 m. A bare soil treatment (Bare) was conducted during the four sequential years, whereas the green cane trash blanket (GCTB) treatment onset in the first ratoon (Oct-2013) and was carried out for 3 years (Table 9). Agricultural practices were adopted to represent high yield farming systems and to ensure the crop was free from pests, diseases, and nutritional stress. The climate is characterized by hot and humid summers with dry winters (Cwa-Köppen classification), and the soil classified as Typic Hapludox. The experiment was irrigated by a center-pivot, based on monitoring the soil moisture by Frequency Domain Reflectometry (FDR) and the evapotranspiration by Bowen Ratio Method (BRM) in both treatments.

Table 9. Description of seasons, planting and harvesting dates, duration in days, treatments, and measurements variables of the field experiment in Piracicaba, Brazil.

Season	Planting	Harvest	Duration	Variables	Treatments
Plant Cane	10/16/2012	10/15/2013	364	SDM,SFM,TIL,LAI *and* POL	Bare
1st Ratoon	10/15/2013	07/15/2014	273	SDM,SFM,TIL,LAI *and* POL	Bare and GCTB
2nd Ratoon	07/15/2014	06/08/2015	328	SDM,SFM,TIL,LAI *and* POL	Bare and GCTB
3rd Ratoon	06/08/2015	06/08/2016	365	SDM,SFM,TIL,LAI and POL	Bare and GCTB

Green cane trash blanket (GCTB), stalk dry mass (SDM) and stalk fresh (SFM) of, leaf area index (LAI), sucrose concentration in fresh matter (POL) and tillering (TIL).

As the eFAST method requires a high computational time, we divided our study into two steps. In the first step (STp1) we simulate sugarcane plant cane (bare soil) and first ratoon with GCTB and bare, testing different sample sizes and two sets of parameter ranges (Table 10). The main objective of STp1 was to define what minimum sample size is needed to obtain a reliable GSA and then apply them to the second step (STp2) of the study. In the STp1 we ran the GSA using as reference the end-of-cycle values of the variables: stalk dry mass (SDM), stalk fresh mass (SFM), leaf area index (LAI), sucrose concentration in the fresh matter (POL). In the STp2, after defining the most adequate sample size, we performed a long simulation considering the different ratoons and ran the GSA in function of the daily values simulated in the whole season of each variable (SDM, SFM, POL, LAI, and tillering (TIL)).

Table 10. Description of the processes performed in the first step (STp1) and the second step (STp2); the simulated season, the sample size evaluated, soil cover type, number of repetitions (NR), and parameter range set (PRS), as described in Table 11 and section 3.2.6.

Step	Season	Sample Size	Treatments	NR	PRS
STp 1	Plant Cane1st Ratoon	65, 129, 257, 513, 1025, 2049, 4097	Bare	10	PRS 1 and PRS 2
STp 2	Plant Cane to 3rd Ratoon	2049	Bare and GCTB	1	PRS 1 and PRS 2

3.2.3. Sensitivity analysis

3.2.4. Extended Fourier amplitude sensitivity test

The eFAST is an algorithm that combines two GSA methods: the Fourier Amplitude Sensitivity Test (FAST) and the Sobol (Saltelli et al., 1999; Saltelli et al., 2010), which in turn, use the model output variance principle. While FAST can scan the entire parameter space and obtain quantitative sensitivity measures in terms of the main sensitivity index (S_i) of each parameter to output variance, the Sobol calculates the total sensitivity index (ST_i) and provides an indication of the overall effect of a given parameter, considering all possible interactions of that parameter with others (Sobol, 2001). The method is based on the decomposition of the model's output variance, determining which fraction of the variance can be explained by the variation in each input parameter. This variation is quantified using the statistical notion of variance (analogous to ANOVA):

$$\sigma^2 = \sum_{i=1}^{N} \frac{(y_i - \underline{y})^2}{(N-1)} \quad (8)$$

where N is number of models runs, y_i is i_{th} model output, and $\underline{y}$ sample mean. Partitioning of variance in eFAST works by varying different parameters at different frequencies, encoding the identity of parameters in the frequency of their variation. In recent years, due to these advantageous properties, eFAST has become more popular in hydrological, ecological, and agronomy modeling (Li et al., 2019; Reusser et al., 2011; Varella et al., 2010b; Xing et al., 2017). We implemented the SAMUCA model in the sensitivity R-package available at: https://cran.r-project.org/web/packages/sensitivity/index.html for applied the method eFAST.

The main sensitivity index (S_i) of a given parameter (i) is calculated as the variance at a particular parameter's unique frequency (and harmonics of that frequency) divided by total variance (VAR_t). First, variance (VAR_i) is calculated from the Fourier coefficients at the frequency of interest (j):

$$VAR_i = 2(A_j^2 + B_j^2) \quad (9)$$

$$A_j = \int_{-\pi}^{\pi} f(s)\, cos\,(js)\, ds \quad (10)$$

$$B_j = \int_{-\pi}^{\pi} f(s)\, sin\,(js)\, ds \quad (11)$$

where s is a scalar variable within the range $-\infty < s < +\infty$; A_j and B_j are the Fourier coefficients (or Fourier a amplitude) over the domain of integer frequencies $j \in \{-\infty, \dots, -1, 0, 1, \dots, \infty\}$. Thus, the S_i is calculated as a fraction VAR_t:

$$S_i = \frac{VAR_i}{VAR_t} \quad (12)$$

the S_i represents the fraction of the output variance of the model explained by the input variation of a given parameter. The ST_i is calculated as the remaining variance after the complementary set contribution is removed. Thus, to estimate ST_i for the given parameter i, the eFAST algorithm first calculates the sensitivity indices except for parameter i using the identification frequencies.

$$ST_i = \frac{VAR_t - VAR_{-i}}{VAR_t} \quad (13)$$

where VAR_{-i} is the sum of all the variance terms that do not include the parameters i.

S_i and ST_i must vary between 0 and 1, where the effects are greater when the indices reach values to 1 whereas values close to 0 indicate negligible effects. The ST_i considers both the S_i and the interactions between the parameters, such interactions that can therefore be evaluated by the difference between the ST_i and the S_i. The two sensitivity indices S_i and ST_i are equal if the effect of the parameter i on the model output is independent of the values of the other parameters: in this case, there is no interaction between this parameter and the others, and the model is additive into parameter i. Only parameters that had S_i >0.05 and ST_i > 0.1were considered significant and relevant in the GSA, as Dejonge et a., (2012) and Xing et al., (2017).

3.2.5. Sample size

The sampling technique is a key factor to explore the domain of interest, being the sample size (SZ) defined as the number of evaluations of the model. In some cases, many evaluations of the model are required, and this can restrict the use of the method. Thus, the relationship between sample size and the convergence of the sensitivity measure is of utmost importance. In this sense, to investigate the dimension of the SZ in the convergence of the sensitivity indices, a sensitivity analysis was performed with different SZ. For this purpose, seven cases of SZ were used: 65, 129, 257, 513, 1025, 2049, and 4097. We performed 10 repetitions for each SZ considering 1st ratoon for the Bare and GCTB treatments. We adopted two criteria to define the most adequate sample size, the first was to calculate the sum of S_i and observe if it converges to 1; if it did not meet this criterion, the sample size would be discarded. The second was to calculate the mean and standard deviation of S_i based on the 10 repetitions. Thus, we determined the smallest sample size that was suitable, where the standard deviation of Si was small enough not to change the order of importance of the parameters. In this analysis, we consider two ranges of variation of parameters, such ranges are described in the following item (3.2.6).

3.2.6. Parameters range set

The GSA is affected by the uncertainty range of the parameters (Wang et al. 2013), so we investigated this factor in the SAMUCA model by constructing two ranges of genotype parameters. The first interval set (PRS1) was constructed based on the literature, containing the maximum and minimum values of each genotype parameter, regardless of the sugarcane cultivar. For the second set of range parameters (PRS2) we considered the studies of Wang et al., (2013), Zhen et al., (2019), and Jin et al., (2018), and concluded that the order of importance of the parameters converges to above 10% disturbance. Thus, in order not to use an excessively small disturbance that would cause inconsistency in the GSA results, and to avoid an excessively large disturbance to generate parameters values outside their genotype reality, we choose to cause a ±25% perturbation in the values calibrated by Vianna et al. (2020). In Table 11 we presented the description of the parameters, and the reference used to elaborate the PRS1. In sequence, the values of the range parameters meet in Table 12, with the respective coefficient of variation (CV).

Table 11. Description of parameters and reference of range parameters used for the PRS1 construction; the values of parameters are shown in Table 4.

Parameters	Description	Reference
amax	Assimilation rate at light saturation point (*µmol. m^{-2}. s^{-1}*)	Sage et al., (2013)
chudec	Heat units for start of tiller abortion *(°C. d)*	Liu et al., (1998)
chumat	Heat units for population establishment *(°C. d)*	Zhou and Shoko, (2011)/Marin and Jones, (2014)
chupeak	Heat units for population peak *(°C. d)*	Coelho et al., (2020); Marin et al., (2017) Nassif et al., 2012)
chustk	Heat units for start culm elongation *(°C. d)*	Marin et al., (2017); /Singels and Bezuidenhout, (2002)
eff	Carboxylation efficiency *(µmol. m^{-2}. s^{-1}/ µmol.m^{-2}. s^{-1})*	Sage et al., (2013)
end_tt_it_gro	Thermal time for completion of internode growth *(°C. d)*	Lingle, (1999)
end_tt_lf_gro	Thermal time for completion of leaf growth *(°C. d)*	Smit and Singels, (2006)
init_lf_area	Initial leaf area of first appeared leaf *(cm^2)*	Zhou et al., (2003)
max_ini_la	Initial leaf area of leaves appeared after top parts formation *(cm^2)*	Zhou et al., (2003)
max_it_dw	Maximum dry biomass of internodes *(g)*	Lingle, (1999)
maxdgl	Maximum number of developed green leaf a tiller can hold *(# . $tiller^{-1}$)*	Vianna et al., (2020)
maxgl	Maximum number of green leaf a tiller can hold *(# . $tiller^{-1}$)*	Marin et al., (2015)
mid_tt_it_gro	Thermal time where internodes can achieve half of its maximum biomass *(°C. d)*	Lingle, (1999)
mid_tt_lf_gro	Thermal time where leaves can achieve half of its maximum biomass *(°C. d)*	Smit and Singels, (2006)
mla	Maximum leaf area *(cm^2)*	Marin et al. (2014)
n_lf_it_from	Number of leaves appeared before internode formation *(# . $tiller^{-1}$)*	Vianna et al., (2020)
n_lf_stk_em	Number of leaves appeared before stalks emerges at soil surface *(# . $tiller^{-1}$)*	Vianna et al., (2020)
phyllochron	Phyllochron interval for leaf appearance *(°C. d)*	Marin et al., (2015)/Inman-Bamber, 1994
plastochron	Thermal time required for the appearance of phytometer *(°C. d)*	Marin et al., (2015)/Inman-Bamber, 1994
popmat	Number of tillers on maturation *(tiller. m^{-2})*	Marin and Jones, (2014)
poppeak	Maximum number of tillers *(tiller. m^{-2})*	Marin et al., (2015)
sla	Specific leaf area *(cm^2.g^{-1})*	Ehara et al., (1994)
tillochron	Thermal time required for emergence of new tiller *(°C. d)*	Bezuidenhout, (2000); Zhou and Shoko, (2011)

Table 12. Calibrated values for genotype RB867515 (μ), first set of parameters (PRS1) based on literature and second set of parameters (PRS2) perturbation ± 25% in relation to μ; CV is the coefficient of variation with respect to μ.

Parameters	μ	PRS1			PRS2		
		Min	Max	CV (%)	Min	Max	CV (%)
amax	44.9	41.3	60.7	22%	33.7	56.1	25%
chudec	1600	1200	1800	19%	1200	2000	25%
chumat	1600	1500	2850	42%	1200	2000	25%
chupeak	1400	404	1950	55%	1050	1750	25%
chustk	650	400	1050	50%	488	813	25%
eff	0.069	0.04	0.08	29%	0.05	0.09	25%
end_tt_it_gro	1200	800	1400	25%	900	1500	25%
end_tt_lf_gro	1300	1100	1500	15%	975	1625	25%
init_lf_area	15	10	30	67%	11	19	25%
max_ini_la	120	80	180	42%	90	150	25%
max_it_dw	28	18	35	30%	21	35	25%
maxdgl	6	6	12	50%	5	8	25%
maxgl	11	10	12	9%	8	14	25%
mid_tt_it_gro	400	380	600	28%	300	500	25%
mid_tt_lf_gro	700	400	800	29%	525	875	25%
mla	600	450	800	29%	450	750	25%
n_lf_it_from	3	2	6	67%	2	4	25%
n_lf_stk_em	4	3	8	63%	3	5	25%
phyllochron	132	107	169	23%	99	165	25%
plastochron	132	107	169	23%	99	165	25%
popmat	9.5	8	12	21%	7	12	25%
poppeak	27	17	30	24%	20	34	25%
sla	120	100	121	9%	90	150	25%
tillochron	69	48.1	134.8	63%	52	86	25%

3.3. Results

3.3.1. Sample size

The sample sizes of 65 and 127 showed relatively high variability, both in the sensitivity indices and in the order of the main parameters explaining the model variance (Figure 13 and 14). For the lower sample sizes (65 and 127) evaluated, it was not possible to obtain sensitivity indices and so accurately quantify the order of importance of the parameters because the sum of the sensitivity indices, S_i and ST_i, diverged from 1. In this case, for these sample sizes, the eFAST method was not able to quantify the sensitivity indices (Tables 13 and 14). For the sample size of 257, however, the minimum sample size required varied as a function of the output variable and the set PRS1 and PRS2 used.

We observed that the convergence of the sensitivity index varied according to the output variable in function of the sample size available, and for the variables TIL and SDM, PRS1 and PRS2 influenced the sample size (Tables 13 and 14). In the case of PRS1, for TIL, the minimum required sample size was 257 and for SDM, it was 1025, what were smaller than the 513 and 2049 obtained in PRS2 for TIL and SDM, respectively (Table 13 and 14; Figure 5 and 6). For POL and LAI, the order of importance of parameters is no longer affected after 513 onwards for PRS1 (Table 13). However, in PRS2, for POL, the order of importance of the parameters was not changed after the sample size of 513 onwards. For LAI, however, the order of importance only stabilized at the sample size of 2049 and 4097 (Table 14). The SDM and SFM variables required larger sample sizes, and regardless of the parameter range, the most appropriate minimum sample size for the application of eFAST was 2049 for the variables considered in this analysis.

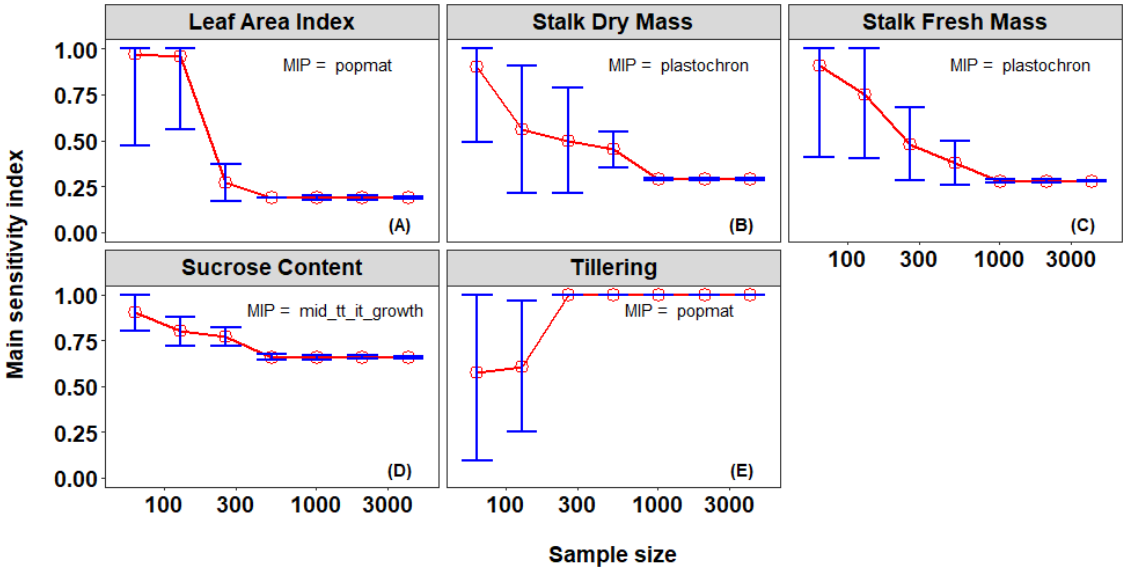


Figure 5. Evolution of sensitivity index of the most important parameter (MIP) with increasing sample size for variables leaf area index (A), mass stalk dry (B) and fresh (C), sucrose content (D), and tillering (E) for PRS1; red line is the average of the 10 simulations for each sample size and blue represents the max and min S_i of each sample size.

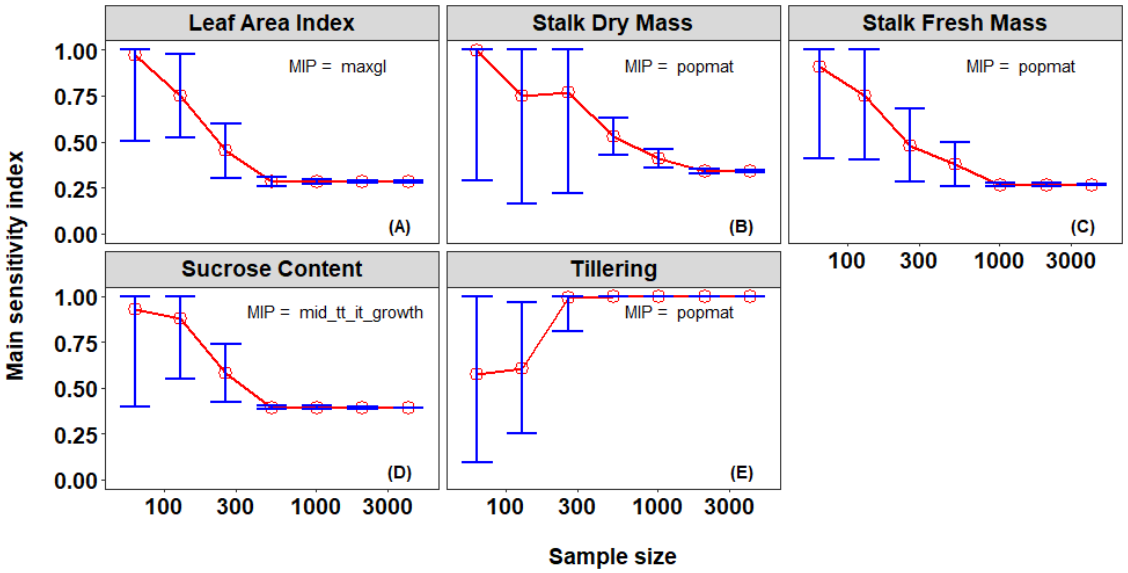


Figure 6. Evolution of sensitivity index of the most important parameter (MIP) with increasing sample size for variables leaf area index (A), mass stalk dry (B) and fresh (C), sucrose content (D), and tillering (E) for PRS2. red line is the average of the 10 simulations for each sample size and blue represents the max and min S_i of each sample size.

Table 13. The order of importance of the genotype parameters for the PRS1 interval, considering the main sensitivity index (S_i) and treatment Bare.

Variable	Rank	Sample Size						
		65	129	257	513	1025	2049	4097
SDM	1	-	-	-	-	*plastochron*	*plastochron*	*plastochron*
	2	-	-	-	-	*max_it_dw*	*max_it_dw*	*max_it_dw*
	3	-	-	-	-	*n_lf_it_from*	*n_lf_it_from*	*n_lf_it_from*
	4	-	-	-	-	*eff*	*eff*	*eff*
	5	-	-	-	-	*popmat*	*popmat*	*popmat*
	6	-	-	-	-	*n_lf_stk_em*	*n_lf_stk_em*	*n_lf_stk_em*
SFM	1	-	-	-	-	*plastochron*	*plastochron*	*plastochron*
	2	-	-	-	-	*mid_tt_it_gro*	*mid_tt_it_gro*	*mid_tt_it_gro*
	3	-	-	-	-	*max_it_dw*	*max_it_dw*	*max_it_dw*
	4	-	-	-	-	*end_tt_it_gro*	*end_tt_it_gro*	*end_tt_it_gro*
	5	-	-	-	-	*eff*	*eff*	*eff*
	6	-	-	-	-	*n_lf_it_from*	*n_lf_it_from*	*n_lf_it_from*
POL	1	-	-	-	*mid_tt_it_gro*	*mid_tt_it_gro*	*mid_tt_it_gro*	*mid_tt_it_gro*
	2	-	-	-	*end_tt_it_gro*	*end_tt_it_gro*	*end_tt_it_gro*	*end_tt_it_gro*
TIL	1	-	-	*popmat*	*popmat*	*popmat*	*popmat*	*popmat*
LAI	1	-	-	-	*popmat*	*popmat*	*popmat*	*popmat*
	2	-	-	-	*mla*	*mla*	*mla*	*mla*

Table 14. The order of importance of the genotype parameters for the PRS2 interval, considering the main sensitivity index (S_i) and treatment Bare.

Variable	Rank	Sample Size						
		65	129	257	513	1025	2049	4097
SDM	1	-	-	-	-	-	*popmat*	*popmat*
	2	-	-	-	-	-	*plastochron*	*plastochron*
	3	-	-	-	-	-	*max_it_dw*	*max_it_dw*
	4	-	-	-	-	-	*eff*	*eff*
SFM	1	-	-	-	-	*popmat*	*popmat*	*popmat*
	2	-	-	-	-	*plastochron*	*plastochron*	*plastochron*
	3	-	-	-	-	*max_it_dw*	*max_it_dw*	*max_it_dw*
	4	-	-	-	-	*mid_tt_it_gro*	*mid_tt_it_gro*	*mid_tt_it_gro*
	5	-	-	-	-	*eff*	*eff*	*eff*
	6	-	-	-	-	*end_tt_it_gro*	*end_tt_it_gro*	*end_tt_it_gro*
POL	1	-	-	-	*mid_tt_it_gro*	*mid_tt_it_gro*	*mid_tt_it_gro*	*mid_tt_it_gro*
	2	-	-	-	*end_tt_it_gro*	*end_tt_it_gro*	*end_tt_it_gro*	*end_tt_it_gro*
TIL	1	-	-	-	*popmat*	*popmat*	*popmat*	*popmat*
LAI	1	-	-	-	*maxgl*	*maxgl*	*maxgl*	*maxgl*
	2	-	-	-	*popmat*	*popmat*	*popmat*	*popmat*
	3	-	-	-	*mid_tt_lf_gro*	*mid_tt_lf_gro*	*mid_tt_lf_gro*	*mid_tt_lf_gro*

3.3.2. Crop features

We consider two key points to evaluate; (i) the order of importance of the parameters and (ii) the values of S_i and ST_i in both parameter ranges (PRS1 and PRS2). We noticed that GCTB only affects the order of the main parameter only in PRS2, while S_i and ST_i values were slightly affected between GCTB and Bare treatments.

In PRS1, the results between Bare and GCTB were similar since the main parameter for all variables was maintained regardless of the presence or absence of GCTB (Figure 7). For example, the main parameter for SDM and SFM was the *plastochron*, and it was responsible for explaining 29% and 31.6% of the variance in the Bare and GCTB treatments, respectively; for SFM the explained variance was 27.9% and 30.4% for Bare and GCTB, respectively (Table 15 and 16). The GCTB only influenced the order of importance of the parameters of the variable SDM, which in the Bare treatment the parameter *n_lf_stk_em* was the 6th parameter (explaining 5.4% of the variance) and in the treatment with GCTB it was the 3rd (explaining 12.7% of the variance) (Table 15 and 16). Furthermore, in the SFM variable, the *plastochron* parameter appears to have greater relevance in the presence of GCTB compared to Bare, as the percentage explained only by it increased, decreasing the values of the parameters *mid_tt_it_gro* and *max_it_dw* (Tables 15 and 16).

In PRS2, the presence of GCTB affected the results in relation to the analysis for the variables SDM, SFM and LAI, while for TIL and POL there was no influence of GCTB (Figure 7). In relation to LAI, the presence of GCTB implied the inclusion of the parameter *mla* between the significant parameters, that is, going from 3 parameters in Bare (*maxgl, popmat* and *mid_tt_lf_gro*) to 4 parameters in treatment GCTB (*maxgl, popmat, mid_tt_lf_gro* and *mla*) (Table 15 and 16). The SDM main parameter in the Bare treatment was the *popmat* (explaining 33.9% of the variance), while in the GCTB treatment the main parameter was the *plastochron* (explaining 29.0% of the variance) (Table 15 and 16). Thus, like in SFM variable, where the main parameter was *popmat* in bare treatment (26.6% of the variance) and *plastochron* in GCTB treatment (30.5% of the variance).

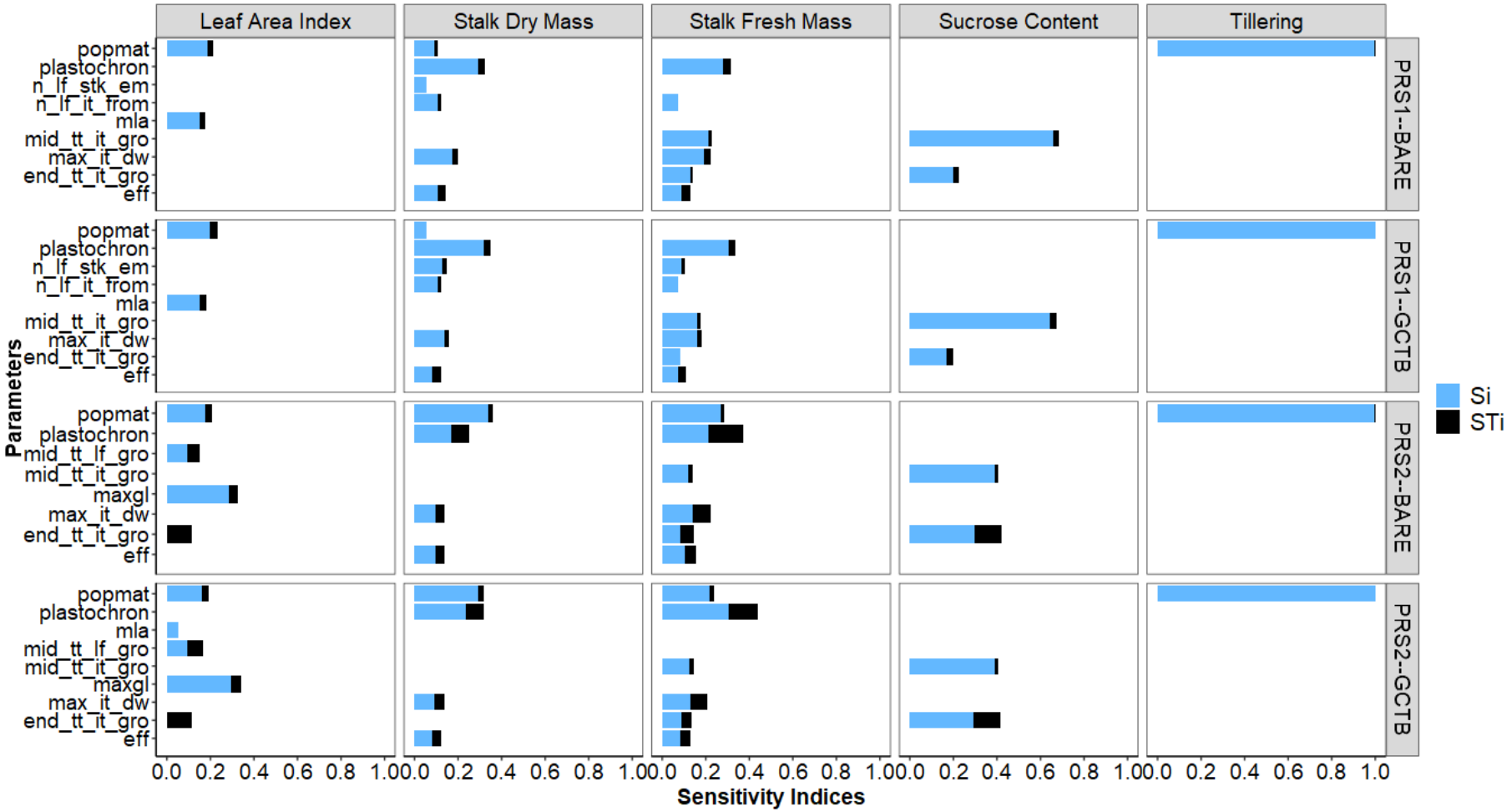


Figure 7. Main sensitivity index (S_i) and total sensitivity index (ST_i) for PRS1 and PRS2 for the sample size of 2049 and bare and GCTB treatments; the sensitivity analysis was calculated for the end-of-season value of each output variable.

3.3.3. Effect of range parameters

For SDM, considering the sample size of 2049, we observed a reduction in the number of significant parameters between PRS1 and PRS2. In PRS1 there were 6 significant parameters (*plastochron, max_it_dw, n_lf_it_from, eff, popmat,* and *n_lf_ stk_em*), which together were responsible for explaining 81.6% (Table 15) and 82.2% (Table 16) of the variance in the bare and GCTB treatments, respectively. In the case of PRS2, only 4 parameters (*popmat, plastochron, max_it_dw*, and *eff*) were significant, responsible for explaining 70% (Table 15) and 70.1% (Table 16) of the variance in the bare and GCTB treatments, respectively.

Among the output variables analyzed, SFM was the one with the highest number of significant parameters, from 6 to 7 parameters (Table 15 and 16). In PRS1 there were 6 significant parameters in the bare treatment (*plastochron, mid_tt_it_gro, max_it_dw, end_tt_it_gro, eff*, and *n_lf_it_form*), which explained 96.8% (Table 15) of the variance, while in the GCTB treatment there were 7 parameters (*plastochron, max_it_dw, mid_tt_it_gro, n_lf_stk_em , end_tt_it_gro, eff,* and *n_lf_it_from*), which explain 93.3% of the variance (Table 16). In the case of PRS2, there were 6 significant parameters in both treatments, responsible for explaining around 93.1% (Table 15) and 95% (Table 16) of the variance in the bare and GCTB treatment, respectively. In the bare treatment, two results should be highlighted: (i) *popmat* was not a significant parameter in PRS1, while in PRS2 it was the most important parameter (Table 15). On the other hand, in the GCTB treatment, popmat remained irrelevant in PRS1 and was the second most important parameter in PRS2; the most important parameter was *plastochron* (Table 16).

The variable LAI in PRS1 presented only two significant parameters (parameters had $S_i > 0.05$) regardless of treatment (Bare or GCTB); the two main parameters were popmat and mla (Tables 15 and 16). However, in PRS2 the number of significant parameters increased, to 3 (*maxgl, popmat*, and *mid_tt_lf_growth*; Table 15) in bare and to 4 (*maxgl, popmat, mid_tt_lf_growth*, and *mla*; Table 16). The biggest discrepancy is in the variance explained by the significant parameters, where in PRS1 they were 33.9% (Table 15) and 34.6% (Table 16), and in PRS2 they were 55.2% (Table 15) and 60% (Table 16).

For variables TIL and POL there was no difference in the order of importance of the parameters in relation to PRS1 and PRS2. For the TIL variable, there was no difference between PRS1 and PRS2, being *popmat* responsible for explaining 99.8% of the variance regardless of treatment. In the case of the POL, there was no change in the group of significant parameters or in their order, with the parameters *mid_tt_it_growth* and *end_tt_it_growth* being the only significant ones, regardless of the treatment and the PRS1 and PRS2 set (Tables 15 and 16). However, we noticed that *mid_tt_it_growth* in PRS1 explained 65.8% (Table 15) and 64.2% (Table 16) of the variance, while in PRS2 it was 39.2% (Table 15 and 16).

Table 15. Relative value of the model output variance (σ^2) explained individually by each parameter, and the variance sum ($\sum \sigma^2$) of the parameters; we only considered the parameters that presented S_i >0.05 and sample size of 2049 in treatment Bare.

Variable	PRS1				PRS2			
	Parameters	σ^2	Rank	$\sum \sigma^2$	Parameters	σ^2	Rank	$\sum \sigma^2$
SDM	*plastochron*	29.0%	1°	29.0%	*popmat*	33.9%	1°	33.9%
	max_it_dw	17.2%	2°	46.2%	*plastochron*	16.7%	2°	50.6%
	n_lf_it_from	10.6%	3°	56.8%	*max_it_dw*	9.8%	3°	60.4%
	eff	10.5%	4°	67.3%	*eff*	9.6%	4°	70.0%
	popmat	8.9%	5°	76.2%	-	-	-	-
	n_lf_stk_em	5.4%	6°	81.6%	-	-	-	-
SFM	*plastochron*	27.9%	1°	27.9%	*popmat*	26.6%	1°	26.6%
	mid_tt_it_gro	21.1%	2°	49.0%	*plastochron*	21.3%	2°	47.9%
	max_it_dw	18.9%	3°	67.9%	*max_it_dw*	14.2%	3°	62.1%
	end_tt_it_gro	12.7%	4°	80.6%	*mid_tt_it_gro*	11.9%	4°	74.0%
	eff	9.0%	5°	89.6%	*eff*	10.6%	5°	84.6%
	n_lf_it_from	7.1%	6°	96.8%	*end_tt_it_gro*	8.5%	6°	93.1%
POL	*mid_tt_it_gro*	65.8%	1°	65.8%	*mid_tt_it_gro*	39.2%	1°	39.2%
	end_tt_it_gro	20.1%	2°	85.9%	*end_tt_it_gro*	29.8%	2°	69.1%
TIL	*popmat*	99.8%	1°	99.8%	*popmat*	99.8%	1°	99.8%
LAI	*popmat*	18.8%	1°	18.8%	*maxgl*	28.3%	1°	28.3%
	mla	15.1%	2°	33.9%	*popmat*	17.5%	2°	45.8%
	-	-	-	-	*mid_tt_lf_gro*	9.4%	3°	55.2%

PRS1 and PRS2 are the different sets of tested parameter ranges

Table 16. Relative value of the model output variance (σ^2) explained individually by each parameter, and the variance sum ($\sum \sigma^2$) of the parameters; we only considered the parameters that presented S_i >0.05 and sample size of 2049 in treatment GCTB.

Variable	PRS1				PRS2			
	Parameters	σ^2	Rank	$\sum \sigma^2$	Parameters	σ^2	Rank	$\sum \sigma^2$
SDM	*plastochron*	31.6%	1°	31.6%	*popmat*	29.0%	1°	29.0%
	max_it_dw	14.0%	2°	45.5%	*plastochron*	23.6%	2°	52.6%
	n_lf_stk_em	12.7%	3°	58.3%	*max_it_dw*	9.5%	3°	62.1%
	n_lf_it_from	10.4%	4°	68.7%	*eff*	8.0%	4°	70.1%
	eff	8.0%	5°	76.7%	-	-	-	-
	popmat	5.5%	6°	82.2%	-	-	-	-
SFM	*plastochron*	30.4%	1°	30.4%	*plastochron*	30.5%	1°	30.5%
	max_it_dw	15.9%	2°	46.3%	*popmat*	21.8%	2°	52.3%
	mid_tt_it_gro	15.9%	3°	62.2%	*max_it_dw*	13.3%	3°	65.6%
	n_lf_stk_em	8.7%	4°	70.9%	*mid_tt_it_gro*	12.3%	4°	77.9%
	end_tt_it_gro	8.2%	5°	79.1%	*end_tt_it_gro*	8.8%	5°	86.7%
	eff	7.2%	6°	86.2%	*eff*	8.3%	6°	95.0%
	n_lf_it_from	7.1%	7°	93.3%	-	-	-	-
POL	*mid_tt_it_gro*	64.2%	1°	64.2%	*mid_tt_it_gro*	39.2%	1°	39.2%
	end_tt_it_gro	16.7%	2°	80.9%	*end_tt_it_gro*	29.3%	2°	68.5%
TIL	*popmat*	99.8%	1°	99.8%	*popmat*	99.8%	1°	99.8%
LAI	*popmat*	19.8%	1°	19.8%	*maxgl*	29.6%	1°	29.6%
	mla	14.8%	2°	34.6%	*popmat*	15.8%	2°	45.5%
	-	-	-	-	*mid_tt_it_gro*	9.5%	3°	55.0%
	-	-	-	-	*mla*	5.0%	4°	60.0%

PRS1 and PRS2 are the different sets of tested parameter ranges

3.3.4. Time-dependent effects on global sensitivity analysis

Having defined the appropriate optimum sample size (2049), we performed an analysis considering a temporal variation in the different stages of the sugarcane crop cycle (plant cane to 3rd ratoon). We observed that the GSA over time provided two important results: (i) the sugarcane season did not affect the order of importance of parameters; (ii) at different times of sugarcane growth, there were different parameters responsible for explaining the greater variance of the model. The order of parameters was not changed compared to ratoons; for example, for SDM we noticed that the same parameters have the same degree of importance in all ratoons (Figure 8). The same pattern was repeated for all variables and their respective parameters. However, the growth period in which a sensitivity analysis was performed affected which parameter explained the greatest model variance.

For the variable SDM, the order of parameters was not altered in relation to ratoons, regardless of the set of parameters (PRS1 and PRS2) and treatment (Bare and GCTB) (Figure 8). However, during the different growth phases, there was a change in the order of the main parameter. For example, during the crop establishment and development phase, the *plastochron* parameter was responsible for explaining more than 50% of the SDM variance, becoming the main parameter in this period when using the PRS2 set. On the other hand, at the end of the maturation phase, the *plastochron* was the second most important parameter, behind the *popmat* with the PRS2 set (Figure 8 - PRS2 Bare and GCTB). In PRS1, during the emergence phase until mid-establishment, the main parameter was *n_lf_it_from*, later the *plastochron* became the main parameter until the end of the growth phase (Figure

8 - PRS1 Bare and GCTB). Furthermore, in short periods and isolated in the growth phase, the parameters *n_lf_it_from*, *end_tt_lf_gro* and *n_lf_ stk_em* were significant using the PRS2 set (Figure 8 - PRS2 Bare and GCTB); these three parameters were not reported in Tables 15 and 16 as they were not significant at the end of the growth phase. This pattern was not observed in PRS1 (Figure 8 - PRS1 Bare and GCTB), that is, all significant parameters presented in Tables 15 and 16 appeared in the analysis throughout the growth phases.

For SFM, *plastochron* had a greater impact on the simulation regardless of the parameter set (PRS1 and PRS2) or treatment (GCTB and Bare; Figure 9). Among the sets of parameters, PRS1 and PRS2, there was the inclusion of the *popmat* parameter in PRS1 in the Bare treatment; however, its influence was minimal (5%). For PRS2, the *popmat* remained relevant during most parts of the growth phase, but in our view not enough to have more impact on the SFM simulations. Finally, the *end_tt_lf_gro* parameter was significant at specific moments in PRS2, both in Bare and in GCTB, a situation not observed in PRS1 (Figure 9).

The LAI variable, contrary to what was seen in sections 3.2 and 3.3, where the maximum number of significant parameters was 4 (Table 15 e 16), was influenced by fifteen parameters at different growth phases (Figure 10). However, this large number of parameters is due to one-off events, because they did not present a pattern or a period of influence in sequence during the growth phase. In PRS1, we observed that *tillochron*, *plastochron*, *mla* and *init_leaf_area* parameters had greater relevance, as they explained, individually, more than 40% of the variance of LAI at different moments of emergence phase and of establishment phase (Figure 10; PRS1 Bare and GCTB). In the establishment and maturity stages, the variance of the LAI limited to the *mla* and *popmat* parameters with the PRS1 set and, in our view, the *mla* parameter was the most important parameter as it was significant in all growth phases (Figure 10 - PRS1 Bare and GCTB). In PRS2, the emergence and establishment phases were dominated by *plastochron* and *mla*, and in the development and maturation phases, the parameters *maxgl* and *mid_tt_lf_gro* were responsible for explaining most of the variance (Figure 10 - PRS2 Bare and GCTB). In our view, considering the PRS2 set, the most important parameter was *maxgl* because it explained more than 30% of the variance in almost all growth phases. This result showed the uncertainty present in the range of the chosen parameters, since *maxgl* was not significant at growth phases in the PRS1 set (Figure 10 - PRS1 Bare and GCTB).

For the variable POL, as in LAI, we had the inclusion of several parameters depending on the of the analyzed growth phase (Figure 11). In this case, in PRS1, in addition to *mid_tt_it_gro* and *end_tt_it_gro*, there was the inclusion of *plastochron*, *n_lf_stk_em* and *n_lf_it_from* (Figure 11; PRS1 Bare and GCTB). In PRS2, the parameters *plastochron*, *n_lf_stk_em*, *n_lf_it_from* and *end_tt_lf_gro* were included (Figure 11; PRS2 Bare and GCTB). For PRS1, the parameter *end_tt_it_gro* was not significant in the 1st ratoon (Figure 11; PRS1 Bare and GCTB), evidencing a combination of the range of parameters with the weather conditions for that season. Yet, for PRS2, *end_tt_lf_gro* was significant, while in PRS1 the same was not observed.

The TIL variable had two main parameters, from the beginning to the middle of the cycle, that was *tillochron* and from the middle to the end of the cycle *popmat* (Figure 12). In the first half of the cycle, the *tilochron* influence was around 90%, while in the second half the *popmat* explained more than 90% of the variance. In the transition period between these two phases of the crop cycle, that takes place between the end of the establishment and half of the development of the crop, we observed that there was a punctual influence of some other parameters, such as *mla* and *plastochron*. However, during this transition, there were strong indications of the effect of climate on GSA, as these two parameters (*mla* and *plastochron*) were not in all ratoons. To corroborate this result, we did not observe relevant influences of the parameter range (PRS1 or PRS2) and the evaluated treatment (Bare or GCTB).

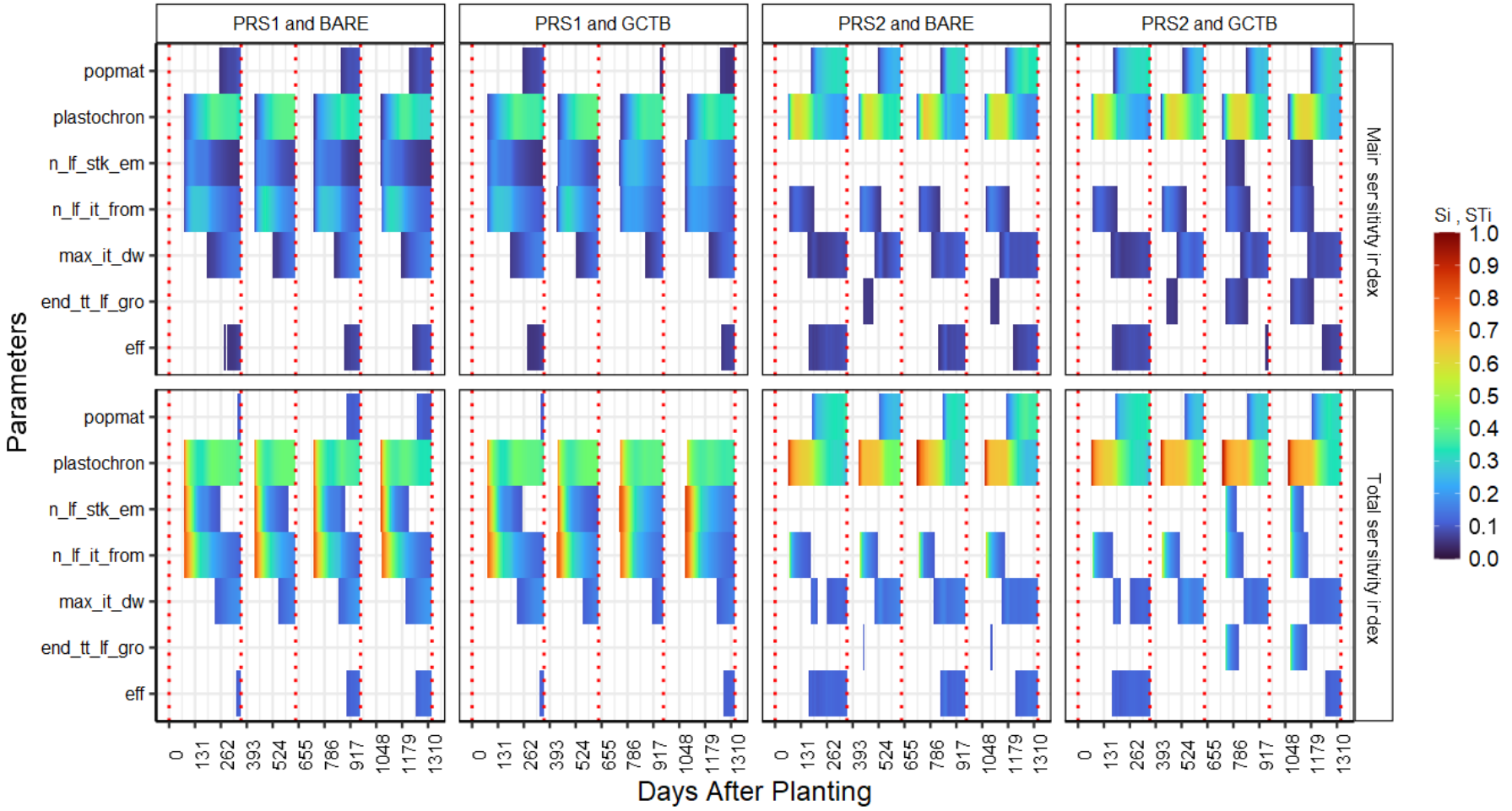


Figure 8. The main sensitivity index (S_i) and total sensitivity index (ST_i) calculated in PRS1 to Plant Cane (PC), 1st,2nd, and 3rd Ratoons for Stalk Dry Mass.

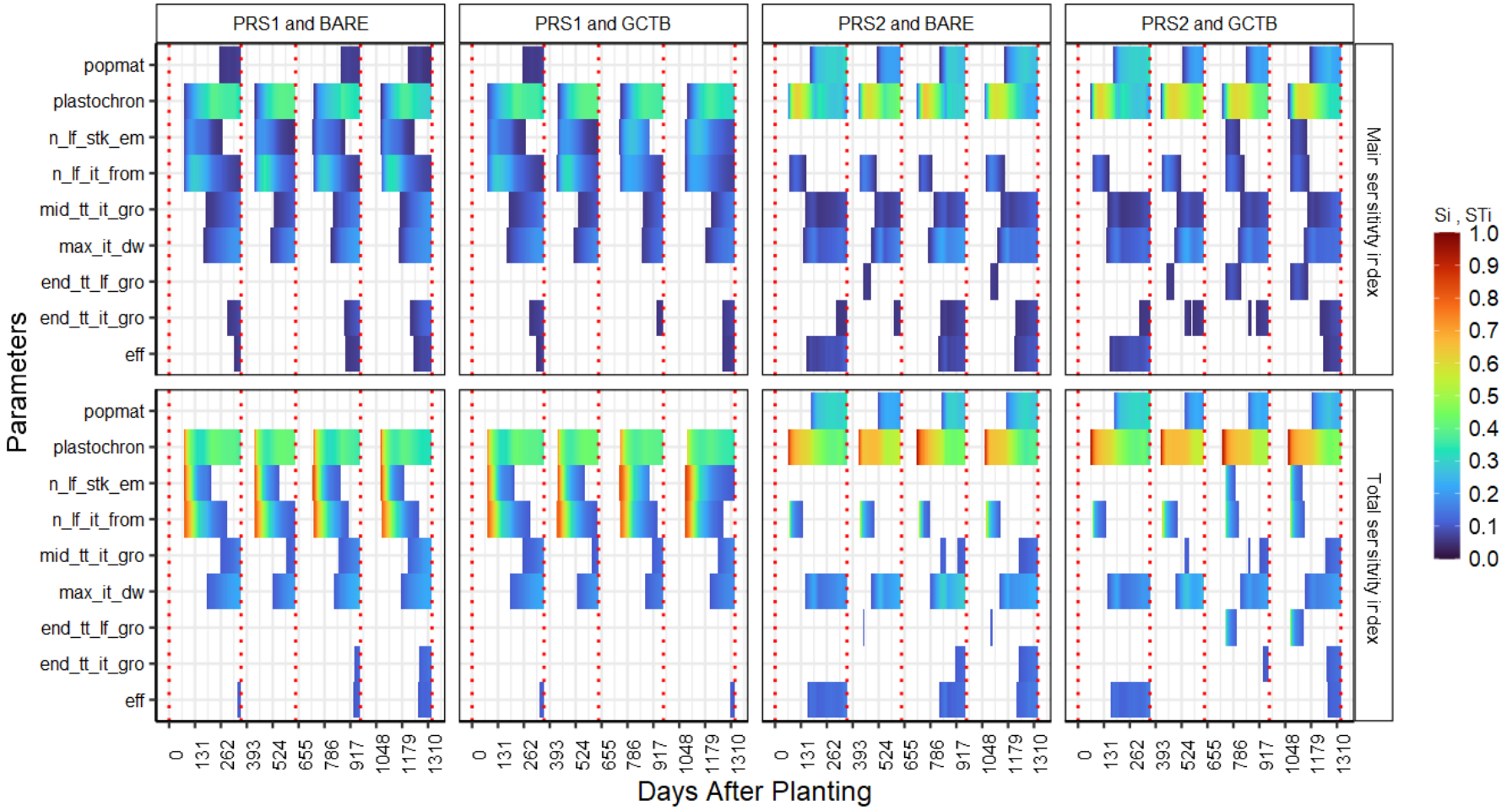


Figure 9. The main sensitivity index (S_i) and total sensitivity index (ST_i) calculated in PRS1 to Plant Cane (PC), 1st,2nd, and 3rd Ratoons for Stalk Fresh Mass.

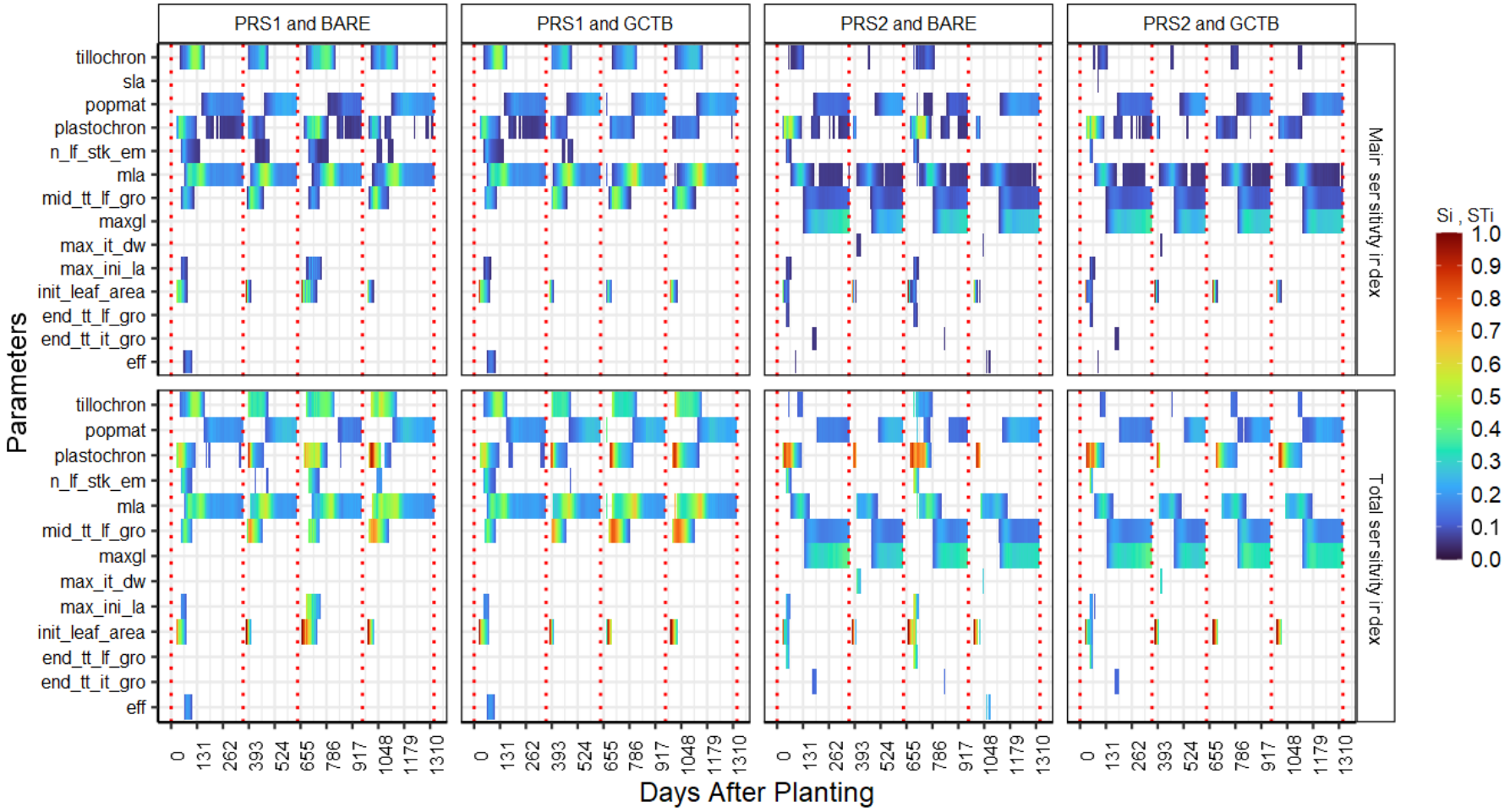


Figure 10. The main sensitivity index (S_i) and total sensitivity index (ST_i) calculated in PRS1 to Plant Cane (PC), 1st,2nd, and 3rd Ratoons for Leaf Are Index

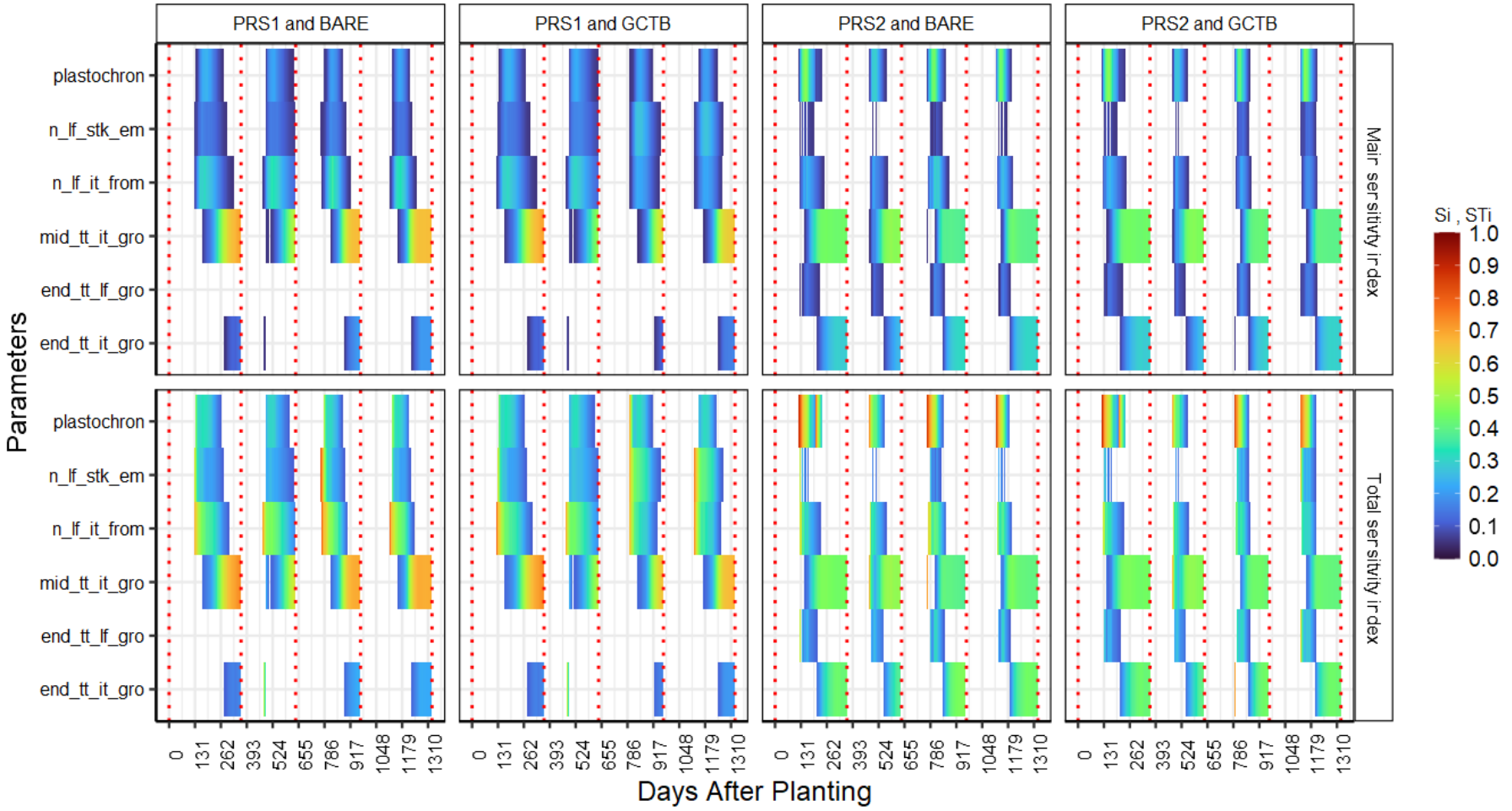


Figure 11. The main sensitivity index (S_i) and total sensitivity index (ST_i) calculated in PRS1 to Plant Cane (PC), 1st,2nd, and 3rd Ratoons for Sucrose content.

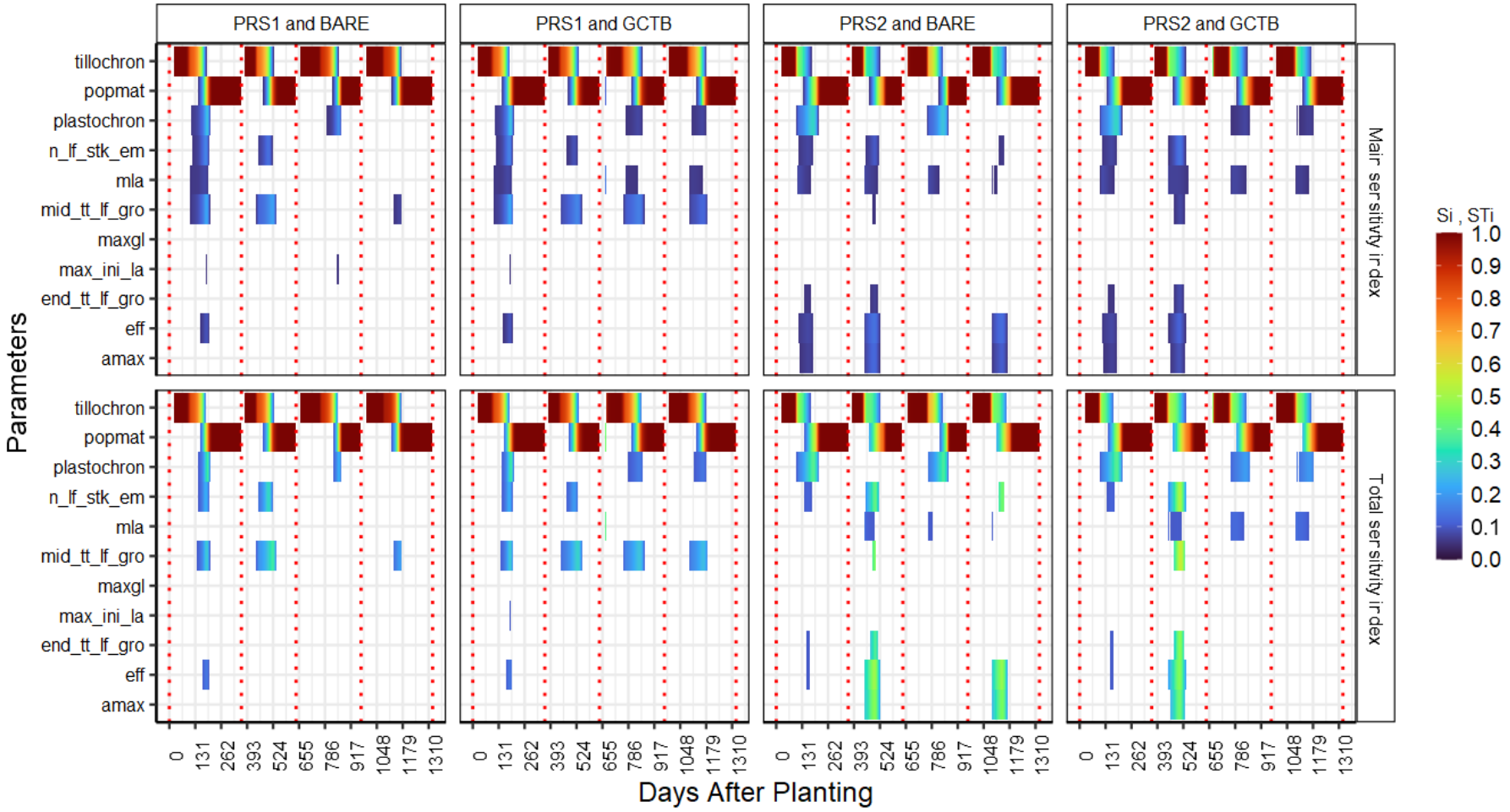


Figure 12. The main sensitivity index (S_i) and total sensitivity index (ST_i) calculated in PRS1 to Plant Cane (PC), 1st,2nd, and 3rd Ratoons for Tillering

3.4. Discussion

We identified that the sample size of 2,049 was the most suitable for application of the eFAST method with the SAMUCA model, regardless of the variable of interest or the parameter sets (PRS1 or PRS2) (Tables 13 and 14). This result agrees with that reported by Wang et al., (2012), where the sample size of 2,049 produced the most stable sensitivity indices. However, the sample size can be different depending on the variable of interest; for instance, POL required a minimum size of 513 to converge (Table 13 and 14), while SDM and SFM require a sample size of at least 2,049 (Table 5 and 6). Furthermore, for the TIL variable, we noticed that the sample size in PRS1 was 257 and in PRS2 it was 513 (Tables 5 and 6). This difference implied that the computational time might be reduced depending on the output variable, thus, the high computational cost, one of the biggest limitations of the GSA (Gilardelli et al., 2018; Jeuffroy et al., 2006; Marino et al., 2008), can be optimized if previous studies that explore the sample size are carried out.

As far as we know, Wang et al., (2013) was the only study using the eFAST method for a PBCM that evaluated different sample sizes and serves as the main reference source for other GSA studies. However, unlike our study, they determined the sample size based on a single output variable. In addition to the output variable, the sample size was also different depending on the set of parameters applied (PRS1 and PRS2; Tables 13 and 14). To our knowledge, there is no available studies on PBCM considering the relationship between the sample size in different analysis variables and the parameter range, and we can now state that sample size in GSA may vary as a function of: (i) the analyzed output variable and (ii) range parameters.

As in many studies on the influence of the environment on sensitivity indices (Dejonge et al. 2012, Sexton et al. 2017, Zhang et al. 2020), there was an effect of GCTB on SAMUCA sensitivity indices, but it was not sufficient to change the order of importance of the parameters and it did not influence the parameter range (PRS1 and PRS2) (Figures 8, 9, 10, 11, and 12). For example, the SFM had the *plastochron* parameter as the main parameter in PRS1 in both treatments (bare and GCTB), and in PRS2, the main parameter was *plastochorn* in Bare treatment and *popmat* in GCTB treatment. To consider that a GCTB was sufficient to change the order of importance of the parameters, an alteration between the parameters should be observed in both sets of parameters (PRS1 and PRS2; Tables 15 and 16). In the case of GCTB, it has a direct influence on the soil heat flux and on the soil water dynamics (Vianna et al., 2020; Pereira et al., 2020), which would hypothetically refer to the lower effect of GCTB on GSA for the variables LAI, POL, SDM, SFM and TIL; if we would evaluate soil temperature or evapotranspiration, we would possibly have a greater impact of GCTB, as it is directly related to these variables. We assume this based on the results of DeJonge al. (2012), who identified that radiation use efficiency (RUE) as the most important parameter for yield, both for irrigated and rainfed treatments. However, for crop evapotranspiration (ETc), in the irrigated environment, the main parameters were related to the crop, while in the rainfed environment the main parameters were related to the soil (Dejong et al., 2012). Thus, the effect of management on the sensitivity analysis is dependent on the variable of interest and did not affect all simulations variables in our study.

The range of parameters was the main source of uncertainty for the sensitivity analysis, changing the order of parameters for the variables SDM, SFM, and LAI. Wang et al., (2020) and Li et. (2019) have already identified that parameter range affected the order and variance explained by each parameter. Many studies have applied parameter ranges with relative values from GSA studies in different PBCM (Jin et al., 2018; Li et al., 2019; Tan et al., 2016; Vazquez-Cruz et al., 2014), which in our view is a mistake, based on the influence of the parameter range. According to Santelli et al., (1999), the principle of the GSA is to quantify the model uncertainty based on

perturbations on the environmental variables. In this sense, the use of relative values in the parameter ranges without experimental foundations may not represent well the influence of the parameters of a PBCM. For example, for the variable LAI, *mla* and *maxgl* were observed as the main parameters in PRS1 and PRS2, respectively. The CV of these parameters was 29% (*mla*) and 9% (*maxgl*), in PRS1, and 25% for both parameters in PRS2; for this configuration of PRS1, *maxgl* was not significant, but it was the most important parameter in PRS2. Yet, this can be related to the correlation that exists between the parameters, neglected in many Bayesians and PBCM applications (Marin et al., 2017; Pereira et al., 2021), which demonstrated that any uncertainty analysis using relative values seems to be inadequate.

In sensitivity analysis, it is necessary to use a time interval for one evaluation, as changes in the order of importance of main parameters occurred during the growth of the crop. For example, considering our final SFM yield value, the parameters *popmat* and *plastochron* accounted for between 21.8% and 30.4% of the variation, respectively (Table 16). However, the *popmat* is less relevant than the *plastochron*, as the *plastochron* influenced a longer period of the cycle, being responsible for explaining up to 60% of the variance in the first 150 days of each ratoon in the PRS2 set (Figure 9, bare and GCTB). For the TIL variable, this was even more evident, having distinct parameters and explaining over 90% of the variation in different times. In addition, the occurrence of significant different parameters between ratoon years indicated a possible influence of climate variation (Figure 12). Several GSA studies showed the effect of the seasonality in sensitivity indices, independent of the method adopted (Attia et al., 2021; Loos and Lemaître, 2015; Li et al., 2019; Sexton et al., 2017; Tan et al., 2017; Tan et al., 2016; Vazquez-Cruz et al., 2014; Xing et al., 2017). In our study, we performed the GSA at a daily time step from plant cane to 3rd ratoon and identified that classification of the importance of parameters between ratoons did not change (Figure 8, 9, 10, 11, and 12). However, there is evidence that the weather conditions interfered with the sensitivity index, as observed by Anderson et al., 2014; Attia et al., 2021; Gilardelli et al., 2018; Sexton et al., 2017. Thus, for sugarcane models, we must use time intervals to perform the GSA and consider that there is evidence of climate influence on the GSA.

3.5. Conclusion

The results showed that the sample size and parameter range were important for GSA, and that a sample size of at least 2049 was required was necessary for the sensitivity indices to converge regardless of the variable. However, some variables could do with smaller sizes, such as the case of TIL that predicted a sample size of 257. The range of parameters must be carefully investigated, and we demonstrate that the use of relative values without biophysical basis to determine the parameter ranges is inappropriate for the uncertainty analysis, and measured thresholds should always be used, even if from different genotypes, to determine the model's response across the full range of parameters. We did not identify the influence of GCTB and the different ratoons on the order of importance of the parameters, they only slightly affected the values (S_i and ST_i) of the sensitivity indices.

References

Anderson, B., Borgonovo, E., Galeotti, M., Roson, R., 2014. Uncertainty in climate change modeling: Can global sensitivity analysis be of help? Risk Analysis 34, 271–293. https://doi.org/10.1111/RISA.12117

Attia, A., El-Hendawy, S., Al-Suhaibani, N., Tahir, M.U., Mubushar, M., Vianna, M. dos S., Ullah, H., Mansour, E., Datta, A., 2021. Sensitivity of the DSSAT model in simulating maize yield and soil carbon dynamics in arid Mediterranean climate: Effect of soil, genotype and crop management. Field Crops Research 260, 107981. https://doi.org/10.1016/j.fcr.2020.107981

Chu-Agor, M.L., Muñoz-Carpena, R., Kiker, G., Emanuelsson, A., Linkov, I., 2011. Exploring vulnerability of coastal habitats to sea level rise through global sensitivity and uncertainty analyses. Environmental Modelling and Software 26, 593–604. https://doi.org/10.1016/j.envsoft.2010.12.003

Confalonieri, R., 2010. Monte Carlo based sensitivity analysis of two crop simulators and considerations on model balance. European Journal of Agronomy 33, 89–93. https://doi.org/10.1016/J.EJA.2010.03.004

Cukier, R.I., Levine, H.B., Shuler, K.E., 1978. Nonlinear sensitivity analysis of multiparameter model systems. Journal of Computational Physics. https://doi.org/10.1016/0021-9991(78)90097-9

Dejonge, K.C., Ascough Ii, J.C., Ahmadi, M., Andales, A.A., Arabi, M., 2012. Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. Ecological Modelling 231, 113–125. https://doi.org/10.1016/j.ecolmodel.2012.01.024

Dzotsi, K.A., Basso, B., Jones, J.W., 2013. Development, uncertainty, and sensitivity analysis of the simple SALUS crop model in DSSAT. Ecological Modelling 260, 62–76. https://doi.org/10.1016/j.ecolmodel.2013.03.017

Everingham, Y., Sexton, J., Skocaj, D., Inman-Bamber, G., 2016. Accurate prediction of sugarcane yield using a random forest algorithm. Agronomy for Sustainable Development 36. https://doi.org/10.1007/s13593-016-0364-z

Everingham, Y.L., Muchow, R.C., Stone, R.C., Inman-Bamber, N.G., Singels, A., Bezuidenhout, C.N., 2002. Enhanced risk management and decision-making capability across the sugarcane industry value chain based on seasonal climate forecasts. Agricultural Systems 74, 459–477. https://doi.org/10.1016/S0308-521X(02)00050-1

Faivre, R., Leenhardt, D., Voltz, M., Benoît, M., Papy, F., Dedieu, G., Wallach, D., 2009. Spatialising Crop Models BT - Sustainable Agriculture, in: Lichtfouse, E., Navarrete, M., Debaeke, P., Véronique, S., Alberola, C. (Eds.), . Springer Netherlands, Dordrecht, pp. 687–705. https://doi.org/10.1007/978-90-481-2666-8_42

Fraedrich, D., Goldberg, A., 2000. Methodological framework for the validation of predictive simulations. European Journal of Operational Research 124, 55–62. https://doi.org/10.1016/S0377-2217(99)00117-4

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. Environmental Modelling and Software 51, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Gilardelli, C., Confalonieri, R., Cappelli, G.A., Bellocchi, G., 2018. Sensitivity of WOFOST-based modelling solutions to crop parameters under climate change. Ecological Modelling 368, 1–14. https://doi.org/10.1016/J.ECOLMODEL.2017.11.003

Guérif, M., Duke, C.L., 2000. Adjustment procedures of a crop model to the site specific characteristics of soil and crop using remote sensing data assimilation. Agriculture, Ecosystems and Environment 81, 57–69. https://doi.org/10.1016/S0167-8809(00)00168-7

He, J., Jones, J.W., Graham, W.D., Dukes, M.D., 2010. Influence of likelihood function choice for estimating crop model parameters using the generalized likelihood uncertainty estimation method. Agricultural Systems 103, 256–264. https://doi.org/10.1016/j.agsy.2010.01.006

He, J., M. D. Dukes, M.D., J. W. Jones, J.W., W. D. Graham, W.D., J. Judge, J., 2009. Applying GLUE for Estimating CERES-Maize Genetic and Soil Parameters for Sweet Corn Production. Trans ASABE 52, 1907–1921. https://doi.org/10.13031/2013.29218

Hirabayashi, S., Kroll, C.N., Nowak, D.J., 2011. Component-based development and sensitivity analyses of an air pollutant dry deposition model. Environmental Modelling and Software 26, 804–816. https://doi.org/10.1016/j.envsoft.2010.11.007

Hoffman, N., Singels, A., Patton, A., Ramburan, S., 2018. Predicting genotypic differences in irrigated sugarcane yield using the Canegro model and independent trait parameter estimates. European Journal of Agronomy 96, 13–21. https://doi.org/10.1016/j.eja.2018.01.005

Homma, T., Saltelli, A., 1996. Importance measures in global sensitivity analysis of nonlinear models. Reliability Engineering and System Sa]ety 52, 17. https://doi.org/https://doi.org/10.1016/0951-8320(96)00002-6

Inman-Bamber, N.G., Smith, D.M., 2005. Water relations in sugarcane and response to water deficits. Field Crops Res 92, 185–202. https://doi.org/https://doi.org/10.1016/j.fcr.2005.01.023

Iooss, B., Lemaître, P., 2015. A Review on Global Sensitivity Analysis Methods. Operations Research/ Computer Science Interfaces Series 59, 101–122. https://doi.org/10.1007/978-1-4899-7547-8_5

Jeuffroy, M.H., Barbottin, A., Jones, J.W., Leoeur, J., 2006. Crop models with genotype parameters, in: Working with Crop Models. pp. 281–308.

Jin, X., Li, Z., Nie, C., Xu, X., Feng, H., Guo, W., Wang, J., 2018. Parameter sensitivity analysis of the AquaCrop model based on extended fourier amplitude sensitivity under different agro-meteorological conditions and application. Field Crops Research 226, 1–15. https://doi.org/10.1016/j.fcr.2018.07.002

Jones, J.W., Antle, J.M., Basso, B., Boote, K.J., Conant, R.T., Foster, I., Godfray, H.C.J., Herrero, M., Howitt, R.E., Janssen, S., Keating, B.A., Munoz-Carpena, R., Porter, C.H., Rosenzweig, C., Wheeler, T.R., 2017. Toward a new generation of agricultural system data, models, and knowledge products: State of agricultural systems science. Agricultural Systems 155, 269–288. https://doi.org/10.1016/j.agsy.2016.09.021

Jones, M.R., Singels, A., 2018. Refining the Canegro model for improved simulation of climate change impacts on sugarcane. European Journal of Agronomy 100, 76–86. https://doi.org/10.1016/j.eja.2017.12.009

Jones, M.R., Singels, A., Ruane, A.C., 2015. Simulated impacts of climate change on water use and yield of irrigated sugarcane in South Africa. Agricultural Systems 139, 260–270. https://doi.org/https://doi.org/10.1016/j.agsy.2015.07.007

Keating, B.A., Robertson, M.J., Muchow, R.C., Huth, N.I., 1999. Modelling sugarcane production systems I. Development and performance of the sugarcane module. Field Crops Research 61, 253–271. https://doi.org/10.1016/S0378-4290(98)00167-1

Krishnan, P., Aggarwal, P., 2018. Global sensitivity and uncertainty analyses of a web-based crop simulation model (web InfoCrop wheat) for soil parameters. Plant Soil 423, 443–463. https://doi.org/10.1007/s11104-017-3498-0

Kroes, J.G., Van Dam, J.C., Groenendijk, P., Hendriks, R.F.A., Jacobs, C.M.J., 2009. SWAP version 3.2. Theory description and user manual. Alterra. https://edepot.wur.nl/176385

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

LI, Z. hai, JIN, X. liang, LIU, H. long, XU, X. gang, WANG, J. hua, 2019. Global sensitivity analysis of wheat grain yield and quality and the related process variables from the DSSAT-CERES model based on the extended Fourier Amplitude Sensitivity Test method. Journal of Integrative Agriculture 18, 1547–1561. https://doi.org/10.1016/S2095-3119(18)62046-5

Li, Z. hai, Jin, X. liang, Liu, H. long, Xu, X. gang, Wang, J. hua, 2019. Global sensitivity analysis of wheat grain yield and quality and the related process variables from the DSSAT-CERES model based on the extended Fourier Amplitude Sensitivity Test method. Journal of Integrative Agriculture 18, 1547–1561. https://doi.org/10.1016/S2095-3119(18)62046-5

Makowski, D., Hillier, J., Wallach, D., Andrieu, B., Jeuffroy, M.-H., 2006. Parameter Estimation for Crop Models. Working with Dynamic Crop Models 55–100.

Makowski, D., Wallach, D., Tremblay, M., 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. Agronomie 22, 191–203. https://doi.org/10.1051/agro:2002007

Marin, F.R., Jones, J.W., 2014. Process-based simple model for simulating sugarcane growth and production. Scientia Agricola 71, 1–16. https://doi.org/10.1590/S0103-90162014000100001

Marin, F.R., Jones, J.W., Boote, K.J., 2017. A stochastic method for crop models: including uncertainty in a sugarcane model. Agronomy Journal 109, 483–495. https://doi.org/https://doi.org/10.2134/agronj2016.02.0103

Marin, F.R., Jones, J.W., Singels, A., Royce, F., Assad, E.D., Pellegrino, G.Q., Justino, F., 2013. Climate change impacts on sugarcane attainable yield in southern Brazil. Climatic Change 117, 227–239. https://doi.org/10.1007/s10584-012-0561-y

Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E., 2008. A methodology for performing global uncertainty and sensitivity analysis in systems biology. Journal of Theoretical Biology 254, 178–196. https://doi.org/10.1016/j.jtbi.2008.04.011

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33, 161. https://doi.org/10.2307/1269043

Pereira, R.A. de A., Vianna, M. dos S., Nassif, D.S.P., Carvalho, K. dos S., Marin, F.R., 2021. Global sensitivity and uncertainty analysis of a sugarcane model considering the trash blanket effect [WWW Document]. European Journal of Agronomy. https://doi.org/https://doi.org/10.1016/j.eja.2021.126371

Raza, G., Ali, K., Hassan, M.A., Ashraf, M., Khan, M.T., Khan, I.A., 2019. Sugarcane as a Bioenergy Source. Sugarcane Biofuels 3–19. https://doi.org/10.1007/978-3-030-18597-8_1

Razavi, S., Gupta, H. v., 2015. What do we mean by sensitivity analysis? The need for comprehensive characterization of "global" sensitivity in Earth and Environmental systems models. Water Resources Research 51, 3070–3092. https://doi.org/10.1002/2014WR016527

Reusser, D.E., Buytaert, W., Zehe, E., 2011. Temporal dynamics of model parameter sensitivity for computationally expensive models with the Fourier amplitude sensitivity test. Water Resources Research 47. https://doi.org/10.1029/2010WR009947

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. Computer Physics Communications 181, 259–270. https://doi.org/10.1016/J.CPC.2009.09.018

Saltelli, A., Tarantola, S., Chan, K.P.S., 1999. A quantitative model-independent method for global sensitivity analysis of model output. Technometrics 41, 39–56. https://doi.org/10.1080/00401706.1999.10485594

Sexton, J., Everingham, Y., Inman-Bamber, G., 2016. A theoretical and real-world evaluation of two Bayesian techniques for the calibration of variety parameters in a sugarcane crop model. Environmental Modelling and Software 83, 126–142. https://doi.org/10.1016/j.envsoft.2016.05.014

Sexton, J., Everingham, Y.L., Inman-Bamber, G., 2017. A global sensitivity analysis of cultivar trait parameters in a sugarcane growth model for contrasting production environments in Queensland, Australia. European Journal of Agronomy 88, 96–105. https://doi.org/10.1016/j.eja.2015.11.009

Sheng, M., Liu, J., Zhu, A.X., Rossiter, D.G., Liu, H., Liu, Z., Zhu, L., 2019. Comparison of GLUE and DREAM for the estimation of cultivar parameters in the APSIM-maize model. Agricultural and Forest Meteorology 278. https://doi.org/10.1016/j.agrformet.2019.107659

Sinclair, T.R., Seligman, N.G., 1996. Crop Modeling: From Infancy to Maturity. Agronomy Journal 88, 698–704. https://doi.org/10.2134/agronj1996.00021962008800050004x

Singels, A., 2013. Crop models. Sugarcane: physiology, biochemistry, and functional biology 541–577. https://doi.org/https://doi.org/10.1002/9781118771280.ch20

Singels, A., Bezuidenhout, C.N., 2002. A new method of simulating dry matter partitioning in the Canegro sugarcane model. Field Crops Research 78, 151–164. https://doi.org/https://doi.org/10.1016/S0378-4290(02)00118-1

Singels, A., Jones, M., Marin, F., Ruane, A., Thorburn, P., 2013. Predicting Climate Change Impacts on Sugarcane Production at Sites in Australia, Brazil and South Africa Using the Canegro Model. Sugar Tech 16, 347–355. https://doi.org/10.1007/s12355-013-0274-1

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Soetaert, K., Petzoldt, T., 2010. Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME. Journal of Statistical Software 3, 1–28. https://doi.org/10.18637/jss.v033.i03

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. Journal of Hydrology. https://doi.org/10.1016/j.jhydrol.2015.02.013

Sreelash, K., Sekhar, M., Ruiz, L., Tomer, S.K., Guérif, M., Buis, S., Durand, P., Gascuel-Odoux, C., 2012. Parameter estimation of a two-horizon soil profile by combining crop canopy and surface soil moisture observations using GLUE. Journal of Hydrology 456–457, 57–67. https://doi.org/10.1016/j.jhydrol.2012.06.012

Tan, J., Cui, Y., Luo, Y., 2016. Global sensitivity analysis of outputs over rice-growth process in ORYZA model. Environmental Modelling and Software 83, 36–46. https://doi.org/10.1016/j.envsoft.2016.05.001

Thorburn, P.J., Meier, E.A., Probert, M.E., 2005. Modelling nitrogen dynamics in sugarcane systems: Recent advances and applications, in: Field Crops Research. https://doi.org/10.1016/j.fcr.2005.01.016

Valade, A., Ciais, P., Vuichard, N., Viovy, N., Caubel, A., Huth, N., Marin, F., Martiné, J.-F., 2014. Modeling sugarcane yield with a process-based model from site to continental scale: uncertainties arising from model structure and parameter values. Geosci. Model Dev 7, 1225–1245. https://doi.org/10.5194/gmd-7-1225-2014

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environmental Modelling & Software 25, 310–319. https://doi.org/https://doi.org/10.1016/j.envsoft.2009.09.012

Vazquez-Cruz, M.A., Guzman-Cruz, R., Lopez-Cruz, I.L., Cornejo-Perez, O., Torres-Pacheco, I., Guevara-Gonzalez, R.G., 2014. Global sensitivity analysis by means of EFAST and Sobol' methods and calibration of reduced state-variable TOMGRO model using genetic algorithms. Computers and Electronics in Agriculture 100, 1–12. https://doi.org/10.1016/J.COMPAG.2013.10.006

Vianna, M. dos S., Nassif, D.S.P., dos Santos Carvalho, K., Marin, F.R., 2020. Modelling the trash blanket effect on sugarcane growth and water use. Computers and Electronics in Agriculture 172, 105361. https://doi.org/10.1016/j.compag.2020.105361

Wallach, D., Makowski, D., Jones, J.W., Brun, F., Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2019. Data Assimilation for Dynamic Models. Working with Dynamic Crop Models 487–518. https://doi.org/10.1016/B978-0-12-811756-9.00014-9

Wang, J., Li, X., Lu, L., Fang, F., 2013. Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. Environmental Modelling and Software 48, 171–182. https://doi.org/10.1016/j.envsoft.2013.06.007

Xing, H. min, Xu, X. gang, Li, Z. hai, Chen, Y. jin, Feng, H. kuan, Yang, G. jun, Chen, Z. xia, 2017. Global sensitivity analysis of the AquaCrop model for winter wheat under different water treatments based on the extended Fourier amplitude sensitivity test. Journal of Integrative Agriculture 16, 2444–2458. https://doi.org/10.1016/S2095-3119(16)61626-X

Xu, C., Gertner, G., 2011. Understanding and comparisons of different sampling approaches for the Fourier Amplitudes Sensitivity Test (FAST). Computational Statistics and Data Analysis 55, 184–198. https://doi.org/10.1016/j.csda.2010.06.028

Zhang, Y., Arabi, M., Paustian, K., 2020. Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. Environmental Modelling and Software 126. https://doi.org/10.1016/j.envsoft.2020.104642

4. GLOBAL SENSITIVITY ANALYSIS OF GENOTYPE AND SOIL PARAMETERS IN A SUGARCANE MODEL FOR CONTRASTING PRODUCTION ENVIRONMENTS ACROSS BRAZIL

Abstract

Due to climate change issues and the limitations for expanding agricultural areas, Brazilian farms are constantly pushed to increase production efficiency. The use of process-based models (PBCM) to test cultivars, management options and environmental evaluation are currently a reality. For model applications such as these, it is important to determine which parameters explain the greatest variation in simulations in different production environments. In this sense, our study determined which are the most important genotype parameters for the SAMUCA crop model in the main producing regions in Brazil, for irrigated and rainfed conditions, and the importance of soil parameters in those environments. We performed a global sensitivity analysis (GSA) and calculated two sensitivity indices: the extended Fourier Amplitude Sensitivity Test (eFAST) and Partial Rank Correlation Coefficient (PRCC). A total of 31 parameters were analyzed, 24 of which genotype and 7 from the soil, and we concluded that only 13 parameters were significant, regardless of the output variable. Furthermore, we confirmed that the weather mainly affected the parameters plastochron and eff, which are important for fresh and dry stalk mass variables. In environments with well rainfall distribution the plastochron was the main parameter, while in environments subjected to greater water stress, the eff was the most important parameter. We noticed that any soil parameter was important for irrigated conditions, while for rainfed, the field capacity and the permanent wilting point were relevant in environments with poor rainfall distribution and shallow soils. Rainy places with deep soils showed no sensitivity to soil parameters as well.

Keywords: SAMUCA, PRCC, eFAST

4.1. Introduction

Sugarcane occupies 8.3 Mha in Brazil with an average yield of 70.3 Mg ha-1 (CONAB, 2022) in diverse areas in terms of climate, soil, management, and cultivars. The Brazilian sugarcane industry is in constant development, testing and releasing new cultivars and management practices to increase the yield. It is consensus that environment and genotypes control the sugarcane yield level (Leal et al., 2013; Vianna et al., 2020) and interacts with farming practices (Soares and Marin, 2021), interrow spacing (Gasparotto et al., 2020) and crop water responses (Inman-Bamber and Smith, 2005). Process-based crop models (PBCM) are able to assess cultivars performances in terms of yield (Inman-Bamber et al., 2012; Jeuffroy et al., 2006), the effect of green cane trash blanket (GCTB), interrow spacing and soil processes on crop growth (Vianna et al., 2020), and irrigation strategies (Coelho et al., 2020) across distinct environments.

The SAMUCA model (Marin & Jones, 2014; Marin et al., 2017; Vianna et al., 2020) was developed to consider important soil-plant-atmosphere processes to the crop farming system, and it was so far mainly tested for one of the most cultivated cultivar in Brazil, the RB867515 (RIDESA, 2018). It is well known that the development of new cultivars is a matter of time, motivated by different strategies such as greater resistance to water stress (Natarajan et al., 2020), pests and diseases (Cursi et al., 2022; Zhang et al., 2021), greater yield in irrigated areas (B Péné, 2019; Béhou and Péné, 2020) and for adaptation to the climate change (Marin et al., 2014; Zhao and Li, 2015). Yet, sugarcane crop has been expanded to new areas with different climates and soils in Brazil (Arruda et al., 2017). To implement new cultivars in PBCM, the determination of which soil and genotype parameters are important for assuring good model performance and highly relevant for model calibration.

Usually, sugarcane PBCM's have many genotype parameters compared to other PBCM's developed for annual crops such as maize and wheat. For instance, DSSAT/CASUPRO, DSSAT/CANEGRO, and SAMUCA are PCBM's dedicated to sugarcane, respectively 33, 18, and 24 genotype parameters, while DSSAT/CERES-MAIZE and DSSAT/CERES-WHEAT only 6 and 7 parameters, respectively. PBCM's also have many soil parameters, which also demand specific calibration based on field measurements. Calibrating these parameters requires complex experiments and much time, which increasingly motivates the use of statistical methods to optimize the calibration process (He et al., 2009; Marin et al., 2011; Maulidiani et al., 2018). However, it is interesting to reduce the number of parameters to be measured or estimated for including new cultivars or soil profiles into the PBCM. Parameters that are influential but are not easily measurable are ideal candidates for statistical calibration, while parameters that do not influence model outputs or do not vary greatly could remain fixed to default values (Sexton et al., 2017). Sensitivity analysis (SA) is a tool that can determine which parameters are most important to the model (Wallach et al., 2018). The SA can be divided into two groups: the local sensitivity analysis (LSA) and the global sensitivity analysis (GSA). The LSA consists of changing a single parameter at a time, while the remaining parameters are kept at their reference values; in other words, this method is based on the local derivatives of the model's output concerning the variation of a single parameter, which indicates how fast the output varies around the reference parameter values. The GSA is more robust and allows to evaluate the entire uncertainty range of parameters, considering changes in all parameters along with their range, as well as the interactions among parameters (Saltelli et al. 1999). The GSA methods can also be classified into three groups: screening, regression, and variance; all of these follow the Bayesian sampling principle. The most used screening method is the Morris method, which permits to define the most important model parameters and it is often considered a qualitative method (Dejonge et al., 2012; Morris, 1991). Regression methods, such as the Partial Rank Correlation Coefficient (PRCC), provide the correlation between the model's output and the selected parameters and they are mandatory when the parameter and the model have a monotonic relationship (Krishnan and Aggarwal, 2018; Marino et al., 2008). Finally, the methods based on the variance are the most commonly used, and the three main are: Sobol (Sobol, 2001a), Fourier Amplitude Sensitivity Test (FAST) (Cukier et al., 1978), and the extended Fourier Amplitude Sensitivity Test (eFAST) (Saltelli et al., 1999); they provide the parameters causing the greater variability for the model's output and usually have a high computational cost.

Studies have shown that a GSA can be dependent on the method used (Marino et al., 2008), the range of parameters (Wang et al., 2013), the time-series for which the output variable is analyzed (Xing et al., 2017), climate (Sexton et al., 2017), soil (Varella et al., 2010) and management (Zhang et al., 2020; Zhao et al., 2014). Pereira et al. (2021) used the PRCC method to determine which parameters are more important for the SAMUCA model, concluding that green cane trash blanket (GCTB) did not influence the importance of parameters. However, Pereira et al. (2021) analysed only the cultivar RB867515 growing in one irrigated site in Brazil and did not consider the GSA throughout the crop cycle, being these two aspects demonstrably determinant for the GSA results (Jin et al., 2018; Silvestro et al., 2017). In Brazil there are different climates and soils in which sugarcane is produced under rainfed and irrigated conditions. To our knowledge, there are no studies evaluating the impact of soil and genotype parameters across different environments and what their real impacts are on the main output variables of a crop model. In this sense, our study aimed (i) to determine which are the most important soil and genotype parameters using a robust set of experiments conducted across different producing environments in Brazil for irrigated and rainfed environments, and (ii) to evaluate the importance of the soil parameters across different producing environments in Brazil.

4.2. Material and Methods

4.2.1. The SAMUCA model

The SAMUCA model creation was motivated by the argument of Sinclair and Seligman (1996) in which they highlighted the importance of developing the proper models for knowledge groups, allowing to deeply understand and represent the mechanisms involved in the simulation process and the uncertainties inherent to the given models. The SAMUCA is a mechanistic crop growth model tested and applied to the environmental conditions and management of sugarcane in Brazil. The most updated version of SAMUCA model was developed by Vianna et al. (2020), which was aimed to reduce the uncertainties around model structure, soil moisture, and heat flow in comparison with previous versions (Marin and Jones, 2014; Marin et al.,2017). Soil moisture is simulated by the widely tested "tipping bucket" method, whereas heat flow is solved numerically according to Kroes et al. (2009). Both processes can also be simulated under the effect of GCTB, which has recently emerged as an important operational practice for Brazilian farmers (Carvalho et al. 2017). Further improvements were also made to the subroutines dedicated to the simulation of carbon partitioning at phytomer level, layered-canopy photosynthesis, tillering and root growth (Laclau and Laclau, 2009; Bezuidenhout et al., 2003; O'Leary, 2000). This new version of the SAMUCA model is also included in the DSSAT platform v4.8.

4.2.2. Soil covering and water treatment simulations

To investigate the effect of weather, soil, and irrigation on the sensitivity analysis and consequently determine the most important parameters for each specific condition, we conducted an extensive literature review to simulate experiments in different sugarcane production environments in Brazil. We also obtained several studies that well described experiments in terms of climates, soils, and cultivars, from which we chose eight experiments across 8 producing regions of Brazil (Figure 13), comprising different planting dates (Table 17), three different soils, and four different climates (Figure 14) according to the Köppen classification. To evaluate the effect of irrigation, we carried out simulations for all sites using irrigated and rainfed treatments, even though these conditions were not present in the field experiment conducted. In this way, we were able to determine which were the main soil parameters for the rainfed and irrigated conditions. We set up the model for automatic irrigation whenever the water content of the soil reached 80% of the total water availability (TAW).

In our simulations, we considered only experiments with plant cane and because of this we adopted no GCTB as plant cane plots usually have bare soil. Yet, according to Pereira et al., (2021) soil covering did not affect the GSA in SAMUCA, with no change in the order of importance of the genotype parameters among plant cane and ratoons.

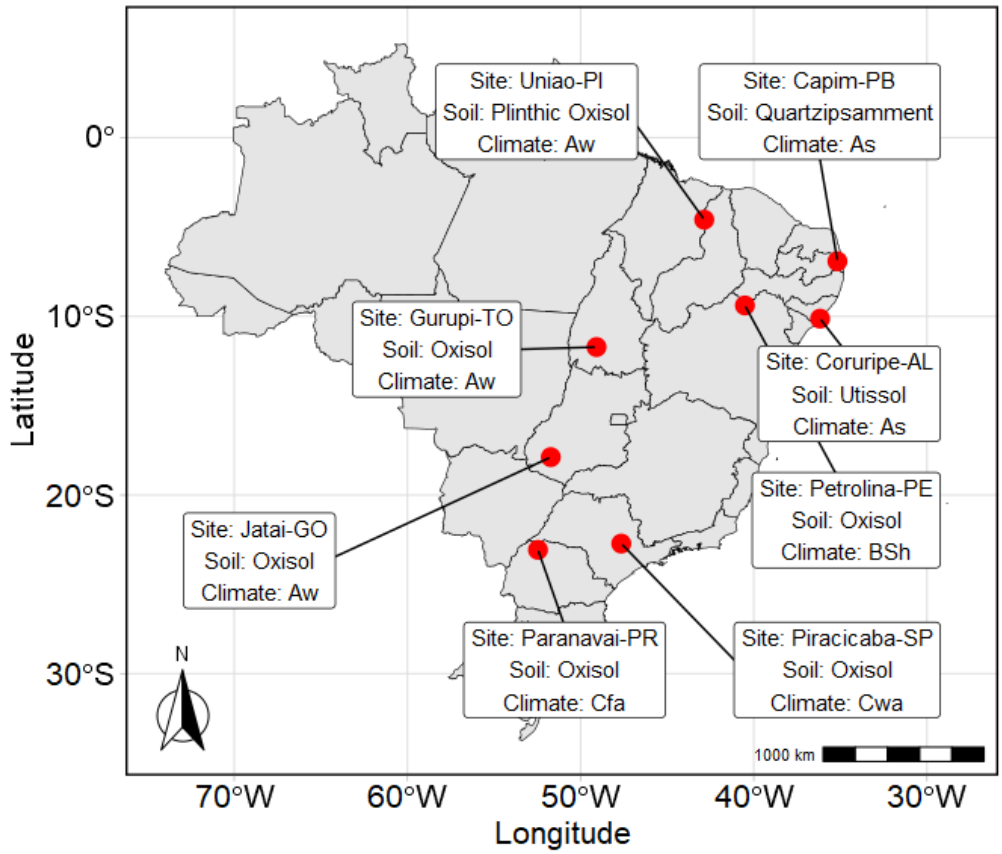


Figure 13. Geographic location, soil and climate of the locations considered for the global sensitivity analysis of the model SAMUCA.

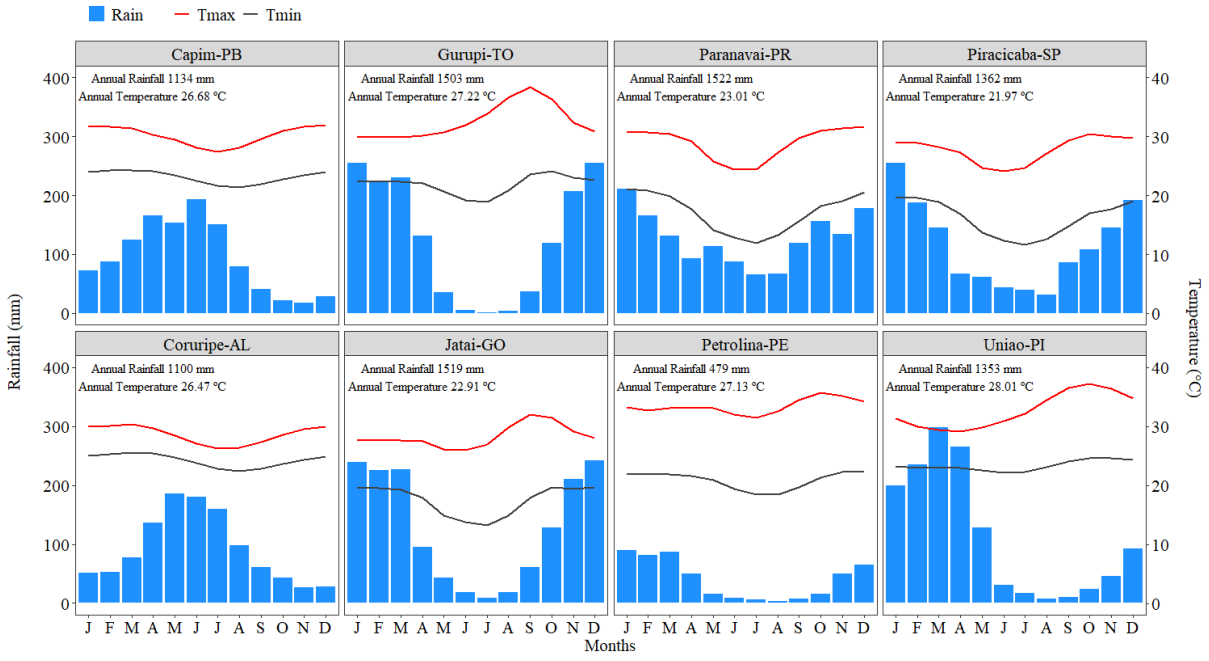


Figure 14. Monthly variation of mean maximum and minimum temperatures (Tmax and Tmin) and rainfall (Rain) of the eight simulated location

Table 17. Description of site, geography information, soil taxonomy, wise soil reference, depth, cultivar, planting and harvesting dates (PaH)

Site	Geographical Definition	Soil Taxonomy	Wise Reference	Depth (cm)	Cultivar	PaH
♆ Coruripe, AL	10°13'S / 36°17'W / 016 m asl	Utissol	Haplic Acrisol	55	RB867515	09/16/2005 to 10/15/2006
♌ Capim, PB	06°57'S / 35°08'W / 59 m asl	Quartzipsamment	Ferralic Arenosol	70	RB867515	07/15/2007 to 07/15/2008
♂ Jataí, GO	17°52'S / 51°43'W / 731 m asl	Oxisol	Haplic Ferrasol	120	RB867515	07/26/2009 to 07/26/2010
♆ União, PI	04°35'S / 42°51'W / 52 m asl	Plinthic Oxisol	Humic Plinthosol	120	RB867515	09/29/2007 to 06/16/2008
♂ Piracicaba, SP	22°52'S / 47°30'W / 560 m asl	Oxisol	Haplic Ferrasol	120	RB867515	10/16/2021 to 10/15/2013
♓ Gurupi, TO	11°43'S / 49°04'W / 283 m asl	Oxisol	Haplic Ferrasol	120	RB832846	09/15/2015 to 09/15/2016
♇ Paranavaí, PR	23°04'S / 52°25'W / 446 m asl	Oxisol	Haplic Ferrasol	120	RB72454	07/15/2008 to 07/15/2009
♂ Petrolina, PE	09°23'S / 40°30'W/ 376 m asl	Oxisol	Haplic Ferrasol	120	RB867515	07/26/2009 to 07/26/2010

♆ Marin et al., 2015; ♌ Vianna and Sentelhas (2015); ♂ Vianna et al., (2020); ♇ de Souza et al., (2015); ♓ Barros et al., (2012); PaH is planting and harvest date.

4.2.3. Soil parameters range

The definition of soil parameters was based on the information from experimental descriptions available in the literature (Table 18). However, to deal with the lack of information in some regions, as mentioned in Marin et al. (2015), we used the WISE database developed by the International Soil Reference and Information Center (ISRIC) (Batjes, 2009; Gijsman et al., 2007) for gap filling the soil profiles with missing data of representative soil class. For each soil profile, we calculated a variation rate for each soil layer in relation to the top soil layer (ΔP) (Eq.14). We thus performed a non-parametric regression for each of the parameters as a function of depth (Figure 15) for each soil profile considering the following parameters: wilting point (WPp), field capacity (FCp), saturation point (STp), saturated hydraulic conductivity (Ksat), clay content (Pclay), silt content (Psilt), sand content (Psand). We then considered a mean curve (black lines in Figure 15) and 95% confidence interval (colored area) to determine the mean, maximum and minimum values of the soil parameters in each layer (Appendix 1). Finally, we randomly sampled only the first layer of the soil profiles and used the regression equations for estimating the parameters of the deeper layers.

$$P_s(n) = P_g + \Delta P(n) \quad (14)$$

where $P_s(n)$ is the value assigned to the soil parameter for layer *n*, based on the value generated for the first layer, P_g; $\Delta P(n)$ is the variation rate between P_g and layer *n*.

Table 18. Soil parameters obtained experimentally (X) and estimated using pedotransfer functions (*).

Site	Field capacity	Wilting point	Saturation point	Saturated hydraulic conductivity	Content clay	Content silt	Content sand
✠ Coruripe, AL	*	*	*	*	*	*	*
ϥ Capim, PB	X	X	X	*	X	X	X
♂ Jataí, GO	*	*	*	*	*	*	*
✠ União, PI	*	*	*	*	*	*	*
♂ Piracicaba, SP	X	X	X	X	X	X	X
✖ Gurupi, TO	*	*	*	*	X	X	X
✯ Paranavaí, PR	X	X	*	*	X	X	X
♂ Petrolina, PE	*	*	*	*	*	*	*

✠ Marin et al., 2015; ϥ Vianna and Sentelhas (2015); ♂ Vianna et al., (2020); ✯ de Souza et al., (2015); ✖ Barros et al., (2012); PaH is planting and harvest date.

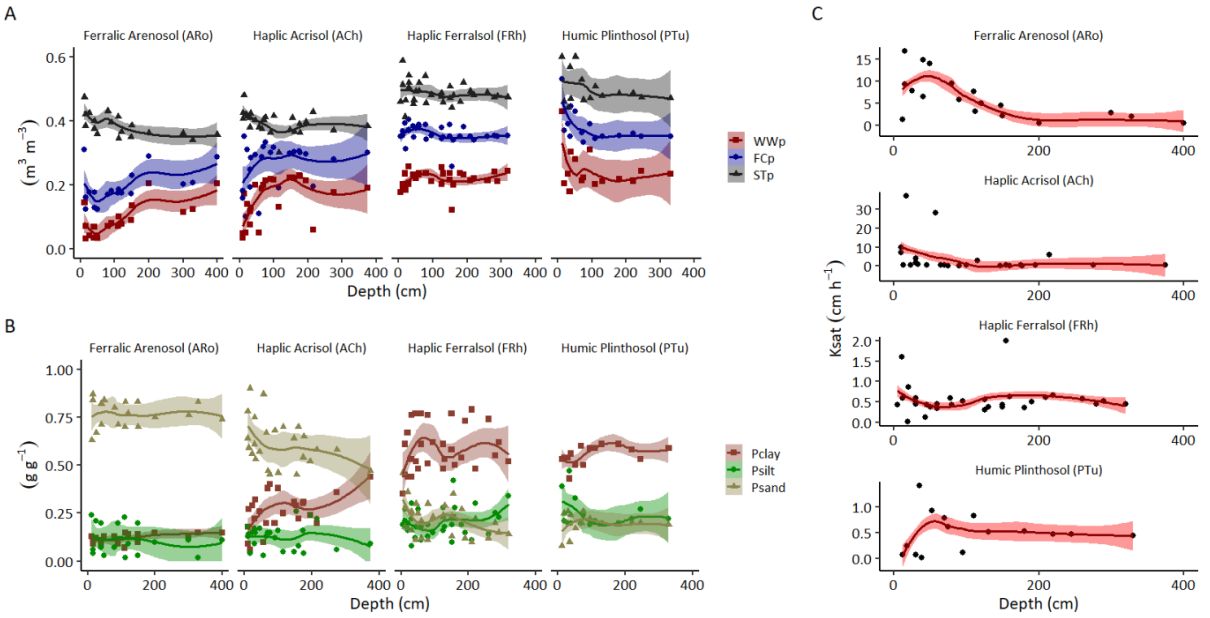


Figure 15. The non-parametric regression for the soil parameters: wilting point (WPp), field capacity (FCp), saturation point (STp), saturated hydraulic conductivity (Ksat), clay content (Pclay), silt content (Psilt), sand content (Psand); dark line represents the mean value and the 95% confidence range and is represented by the colored area.

4.2.4. Genotype parameters range

For GSA, one must respect the variation of parameters observed in the experiments and avoid creating ranges without biophysical foundations (Saltelli et al., 1999), as the parameter range is the greatest source of uncertainty in the GSA results (Wang et al., 2013). As the GSA technique aims at quantification of model uncertainty, the realistic variation of the parameters must be respected by using the measured values and not assumed relative variation ranges (Saltelli et al., 1999). We used the same range of parameters as reported by Pereira et al. (2021) because the set of parameters adopted (Table 19) was based on well-conducted experiments and intensive measurements.

4.2.5. Global sensitivity analyses methods

According to Marino et al. (2008), different GSA methods may result in different outputs, and in this context we use two methods (the Extended Fourier Amplitude Sensitivity Test, eFAST, and the Partial Classification Correlation Coefficient, PRCC) because they have different types of research questions: eFAST answers which parameters cause the greatest variance in the output of the model, while PRCC indicates the degree of correlation between the parameter and the variable of interest. By combining these two methods, we can determine which parameters explain the highest variance and the correlation they have with the output variables. For both methods, we computed the GSA at a daily time step from planting date until harvest considering the output variables: leaf area index (LAI), stalk dry mass (SDM), stalk fresh mass (SFM), sucrose concentration in stalk fresh mass (POL), and tillering (TIL).

4.2.5.1. Extended Fourier Amplitude Sensitivity Test

eFAST is an algorithm that combines two GSA methods: the Fourier Amplitude Sensitivity Test (FAST) and Sobol (Saltelli et al. 1999, 2010), which in turn, use the model output variance principle. While FAST scans the entire parameter space and obtain quantitative sensitivity measures in terms of the main sensitivity index (S_i) of each parameter to output variance, the Sobol calculates the total sensitivity index (ST_i) and provides an indication of the overall effect of a given parameter, considering all possible interactions of that parameter with others (Sobol, 2001b). Therefore, by integrating the merits of FAST and Sobol's algorithms, eFAST offers a method that has the characteristics of high efficiency and precision, as well as the ability to calculate the interaction effects between parameters. The method is based on the decomposition of the model's output variance, determining which fraction of the variance can be explained by the variation in each input parameter. In recent years, due to these advantageous properties, eFAST has become more popular in hydrological, ecological, and agronomy modeling (Li et al., 2019; Reusser et al., 2011; Varella et al., 2010b; Xing et al., 2017). In this study, we implemented the SAMUCA model in the sensitivity R-package available at: https://cran.r-project.org/web/packages/sensitivity/index.html to apply the method eFAST.

Two sensitivity indices can be computed for each parameter S_i (Eq. 15) and ST_i (Eq.16). The S_i and ST_i must vary between 0 and 1, where the effects are greater when the indices reach values close to 1 whereas values close to 0 indicate negligible effects. According to Dejong et al., (2012) and Xing et al., (2017) it is considered that the effect of the parameter can be negligible when $S_i < 0.05$ and $ST_i < 0.1$, that is, indicating that the variance explained by parameter *i* is less than 5% and the variance explained by parameter *i* plus their interaction with the other parameters is less than 10%.

$$S_i = \frac{VAR_i}{VAR_t} \tag{15}$$

$$ST_i = \frac{VAR_t - VAR_{-i}}{VAR_t} \tag{16}$$

where S_i represents the fraction of the output variance of the model explained by the input variation of a given parameter; VAR_i is the estimated conditional variance of the *i-th* parameter. VAR_t is the variance of the output of the variable and VAR_{-i} is the estimated conditional variance, except for the i-th factor is the sum of all the variance terms that do not include the parameters *i*.

The sensitivity indices obtained through FAST can be affected by the boundary conditions of the simulation, such as the presence of GCTB, and by the sample size. In this sense, we carried out a previous study to determine the required sample size and the effect of GCTB on the sensitivity indices obtained by the eFAST method for the SAMUCA model. We tested sample sizes of 65, 129, 257, 513, 1025, 2049, and 4097 for different variables and treatment bare and GCTB. We concluded that the size of 2049 was sufficient to apply eFAST, and the GCTB had minimal influence. The summary of the results is presented in appendix 2 and 3.

Table 19. Cultivar-specific parameters, descriptions, units, and range used for uniform distribution sampling and standard values assumed for initial simulations.

Parameter	Description	Min	θ	Max	Reference
amax	Assimilation rate at light saturation point *(µmol.m^{-2}.s^{-1})*	41.3	44.9	60.7	Sage et al., (2013)
chudec	Heat units for start of tiller abortion *(°C.d)*	1200	1600	1800	Liu et al., (1998)
chumat	Heat units for population establishment *(°C.d)*	1500	1600	2850	Zhou and Shoko, (2011)/Marin and Jones,
chupeak	Heat units for population peak *(°C.d)*	400	1400	1950	Coelho et al., (2020); Marin et al., (2017)
chustk	Heat units for start culm elongation *(°C.d)*	404	650	1050	Marin et al., (2017); /Singels and
eff	Carboxylation efficiency *(µmol.m^{-2}.s^{-1} / µmol.m^{-2}. s^{-1})*	0.040	0.069	0.080	Sage et al., (2013)
end_tt_it_gro	Thermal time for completion of internode growth *(°C.d)*	600	1200	1400	Lingle, (1999)
end_tt_lf_gro	Thermal time for completion of leaf growth *(°C.d)*	1100	1300	1500	Smit and Singels, (2006)
init_lf_area	Initial leaf area of first appeared leaf *(cm²)*	15	10	30	Zhou et al., (2003)
max_ini_la	Initial leaf area of leaves appeared after top parts formation *(cm²)*	80	120	180	Zhou et al., (2003)
max_it_dw	Maximum dry biomass of internodes *(g)*	18	28	35	Lingle, (1999)
maxdgl	Maximum number of developed green leaf a tiller can hold *(#/tiller)*	6	6	12	Vianna et al., (2020)
maxgl	Maximum number of green leaf a tiller can hold (#/tiller)	10.0	12.0	12.0	Marin et al., (2015)
mid_tt_it_gro	Thermal time where internodes can achieve half of its maximum	380	400	700	Lingle, (1999)
mid_tt_lf_gro	Thermal time where leaves can achieve half of its maximum biomass	400	700	800	Smit and Singels, (2006)
mla	Maximum leaf area *(cm²)*	450	600	800	Marin et al. (2014)
n_lf_it_form	Number of leaves appeared before internode formation *(#/tiller)*	3	3	8	Vianna et al., (2020)
n_lf_stk_em	Number of leaves appeared before stalks emerges at soil surface	3	4	8	Vianna et al., (2020)
phyllochron	Phyllochron interval for leaf appearance *(°C.d)*	107	132	169	Marin et al., (2015)/Inman-Bamber, 1994
plastochron	Thermal time required for the appearance of phytometer *(°C.d)*	107	132	169	Marin et al., (2015)/Inman-Bamber, 1994
popmat	Number of tillers on maturation *(tiller/m²)*	8.0	9.5	12.0	Marin and Jones, (2014)
poppeak	Maximum number of tillers *(tiller/m²)*	17.0	22.0	30.0	Marin et al., (2015)
sla	Specific leaf area *(cm^2.g^{-1})*	100.0	120.00	121.00	Ehara et al., (1994)
tillochron	Thermal time required for emergence of new tiller *(°C.d)*	48.1	69.0	134.8	Bezuidenhout, (2000); Zhou and Shoko,

θ is the value calibrated by Vianna et al. (2020) for cultivar RB867515; Max and Min value are range used for random parameters uniform distribution.

4.2.5.2. Partial Rank Correlation Coefficient (PRCC)

We used PRCC because it serves as a complement to eFAST. While eFAST provides the parameters' importance in explaining the largest variance in the model output, the PRCC indicates the correlation between parameter and output variables (Varella et al., 2010; Marino et al., 2008), in our case, for SDM, SFM, TIL, LAI and POL. Unlike Pearson's correlation, when the association between parameter i and variable Y is obtained without the interference of another parameter j, PRCC measures this association taking into account other parameters influence in the output variance (Baba et al., 2004). The method consists of a massive sampling of parameters using the Monte Carlo approach to evaluate the correlation between each parameter and the model output. Therefore, we obtained the linear relationship between the genetic and soil parameters and the multiple model outputs with the PRCC method, where the positive values of PRCC represent direct linear relationship while the negative means an inverse linear relationship. The difference between PRCC and its advantage over Pearson's correlation coefficient is that the PRCC coefficient can be applied in non-linear models. Rank-based methods such as PRCC offer robust SA and easy implementation if the input-output relationship is monotonic (Saltelli et al., 1999). PRCC values range from -1 to 1, measuring the strength of a linear association between an input and an output. In our analysis, we considered only the genotype and soil parameters that were statistically significant at 1% for the components of the sugarcane output model (Marino et al., 2008). In addition, Mukaka (2012) presented different classes of interpretation for PRCC (Table 20) and based on the results obtained by Marino et al. (2008), we considered only the parameters with high and very high correlation (PRCC $\geq |0.7|$), as the parameters with these values explain the high output variance.

Table 20. Rules for interpreting the size of correlation coefficient (Mukaka,2012).

Size of Correlation	Interpretation
0.90 to 1.00 (-0.90 to -1.00)	Very high positive (negative) correlation
0.70 to 0.90 (-0.70 to -0.90)	High positive (negative) correlation
0.50 to 0.70 (-0.50 to -0.70)	Moderate positive (negative) correlation
0.30 to 0.50 (-0.30 to -0.50)	Low positive (negative) correlation
0.00 to 0.30 (-0.00 to -0.30)	Negligible correlation

4.3. Results

4.3.1. Time-series sensitivity analysis

4.3.1.1. Irrigated conditions

The PRCC method identified the same genotype parameters in different locations and did not indicate a significant influence on soil parameters under irrigated conditions (Figure 16). For the TIL variable, two parameters were identified as the most important, and their importance was dependent on the evaluated sample time; the tillochron parameters were inversely related between 0 and 180 days after planting (DAP), while popmat showed a direct relationship between 180 and 360 DAP. (Figure 16). The most critical parameters for SFM and SDM were the *plastochron, n_lf_when_stk_em, n_lf_it_form, and mid_tt_it_gro*, which showed an inverse relation; *eff* had a direct relation

with SFM and SDM (Figure 16). Still, on SDM and SFM variables, the results showed that the importance of the plastochron is lower in the places where the eff becomes more important. This was observed for instance in the final third of the simulation in Capim, Coruripe, Gurupi, Petrolina, and Piracicaba (Figure 16). For POL, *mid_tt_it_gro* and *end_tt_it_gro* were the parameters that had a direct relationship during most parts of the cycle, having the same pattern in all places (Figure 16). Parameters related to LAI were those that most affected the simulations. Mainly *mla* and *plastochron*, where mla shows a positive relationship from 90 to 360 DAP, and *plastochron* negatively affecting from 60 to 90 DAP and positively from 180 to 360 DAP (Figure 16).

The results obtained by the PRCC method agreed with those obtained with the eFAST method, indicating the same pattern observed for genotype parameters and were not influenced by soil parameters (Figure 17). For the variable TIL, for example, more than 90% of the variance was explained by the *tillochron* (60 to 180 DAP) and *popmat* (180 to 360 DAP) for all locations; between 120 and 180 DAP there was an effect of other parameters in Piracicaba, such as *mla* and *plastochron*, but their effect did not exceed 15% (Figure 17). The parameters that explained most the variance of SDM and SFM were *plastochron, n_lf_stk_emerg, n_lf_it_form, mid_tt_it_gro* and *eff*. These five parameters were also observed with the PRCC method, whereas the parameter *popmat* was also influential according to the PRCC method (at the end of the cycle of SDM and SFM in Piracicaba; Figure 17). Overall, for SDM and SFM, more than 60% of the variance over the cycle was explained by *eff* or *plastochron* at all sites (Figure 17). For POL, the parameter *mid_tt_it_gro* explains more than 85% of the variance, starting from DAP 180 for all locations (Figure 17), and *end_tt_it_gro* explained most of the remaining variance (approximately 15%). The LAI was affected by up to 13 genotype parameters, each part of the cycle being affected differently, such as *init_leaf_init* from 0 to 90 DAP (explaining more than 70% of the variance), *plastochron* from 60 to 120 DAP (explaining between 40 to 70 % of variance) and *mla* (explaining between 30 to 50% of the variance in the more part of cycle) (Figure 17).

The ST_i (Figure 18) showed that the interaction of the *sla, phyllochron, n_lf_it_form* and *maxgl* parameters with the other genotype parameters made these parameters momentarily important close to the 70 DAP for the LAI variable in Paranavaí-PR and Petrolina-PE (Figure 18). However, the parameters that dominate the simulation process were the same as previously indicated by the values of S_i. The POL was another variable that, in addition to the parameters already mentioned in the previous paragraph, had the momentary inclusion of the *maxgl* and *max_it_dw* parameters (Figure 18). Finally, for the variables SDM, SFM and TIL, we did not observe changes in comparison with the results obtained by the values of S_i.

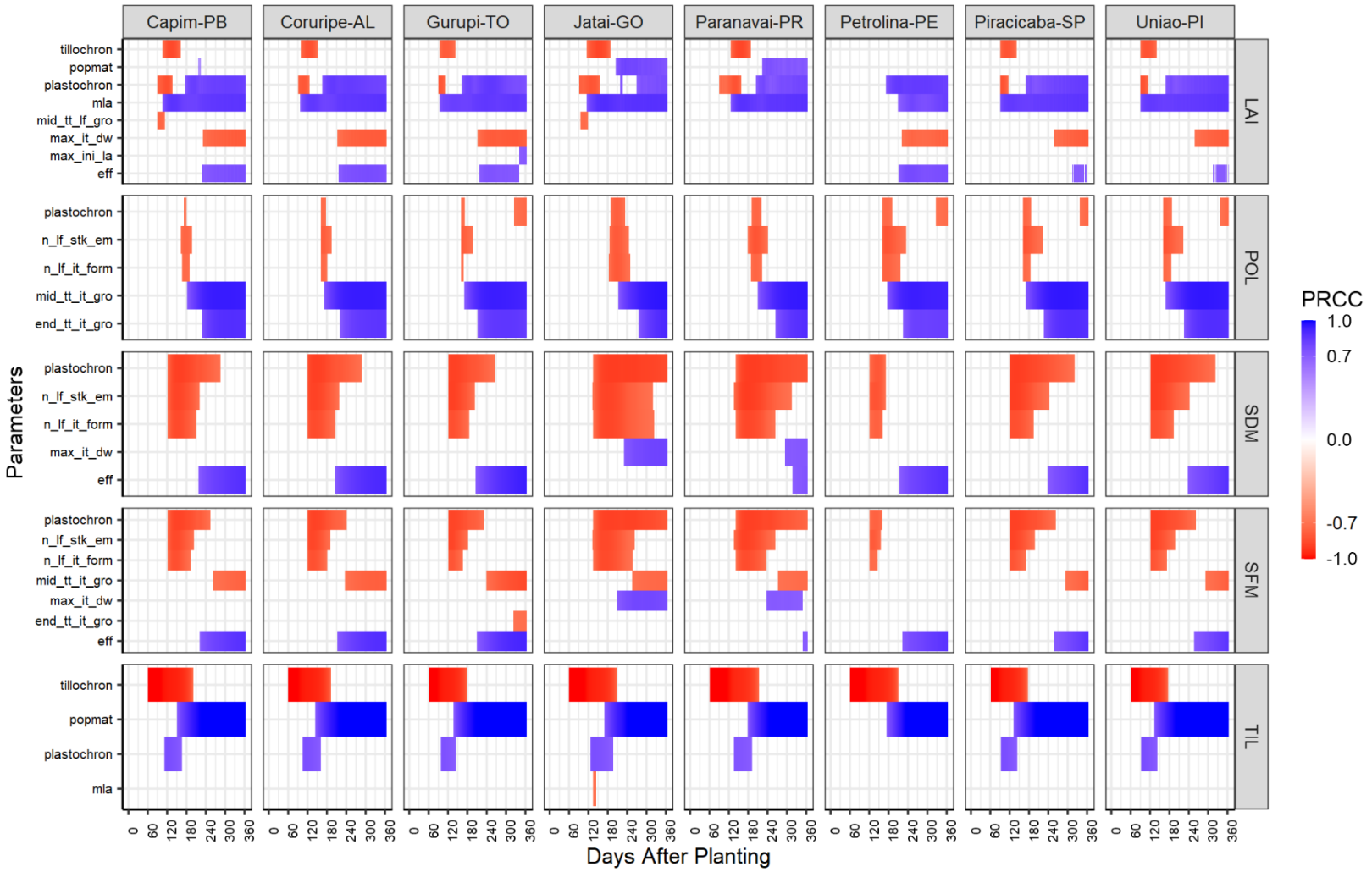


Figure 16. Partial Rank Correlation Coefficient (PRCC) results for time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering (TIL) the treatment irrigation

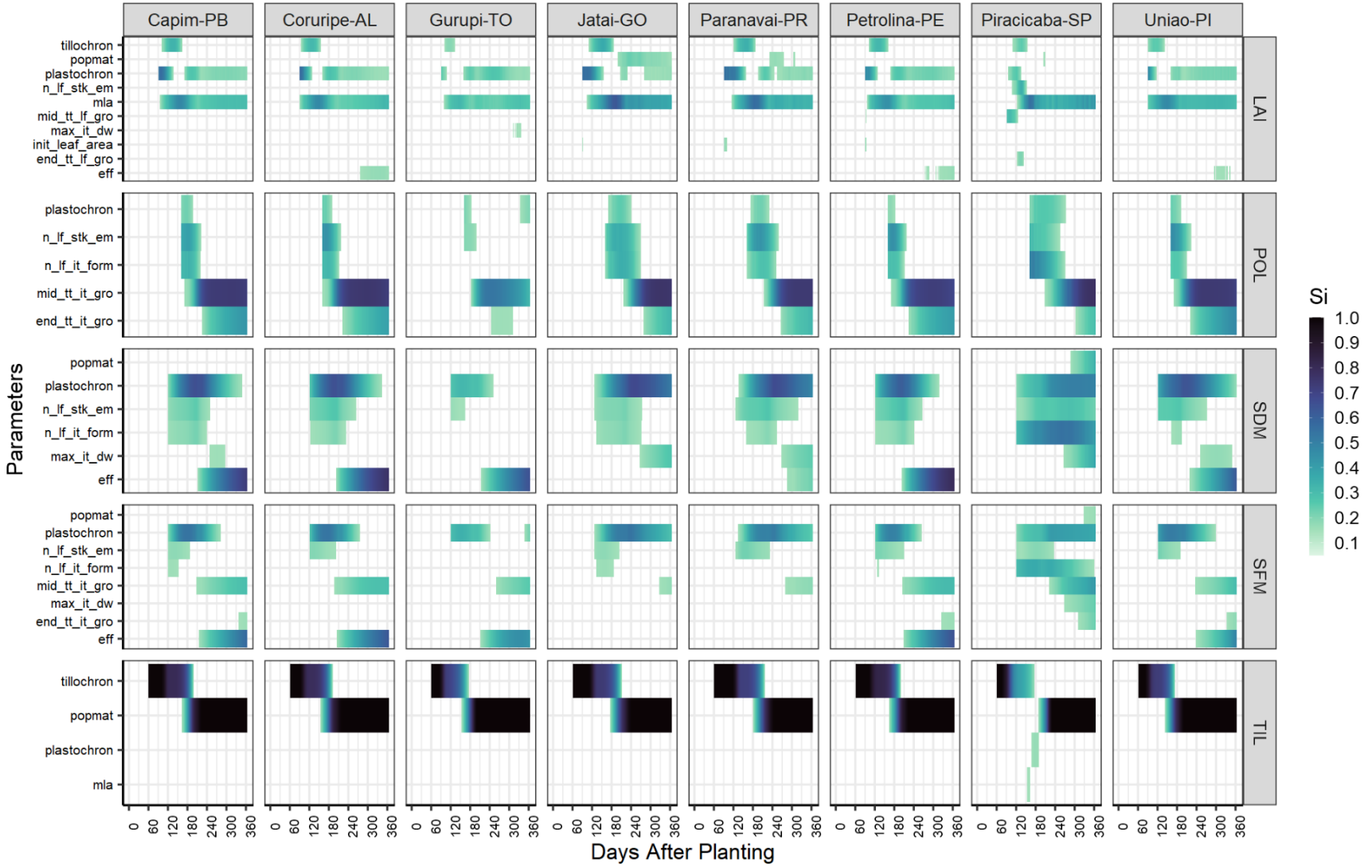


Figure 17. The main sensitivity index (S_i) results for time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering (TIL) the treatment irrigation.

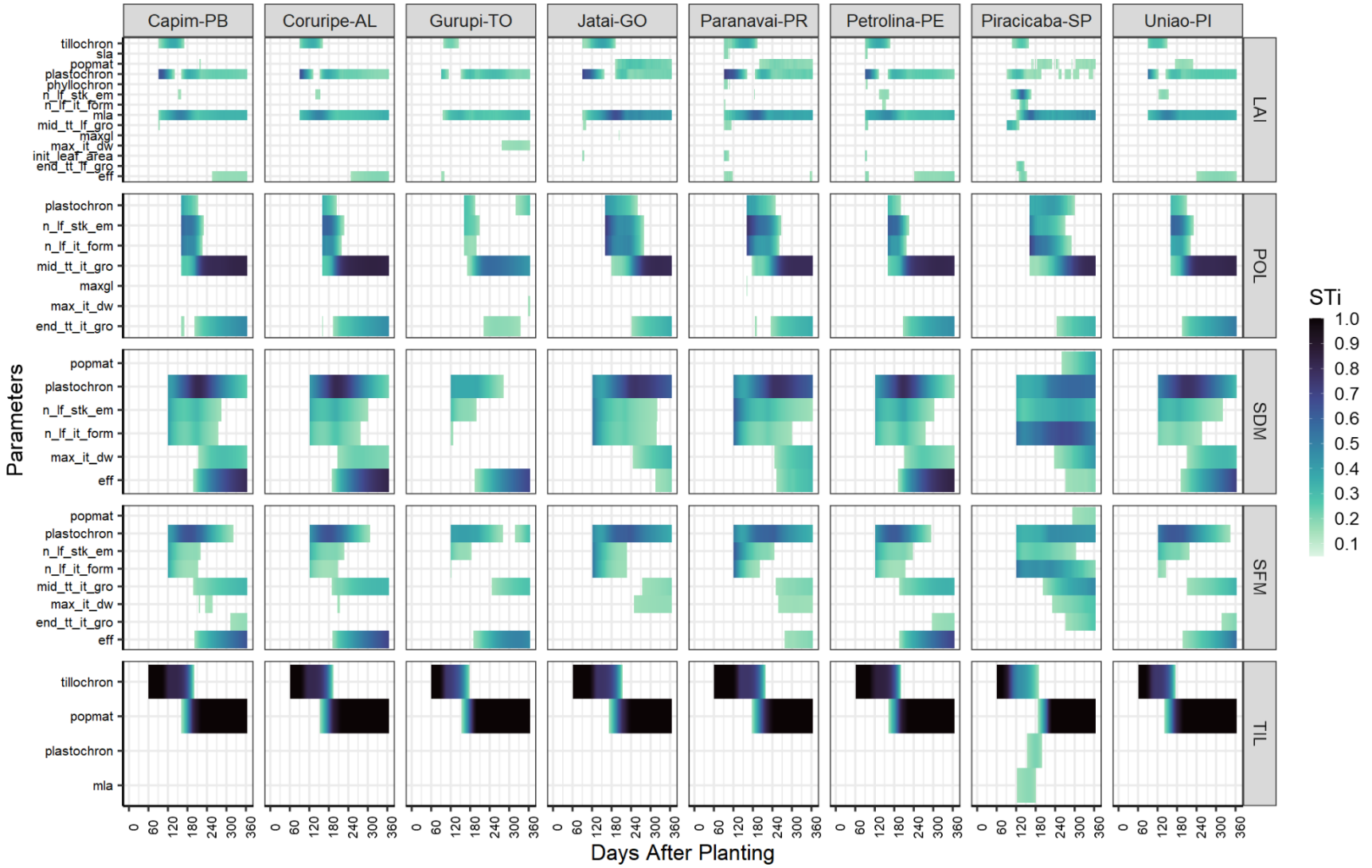


Figure 18. The total sensitivity index (ST_i) results for time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering (TIL) the treatment irrigation

4.3.1.2. Rainfed conditions

Under rainfed conditions, the PRCC method identified the same genotype parameters in different locations. In addition, it highlighted the importance of soil parameters, *WPp* and *FCp*, in environments with greater water restriction. First, the TIL and POL variables had a weak influence of soil parameters (Figure 19). The TIL was inversely related between 0 and 180 DAP to the *tillochron* parameter; later it was directly related to the *popmat* parameter (Figure 19). POL had the parameters *mid_tt_it_gro* and *end_tt_it_gro* as the two most influential, showing a direct relationship during most of the growing cycle and similar pattern in all locations (Figure 19). The most important parameters for SFM and SDM were *plastochron, eff, n_lf_when_stk_em, n_lf_it_form* and *mid_tt_it_gro*, which showed an inverse relationship; *eff* was directly related to SFM and SDM (Figure 19). Yet, for the SDM and SFM variables, the results showed that the importance of the *plastochron* is lower in places where *eff* becomes more important. This was observed, for example, in the final third of the simulation in Capim-PB, Coruripe-AL, Gurupi-TO and Petrolina-PB (Figure 19), some of the hottest places in our data set. Furthermore, in Coruripe-PB and Capim-PB, the parameters *FCp* (directly linear to SDM and SFM) and *WPp* (inversely linear to SDM and SFM) were significant between 150 and 240 DAP (Figure 19). The parameters related to LAI were the ones that most affected the simulations, mainly *mla* and *plastochron*; mla stood out for having a positive relationship from 90 to 360 DAP, and *plastochron* for negatively affect it from 60 to 90 DAP and positively affect it from 180 to 360 DAP (Figure 19). Nevertheless, just as SDM and SFM, LAI was influenced by soil parameters *FCp* (directly linear), and *WPp* (inversely linear), between 120 and 150 DAP.

The results obtained by the eFAST method converged with those obtained by the PRCC, showing the same pattern observed for the genotype and soil parameters in rainfed conditions (Figure 20). For the variable TIL, for example, more than 90% of the variance was explained by *tillochron* (60 to 180 DAP) and *popmat* (180 to 360 DAP) for all locations. We observed an effect of other parameters in Piracicaba between 120 and 180 DAP, such as *mla* and *plastochron*, but their effect did not exceed 15% (Figure 20). In addition to the TIL, the POL was another variable that, under rainfed conditions, did not have any soil parameter explaining part of its variance; in fact, all of its variance was explained by the genotype parameters. For POL, the *mid_tt_it_gro* parameter explained more than 85% of the variance, starting at DAP 180 for all locations (Figure 20), and *end_tt_it_gro* explained most of the remaining variance (approximately 15%). The parameters that most explained the variance of SDM and SFM were *plastochron, n_lf_stk_emerg, n_lf_it_form, mid_tt_it_gro* and *eff*. These five parameters were also observed with the PRCC method, and the only difference in the results was the inclusion of the *popmat* parameter (end of the SDM and SFM cycle in Piracicaba; Figure 20). Overall, for SDM and SFM, more than 60% of the variation across the cycle was explained by *eff* or *plastochron* at all sites (Figure 20). Furthermore, in Capim-PB and Coruripe-PE, the soil parameters (*FCp* and *WPp*) explained between 20 and 40% of the variance between 120 and 180 DAP (Figure 20). LAI was affected by up to 13 genotype parameters, each part of the cycle being affected differently, such as *init_leaf_init* from 0 to 90 DAP (explaining more than 70% of the variance), *plastochron* from 60 to 120 DAP (explaining between 40 and 70% of variance) and *mla* (explaining between 30 and 50% of the variance for most of the cycle) (Figure 20). In addition to genotype parameters, in Capim-PB and Coruripe-AL, soil parameters explained up to 70% of the variance between DAP 60 and 120 (Figure 20).

When analyzing the ST_i (Figure 21), it was observed that the interaction among parameters became greater in the environments with higher water stress in the shallower soils. The number of parameters that explained more than 15% of the variance in TIL, SDM, SFM, and POL increased in Coruripe-AL and Capim-PB in relation to the

remaining sites. For example, for TIL, the *tillochron* and *popmat* parameters explained all variance in all locations with exception of Capim-PB and Coruripe-AL. Moreover, for these two locations, in addition to the soil parameters, *FCp* and *WWp*, the parameters *end_tt_it_gro, init_leaf_area, plastochron*, and *n_lf_it_form* were included.

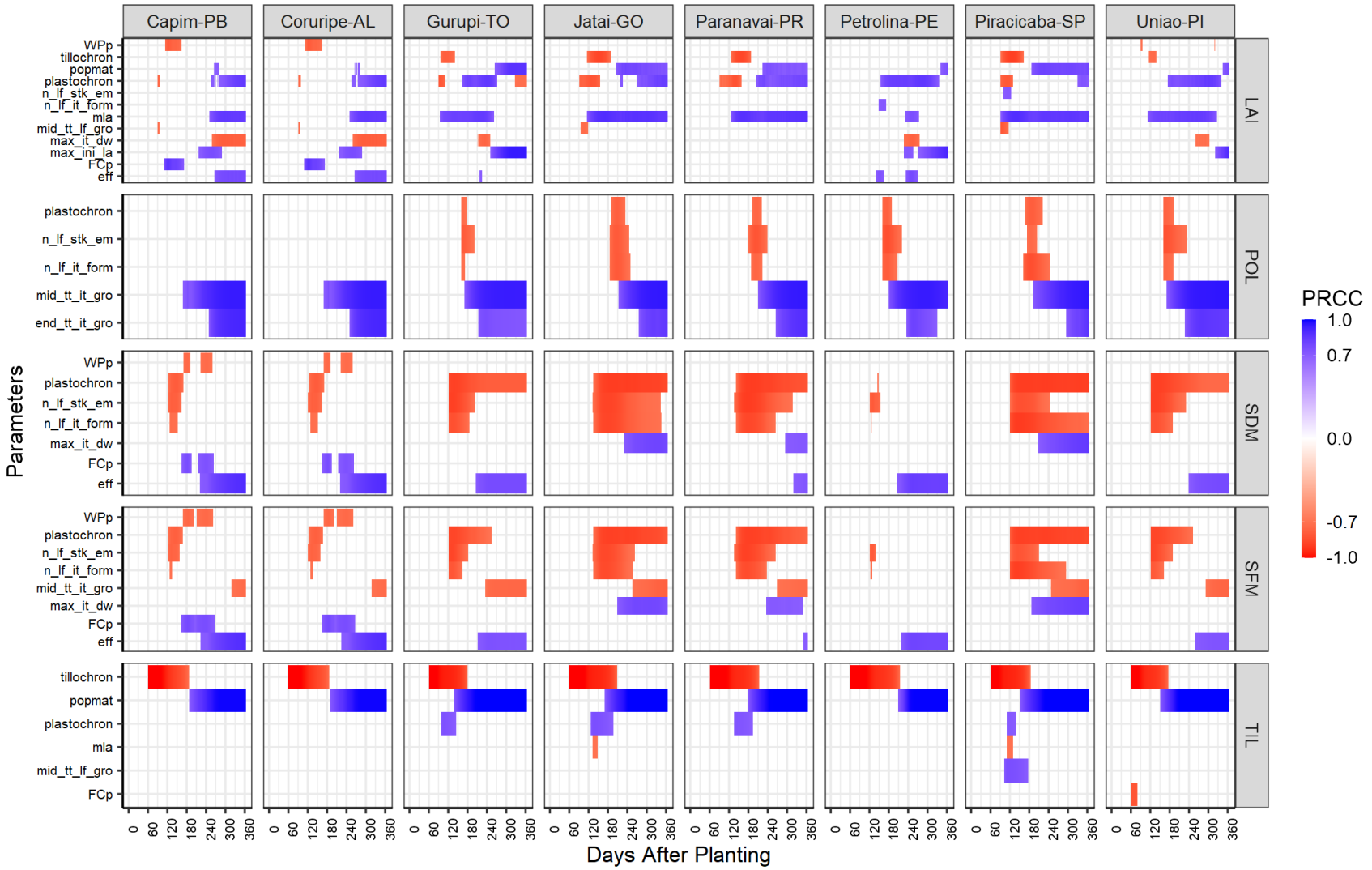


Figure 19. Partial Rank Correlation Coefficient (PRCC) results for time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering (TIL) the treatment rainfed.

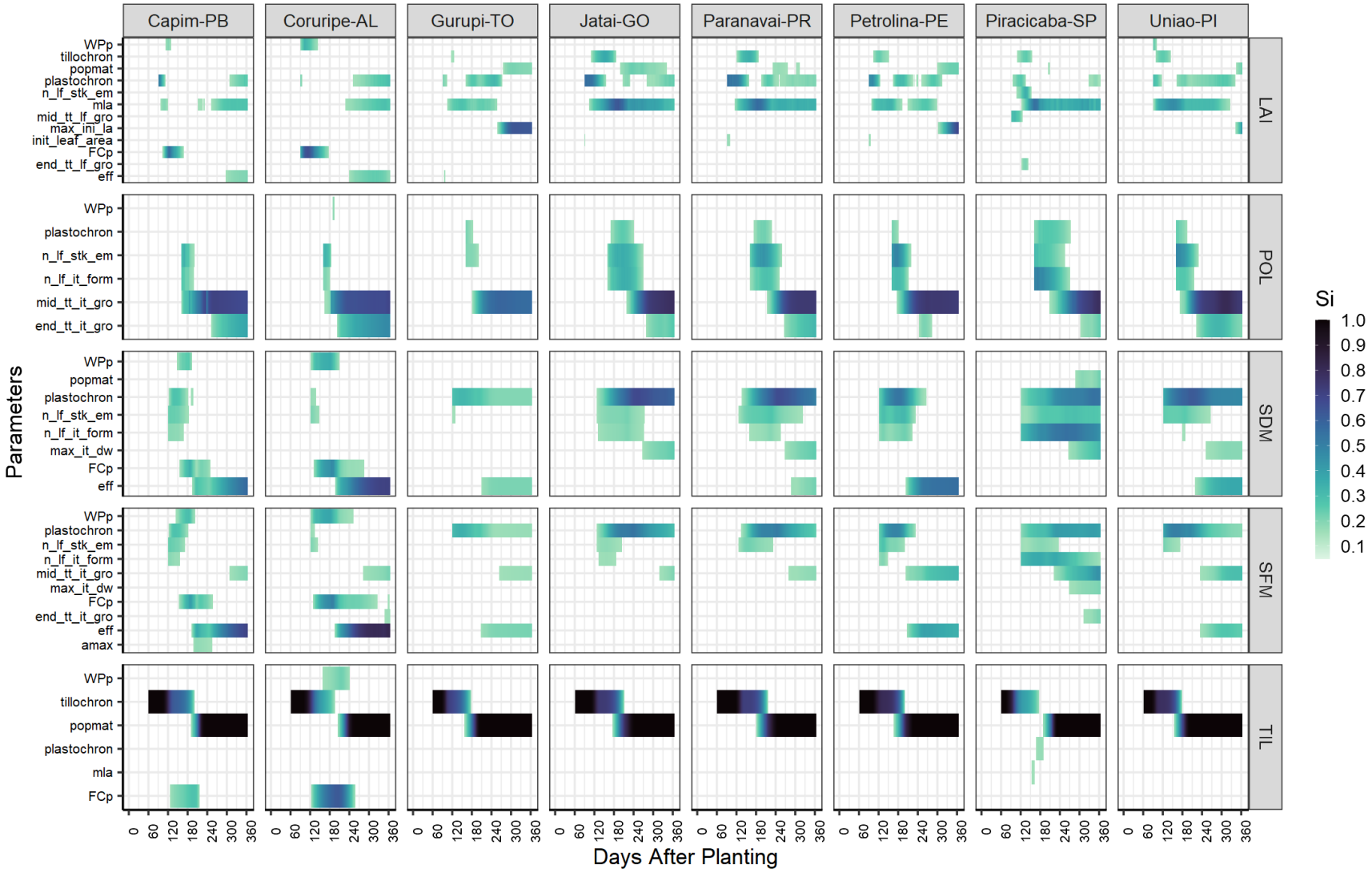


Figure 20. The main sensitivity index (Si) results for time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering (TIL) the treatment rainfed.

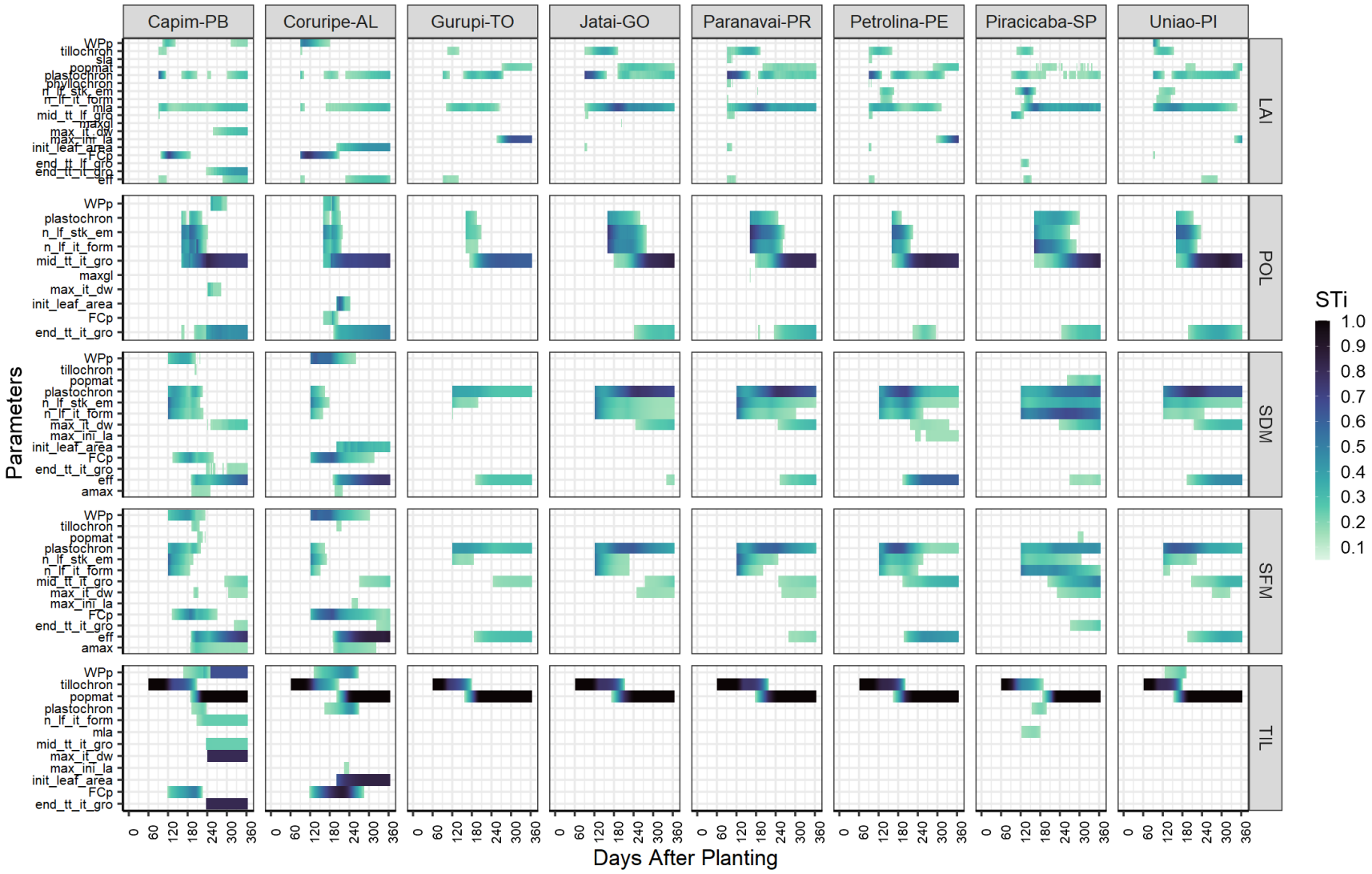


Figure 21. The total sensitivity index (STi) results for time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering (TIL) the treatment rainfed

4.3.2. Rank of most important parameters

To define the parameters that explain the greatest variance of the model, we performed an average over time for the S_i index (Figure 22). It was done because we observed that, at different times in the crop cycle, the parameter responsible for explaining most of the variance could be different. In this way, we were able to identify the most important parameters for each variable in all environments and analyzed the effect of climate on the parameters responsible for explaining the highest output variance.

For the variable LAI, in irrigated environments, the parameter responsible for explaining the highest variance was the *mla*, which on average accounted for 40% of the variance (Figure 22). On the other hand, under rainfed conditions, there was a disparity between the parameters responsible for explaining the higher variance. In Paranavaí-PR, Piraciaba-SP, and Jataí-GO, the *mla* was kept as the main parameter, explaining about 30% of the variance. In União-PI, Petrolina-PE, and Gurupi-TO, however, the *max_ini_lai* explained from 20% to 50% of the variance. The *FCp* parameter was the main one in the regions of Capim-PB and Coruripe-AL, explaining from 30 to 35% of the variance.

POL was the variable that is the least influenced variable among the environments, regarding climate, soil, irrigated and rainfed treatments (Figure 22). We verified that the variance explained by the parameters between the irrigated and rainfed treatments were close, with *mid_tt_it_gro* being the most important parameter, responsible for explaining on average 50% to 60% of the POL variance regardless of climate, soil and water treatment. It should be emphasized again that Coruripe-AL was the only place that indicated the importance of the soil parameter in the rainfed condition for POL; the WPp parameter was responsible for explaining on average 20% of the variance.

The SDM showed to be sensitive to *plastochron* and *eff* parameters in rainfed and irrigated environments (Figure 22). In Capim-PB, Coruripe-AL, Gurupi-TO, and Petrolina-PE, the main parameter was the *eff*, and in Jataí-GO, Paranavaí-PR, Piracicaba-SP, and União-PI the main parameter was the *plastochron*. The variance explained by *eff* varied from 20% (Gurupi-TO) to 45% (Coruripe-AL) in rainfed, and in conditions of irrigation between 35% (Gurupi-TO) and 50% (Capim-PB). The variance explained through *plastochron* ranged from 45%, in Piracicaba-SP, to 55%, in Paranavaí-PR for both rainfed and irrigation treatments.

The main parameters for SFM were *plastochron* and *eff*, and as well as for SDM no influence was observed of conditions rainfed and irrigation about the main parameter (Figure 22). In Jataí-GO, Paranavaí-PR, Piracicaba-SP, and União-PI the main parameter was the *plastochron*, and in Capim-PB, Coruripe-AL, Gurupi-TO, and Petrolina-PE, the main parameter was *eff*. On irrigated conditions, the variance explained by *plastochron* varied between 28% (Piracicaba-SP) e 32% (Jataí-GO), and the *eff* varied between 15% (Gurupi-TO) and 70% (Coruripe-AL). Even in regions where *eff* was the most important parameter, we observed *plastochron* as the second most important. This was not observed in the regions where the *plastochron* was the most important, as in the case of Jataí-GO and Piracicaba-SP, where the *eff* did not appear among the most significant parameters (Figure 22).

The TIL variable was dominated almost in isolation by the *tillochron* and *popmat* parameters in all locations and treatments. Piracicaba also had the *plastochron* and *mla* parameters in both the irrigated and rainfed conditions; and Coruripe-AL and Capim-PB were affected by soil parameters, mainly by *FCp*, which explained less than 20% of the variation in Capim-PB and almost 40% of the variation in Coruripe-AL (Figure 22).

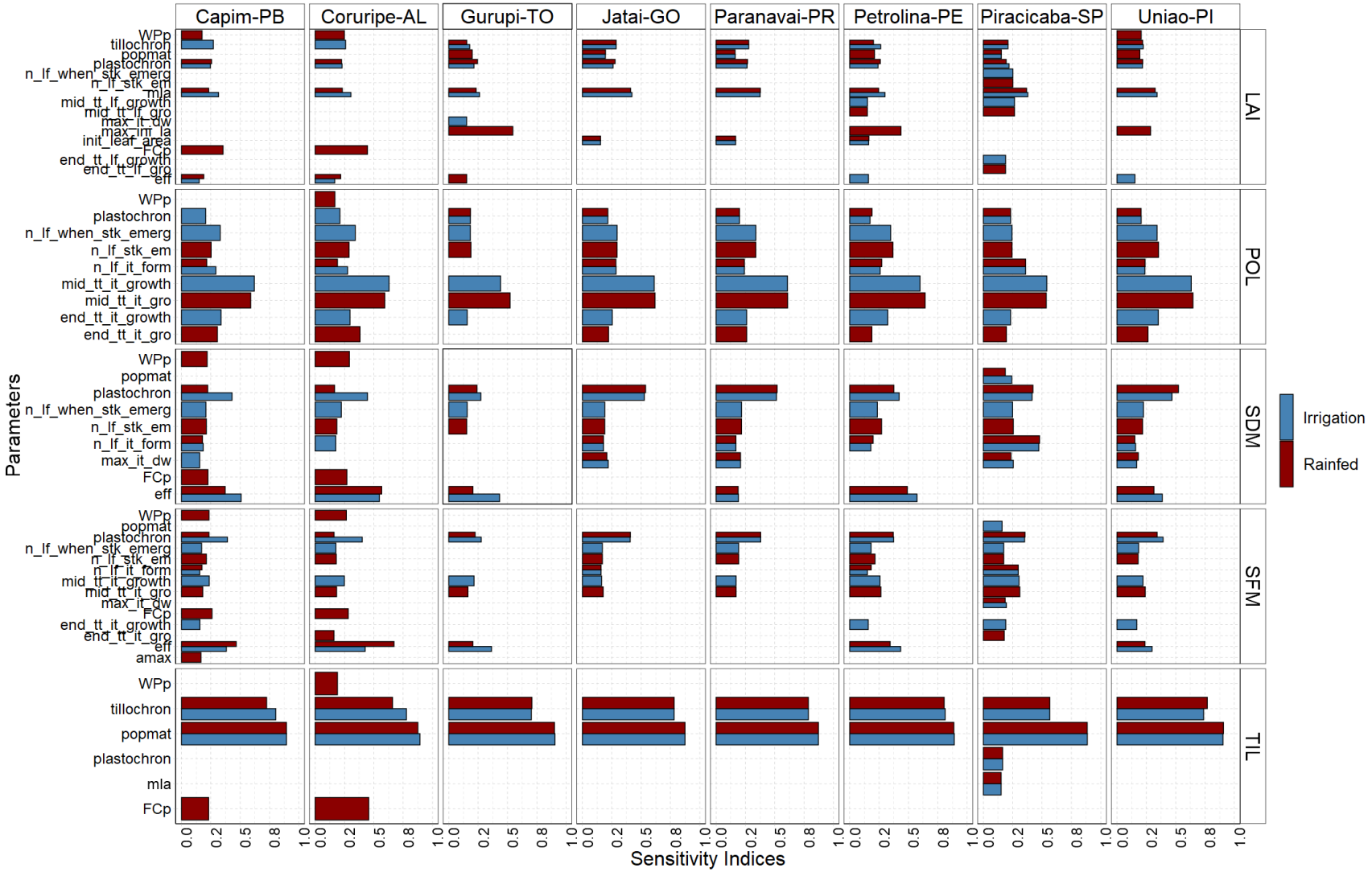


Figure 22. Average of main sensitivity index (Si) time-series in different site for leaf area index (LAI), sucrose concentration (POL) stalk dry (SDM) and fresh (SFM) mass and tillering the treatment irrigation and rainfed

Table 21. Mean values of temperature (TAR), solar radiation (SRAD), and rainfall (RAIN)during the cycle and the parameters responsible for explaining the largest variance in for the largest variance in for stalk fresh (SFM) and stalk dry mass (SDM), leaf area index (LAI), tillering (TIL), and sucrose content (POL).

SITE	TAR	SRAD	RAIN	LAI		POL		SDM		SFM		TIL	
	(°C)	($MJ.m^{-2}. d^{-1}$)	(mm)	irrigation	rainfed	irrigation	rainfed	irrigation	rainfed	irrigation	rainfed	irrigation	rainfed
Capim-PB	25.8	20.1	1440	mla	fcp	mid_tt_it_growth	mid_tt_it_growth	eff	eff	eff	eff	popmat	popmat
Coruripe-AL	25.9	22.3	947	mla	fcp	mid_tt_it_growth	mid_tt_it_growth	eff	eff	eff	eff	popmat	popmat
Gurupi-TO	29.9	20.9	887	mla	max_ini_lai	mid_tt_it_growth	mid_tt_it_growth	eff	eff	eff	eff	popmat	popmat
Jataí-GO	22.8	18.8	1437	mla	mla	mid_tt_it_growth	mid_tt_it_growth	plastochron	plastochron	plastochron	plastochron	popmat	popmat
Paranavaí-PR	22.4	18.0	1142	mla	mla	mid_tt_it_growth	mid_tt_it_growth	plastochron	plastochron	plastochron	plastochron	popmat	popmat
Petrolina-PE	27.0	20.4	420	mla	max_ini_lai	mid_tt_it_growth	mid_tt_it_growth	eff	eff	eff	eff	popmat	popmat
Piracicaba-SP	21.5	18.0	1450	mla	mla	mid_tt_it_growth	mid_tt_it_growth	plastochron	plastochron	plastochron	plastochron	popmat	popmat
União-PI	26.8	20.1	1530	mla	mla	mid_tt_it_growth	mid_tt_it_growth	plastochron	plastochron	plastochron	plastochron	popmat	popmat

4.4. Discussion

The results obtained by the two GSA methods were similar. In both cases, the main parameters were indicated by the two methods, as also observed by (Marino et al., 2008). The combination of these two methods made it possible to evaluate the model's variance throughout the simulation and to identify the correlation of the parameters on the target variables. In GSA studies, some studies pointed out the importance of using more than one method to validate their results (Marino et al., 2008; Wang and Solomatine, 2019). The combination of PRCC and eFAST methods provided a more robust analysis on the uncertainty of the SAMUCA model, and in that sense, it is more suitable to use different and not similar methods to evaluate uncertainties in a crop model. For example, FAST and Sobol are methods focused on obtaining the parameters responsible for explaining the highest variance of the model (Wei, 2013), i.e., applied together, they serve to validate their results and will not probably provide additional information. Moreover, eFAST requires more computational time than PRCC (Vazquez-Cruz et al., 2014; Xing et al., 2017), which should be considered in future studies, since the results were similar for the SAMUCA. Until then, the survey by Marino et al., (2008) was the only one to compare PRCC and eFAST and, like our results, showed agreement between these two methods. The PRCC, on the other hand, was more operationally efficient because of the lower computational cost, but its results depend on the monotonicity relationship between the parameter and an output variable (Krishnan and Aggarwal, 2018; Pereira et al., 2021). Thus, the application of GSA in crop models should consider the limitations of each method, the objectives of the study, and especially the computational cost involved.

Our results showed that the boundary conditions had some influence on the sensitivity indices, being as important to the results as the GSA method adopted (Sexton et al., 2017; Varella et al., 2010; Zhang et al., 2020; Zhao et al., 2014). Regarding soil parameters in irrigated environments, we did not identify any influence on the SAMUCA model output, similarly as reported by Pereira et al, (2021). Under rainfed conditions and in soils with more than 100 cm of depth, in the same way there was no influence of soil parameters. This low influence of soil parameters on plant growth and development variables had already been reported in other models under rainfed conditions, such as STICS (Varella et al., 2012), InfoCrop Wheat (Krishnan and Aggarwal, 2018), and DSSAT-CERES (LI et al., 2019). However, Dejong et al., (2012) identified that soil hydraulic parameters, such as field capacity, permanent wilting point, hydraulic conductivity, and saturation point, were significantly important for other variables, such as transpiration, evaporation, and evapotranspiration for CERES-Maize, and should be considered in future studies on GSA to confirm our results in the SAMUCA model. Finally, soil depth appeared to be more relevant in rainfed conditions than the parameters of field capacity and permanent wilting point. This can be explained as such variables have a greater influence in determining the rooting depth and available water for the crop. The *plastochron*, which is the time interval between the appearance of two successive phytomers (leaf+node), and the *eff*, which is related to the CO2 assimilation rate, were the most important parameters for SDM and SFM; and both parameters had a direct linear relationship with these output variables. In environments without restricted water but adequate levels of temperature and solar radiation, the crop grows with minimum limitation and that was the reason for such parameters appear as relevant in such conditions (Inman-Bamber and Smith, 2005; Marin et al., 2014; Stokes et al., 2016) (Table 21). Interestingly, even under irrigated conditions, in regions with high average temperature and high solar irradiance, combined with shallow soils, the *eff* was the main parameter, followed closely by *plastochron* (Figure 22). We hypothesize that in such conditions, together with transient water stress, the high crop growth rate made those parameters crucial for simulation (Inman-Bamber and Smith, 2005; Stokes et al., 2016).

Interesting to note that in União-PI, with high air temperatures and solar radiation, *plastochron* was the main parameter for SDM and SFM, instead of *eff*, and this could be explained as the deeper soil increases the water storage capacity for the crop. As a result, the frequency of soil irrigation was lower compared to that in Capim-PB, Coruripe-TO, Gurupi-TO, and Petrolina-PE, indicating transient water stress in these sites.

The LAI variable can be considered the variable with the greatest uncertainty due to the large number of influential parameters in both simulated water treatments (Figure 22). It is a consensus that the LAI variable is one of the most complex for simulation, constantly having the lowest statistical performance indexes between simulated data and observed data when compared to other model output variables (Marin et al., 2015b; O'Leary, 2000; Vianna et al., 2020). This could be because the PBCMs are not yet ready to account for all the processes shaping canopy formation, which is generally simplified to a per-area index (e.g., LAI) and fixed parameters (e.g., light extinction coefficient). Another reason is that PBCMs still not properly account for all reduction factors in yield simulations due to the high complexity in measuring the various biotic and abiotic interactions, which are usually accounted as a yield gap factor (Fattori Junior et al., 2022; Lobell et al., 2009; Van Ittersum et al., 2003). This latent empiricism may result in increased number of input parameters to compensate for these yield gaps, which in one hand is important to correct model simulations but, in most cases, end up increasing the uncertainty in the simulation (Gan et al., 2014; Varella et al., 2012). For example, in environments without water restriction, the *mla* was the parameter responsible for explaining the highest variance of LAI, but in places with high water restriction, the *FCp* parameter has explained the highest variance (Table 21). This shows that different parameters, *mla* a growth-related parameter, and *FCp* a soil parameter, can cause significant variations in the simulation. The importance of the *plastochron* parameter should also be highlighted. Even though this parameter did not explain the higher variance of the LAI output, it was the only one that alternated its response throughout the simulation (Figures 16 and 19). In both irrigated and rainfed treatments, the *plastochron* has an inverse linear relationship to LAI until the simulation reaches the maximum LAI values, with a direct linear relationship. Such results show that this variable is extremely complex, and its uncertainty should deserve further efforts in future studies.

Finally, our study analyzed the uncertainty of the SAMUCA model in different sugarcane production environments in Brazil. However, we did not consider the influence of long-term climate variability in the analysis. In addition, it would be interesting in future studies to evaluate soil parameters in rainfed environments about other output variables, such as evapotranspiration. We found this particularly important as our results indicated that the relevance of soil hydraulic parameters, such as *FCp, WPp, Ksat*, and *STp*, seemed to have less importance than the value attributed to soil depth. This information could be quite relevant since in practice we could focus on measuring only one piece of information in the field very accurately instead of several soil parameters altogether. In addition, these results may be linked to the empirical method of soil water balance. Like many PBCMs, SAMUCA uses the tipping bucket method to calculate the water balance in the soil, which can function well under monotonic drying conditions, but may not match the rates and patterns of water uptake by the crop during complex wetting and drying cycles in the field (Jarvis et al., 2022).

4.5. Conclusion

A total of 31 parameters were analyzed, 24 of which were genotype and 7 from the soil, and we concluded that only 13 parameters were significant, regardless of the output variable, climate, soil, and water treatment. They were, not necessarily in this order: *mla, tillochron, max_ini_la, popmat, eff, max_it_dw, mid_tt_lf_gro, init_leaf_area,*

n_lf_it_form, n_lf_when_stk_em, plastochron, FCp and *WPp*. Furthermore, we confirmed that the climate affected the main parameter only for the SDM and SFM variables, with the *plastochron* and *eff* parameters being the most important for these two variables. In environments with well-distributed rainfall, located in the Central part or in the Southern Brazil, the *plastochron* was the main parameter, while in hotter environments with lower soil water storage and so subject to higher water stress, the *eff* was the most important parameter. Regarding the soil parameters, we noticed that no soil parameter was important for the irrigated treatment, but in rainfed conditions the *FCp* and *WPp* were relevant in environments with poor rainfall distribution and shallow soils. Even in rainfed environments but with higher amounts and well-distributed rainfall and the deeper soils, soil parameters were less important for the SAMUCA model. Finally, soil depth may be the most important soil parameter for water dynamics, but we did not include it in the sensitivity analysis, and this is an important open question for future study.

References

Arruda, M.R., Giller, K.E., Slingerland, M., 2017. Where is sugarcane cropping expanding in the brazilian cerrado, and why? A case study. An. Acad. Bras. Cienc. 89, 2485–2493. https://doi.org/10.1590/0001-3765201720150260

B Péné, C., 2019. Screening of Fifteen Sugarcane Genotypes under Irrigation based on Genetic Variations and Heritability for Agro-morphological Traits at Early Selection Stage in Ferké, Ivory Coast. Am. J. Biomed. Sci. Res. 4, 362–371. https://doi.org/10.34297/AJBSR.2019.04.000835

Baba, K., Shibata, R., Sibuya, M., 2004. PARTIAL CORRELATION AND CONDITIONAL CORRELATION AS MEASURES OF CONDITIONAL INDEPENDENCE. Aust. N. Z. J. Stat. 46, 657–664. https://doi.org/10.1111/J.1467-842X.2004.00360.X

Batjes, N.H., 2009. Harmonized soil profile data for applications at global and continental scales: Updates to the WISE database. Soil Use Manag. 25, 124–127. https://doi.org/10.1111/J.1475-2743.2009.00202.X

Béhou, Y., Péné, C., 2020. Genetic Variability and Heritability among Sugarcane Genotypes at Early Stage of the Advanced Selection for some Agronomic Traits in Ferké, Northern Ivory Coast. Agric. Sci. 2, p83–p83. https://doi.org/10.30560/AS.V2N1P83

Bezuidenhout, C., 2000. A model review and proposed mechanistic tiller model for the canegro sugarcane crop model. undefined. Faculty of Engineering at Technikon Natal, Durban.

Coelho, A.P., Dalri, A.B., Fischer Filho, J.A., de Faria, R.T., Silva, L.S., Gomes, R.P., 2020. Calibration and evaluation of the DSSAT/Canegro model for sugarcane cultivars under irrigation managements. Rev. Bras. Eng. Agric. e Ambient. 24, 52–58. https://doi.org/10.1590/1807-1929/agriambi.v24n1p52-58

CONAB, 2022. Acompanhamento da safra brasileira de cana-de-açúcar. Brasília.

Cukier, R.I., Levine, H.B., Shuler, K.E., 1978. Nonlinear sensitivity analysis of multiparameter model systems. J. Comput. Phys. https://doi.org/10.1016/0021-9991(78)90097-9

Cursi, D.E., Hoffmann, H.P., Barbosa, G.V.S., Bressiani, J.A., Gazaffi, R., Chapola, R.G., Fernandes Junior, A.R., Balsalobre, T.W.A., Diniz, C.A., Santos, J.M., Carneiro, M.S., 2022. History and Current Status of Sugarcane Breeding, Germplasm Development and Molecular Genetics in Brazil. Sugar Tech 24, 112–133. https://doi.org/10.1007/S12355-021-00951-1/TABLES/6

Dejonge, K.C., Ascough Ii, J.C., Ahmadi, M., Andales, A.A., Arabi, M., 2012. Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. Ecol. Modell. 231, 113–125. https://doi.org/10.1016/j.ecolmodel.2012.01.024

Ehara, H., Takaruma, T., Tsuchiya, M., 1994. Growth and dry matter production of sugar cane in warm temperate zone of Japan. Japanese J. Trop. Agric. 38, 335–342.

Fattori Junior, I.M., dos Santos Vianna, M., Marin, F.R., 2022. Assimilating leaf area index data into a sugarcane process-based crop model for improving yield estimation. Eur. J. Agron. 136, 126501. https://doi.org/10.1016/J.EJA.2022.126501

Gan, Y., Duan, Q., Gong, W., Tong, C., Sun, Y., Chu, W., Ye, A., Miao, C., Di, Z., 2014. A comprehensive evaluation of various sensitivity analysis methods: A case study with a hydrological model. Environ. Model. Softw. 51, 269–285. https://doi.org/10.1016/j.envsoft.2013.09.031

Gasparotto, L.G., Rosa, J.M., Marin, F.R., 2020. Interrow spacing and sugarcane yield in a diversity of climates: A major review. Agron. J. 112, 4550–4557. https://doi.org/10.1002/AGJ2.20425

Gijsman, A.J., Thornton, P.K., Hoogenboom, G., 2007. Using the WISE database to parameterize soil inputs for crop simulation models. Comput. Electron. Agric. 56, 85–100. https://doi.org/10.1016/j.compag.2007.01.001

He, J., M. D. Dukes, M.D., J. W. Jones, J.W., W. D. Graham, W.D., J. Judge, J., 2009. Applying GLUE for Estimating CERES-Maize Genetic and Soil Parameters for Sweet Corn Production. Trans. ASABE 52, 1907–1921. https://doi.org/10.13031/2013.29218

Inman-Bamber, N.G., 1994. Temperature and seasonal effects on canopy development and light interception of sugarcane. F. Crop. Res. 36, 41–51. https://doi.org/10.1016/0378-4290(94)90051-5

Inman-Bamber, N.G., Lakshmanan, P., Park, S., 2012. Sugarcane for water-limited environments: Theoretical assessment of suitable traits. F. Crop. Res. 134, 95–104. https://doi.org/10.1016/J.FCR.2012.05.004

Inman-Bamber, N.G., Smith, D.M., 2005. Water relations in sugarcane and response to water deficits. F. Crop. Res. 92, 185–202. https://doi.org/10.1016/J.FCR.2005.01.023

Jarvis, N., Larsbo, M., Lewan, E., Garré, S., 2022. Improved descriptions of soil hydrology in crop models: The elephant in the room? Agric Syst 202. https://doi.org/10.1016/J.AGSY.2022.103477

Jeuffroy, M.H., Barbottin, A., Jones, J.W., Leoeur, J., 2006. Crop models with genotype parameters, in: Working with Crop Models. pp. 281–308.

Jin, X., Li, Z., Nie, C., Xu, X., Feng, H., Guo, W., Wang, J., 2018. Parameter sensitivity analysis of the AquaCrop model based on extended fourier amplitude sensitivity under different agro-meteorological conditions and application. F. Crop. Res. 226, 1–15. https://doi.org/10.1016/j.fcr.2018.07.002

Krishnan, P., Aggarwal, P., 2018. Global sensitivity and uncertainty analyses of a web based crop simulation model (web InfoCrop wheat) for soil parameters. Plant Soil 423, 443–463. https://doi.org/10.1007/s11104-017-3498-0

Leal, M.R.L.V., Galdos, M. V., Scarpare, F. V., Seabra, J.E.A., Walter, A., Oliveira, C.O.F., 2013. Sugarcane straw availability, quality, recovery and energy use: A literature review. Biomass and Bioenergy 53, 11–19. https://doi.org/10.1016/j.biombioe.2013.03.007

Lingle, S.E., 1999. Sugar Metabolism during Growth and Development in Sugarcane Internodes. Crop Sci. 39, 480–486. https://doi.org/10.2135/cropsci1999.0011183X0039000200030x

Liu, D.L., Kingston, G., Bull, T.A., 1998. A new technique for determining the thermal parameters of phenological development in sugarcane, including suboptimum and supra-optimum temperature regimes. Agric. For. Meteorol. 90, 119–139. https://doi.org/10.1016/S0168-1923(97)00087-7

Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop Yield Gaps: Their Importance, Magnitudes, and Causes. http://dx.doi.org/10.1146/annurev.environ.041008.093740 34, 179–204. https://doi.org/10.1146/ANNUREV.ENVIRON.041008.093740

Marin, F.R., Jones, J.W., 2014. Process-based simple model for simulating sugarcane growth and production. Sci. Agric. 71, 1–16. https://doi.org/10.1590/S0103-90162014000100001

Marin, F.R., Jones, J.W., Boote, K.J., 2017. A stochastic method for crop models: including uncertainty in a sugarcane model. Agron. J. 109, 483–495. https://doi.org/https://doi.org/10.2134/agronj2016.02.0103

Marin, F.R., Jones, J.W., Royce, F., Suguitani, C., Donzeli, J.L., Filho, W.J.P., Nassif, D.S.P., 2011. Parameterization and evaluation of predictions of DSSAT/CANEGRO for Brazilian sugarcane. Agron. J. 103, 304–315. https://doi.org/10.2134/agronj2010.0302

Marin, F.R., Ribeiro, R. V., Marchiori, P.E.R., 2014. How can crop modeling and plant physiology help to understand the plant responses to climate change? A case study with sugarcane. Theor. Exp. Plant Physiol. 26, 49–63. https://doi.org/10.1007/S40626-014-0006-2/FIGURES/5

Marin, F.R., Thorburn, P.J., Nassif, D.S.P., Costa, L.G., 2015. Sugarcane model intercomparison: Structural differences and uncertainties under current and potential future climates. Environ. Model. Softw. 72, 372–386. https://doi.org/10.1016/j.envsoft.2015.02.019

Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E., 2008. A methodology for performing global uncertainty and sensitivity analysis in systems biology. J. Theor. Biol. 254, 178–196. https://doi.org/10.1016/j.jtbi.2008.04.011

Maulidiani, Rudiyanto, Abas, F., Ismail, I.S., Lajis, N.H., 2018. Generalized Likelihood Uncertainty Estimation (GLUE) methodology for optimization of extraction in natural products. Food Chem. 250, 37–45. https://doi.org/10.1016/j.foodchem.2018.01.023

Morris, M.D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33, 161. https://doi.org/10.2307/1269043

Mukaka, M.M., 2012. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med. J. 24, 69–71.

Nassif, D.S.P., Marin, F.R., Pallone Filho, W.J., Resende, R.S., Pellegrino, G.Q., Antonio, S., 2012. Parametrização e avaliação do modelo DSSAT/Canegro para variedades brasileiras de cana-de-açúcar Parameterization and evaluation of the DSSAT/Canegro model for Brazilian sugarcane varieties, Pesq. agropec. bras. https://doi.org/https://doi.org/10.1590/S0100-204X2012000300001

Natarajan, S., Basnayake, J., Lakshmanan, P., Fukai, S., 2020. Limited contribution of water availability in genotype-by-environment interaction in sugarcane yield and yield components. J. Agron. Crop Sci. 206, 665–678. https://doi.org/10.1111/JAC.12407

O'Leary, G.J., 2000. A review of three sugarcane simulation models with respect to their prediction of sucrose yield. F. Crop. Res. 68, 97–111. https://doi.org/10.1016/S0378-4290(00)00112-X

Pereira, R.A. de A., Vianna, M. dos S., Nassif, D.S.P., Carvalho, K. dos S., Marin, F.R., 2021. Global sensitivity and uncertainty analysis of a sugarcane model considering the trash blanket effect [WWW Document]. Eur. J. Agron. https://doi.org/https://doi.org/10.1016/j.eja.2021.126371

RIDESA, R.I. para o D.S., 2018. Censo varietal Brasil safra 2017/2018 [WWW Document]. URL https://www.ridesa.com.br/censo-varietal (accessed 10.22.21).

Sage, R.F., Peixoto, M.M., Sage, T.L., 2013. Photosynthesis in Sugarcane, Sugarcane: Physiology, Biochemistry, and Functional Biology. https://doi.org/10.1002/9781118771280.ch6

Saltelli, A., Tarantola, S., Chan, P.-S., 1999. A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. Technometrics 41, 39–56. https://doi.org/10.1080/00401706.1999.10485594

Sexton, J., Everingham, Y.L., Inman-Bamber, G., 2017. A global sensitivity analysis of cultivar trait parameters in a sugarcane growth model for contrasting production environments in Queensland, Australia. Eur. J. Agron. 88, 96–105. https://doi.org/10.1016/j.eja.2015.11.009

Silva, M. de A., Jifon, J.L., dos Santos, C.M., Jadoski, C.J., da Silva, J.A.G., 2013. Photosynthetic capacity and water use efficiency in sugarcane genotypes subject to water deficit during early growth phase. Brazilian Arch. Biol. Technol. 56, 735–748. https://doi.org/10.1590/S1516-89132013000500004

Silvestro, P.C., Pignatti, S., Yang, H., Yang, G., Pascucci, S., Castaldi, F., Casa, R., 2017. Sensitivity analysis of the Aquacrop and SAFYE crop models for the assessment of water limited winter wheat yield in regional scale applications. PLoS One 12. https://doi.org/10.1371/JOURNAL.PONE.0187485

Singels, A., Bezuidenhout, C.N., 2002. A new method of simulating dry matter partitioning in the Canegro sugarcane model. F. Crop. Res. 78, 151–164. https://doi.org/https://doi.org/10.1016/S0378-4290(02)00118-1

Smit, M.A., Singels, A., 2006. The response of sugarcane canopy development to water stress. F. Crop. Res. 98, 91–97. https://doi.org/10.1016/j.fcr.2005.12.009

Soares, F.A.A., Marin, F.R., 2021. Crop-specific technology extrapolation domains for Brazil. Bragantia 80, 1–14. https://doi.org/10.1590/1678-4499.20200310

Sobol, I.M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math. Comput. Simul. 55, 271–280. https://doi.org/10.1016/S0378-4754(00)00270-6

Stokes, C.J., Inman-Bamber, N.G., Everingham, Y.L., Sexton, J., 2016. Measuring and modelling CO2 effects on sugarcane. Environ. Model. Softw. 78, 68–78. https://doi.org/10.1016/J.ENVSOFT.2015.11.022

Van Ittersum, M.K., Leffelaar, P.A., Van Keulen, H., Kropff, M.J., Bastiaans, L., Goudriaan, J., 2003. On approaches and applications of the Wageningen crop models. Eur. J. Agron. 18, 201–234. https://doi.org/10.1016/S1161-0301(02)00106-5

Varella, H., Buis, S., Launay, M., Guérif, M., Varella, H., Buis, S., Launay, M., Guérif, M., 2012. Global sensitivity analysis for choosing the main soil parameters of a crop model to be determined. Agric. Sci. 3, 949–961. https://doi.org/10.4236/AS.2012.37116

Varella, H., Guérif, M., Buis, S., 2010. Global sensitivity analysis measures the quality of parameter estimation: The case of soil parameters and a crop model. Environ. Model. Softw. 25, 310–319. https://doi.org/https://doi.org/10.1016/j.envsoft.2009.09.012

Vazquez-Cruz, M.A., Guzman-Cruz, R., Lopez-Cruz, I.L., Cornejo-Perez, O., Torres-Pacheco, I., Guevara-Gonzalez, R.G., 2014. 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. https://doi.org/10.1016/J.COMPAG.2013.10.006

Vianna, M. dos S., Nassif, D.S.P., dos Santos Carvalho, K., Marin, F.R., 2020. Modelling the trash blanket effect on sugarcane growth and water use. Comput. Electron. Agric. 172, 105361. https://doi.org/10.1016/j.compag.2020.105361

Wallach, D., Makowski, D., Jones, J.W., Brun, F., 2018. Working with dynamic crop models: methods, tools and examples for agriculture and environment. Academic Press.

Wang, A., Solomatine, D.P., 2019. Practical experience of sensitivity analysis: Comparing six methods, on three hydrological models, with three performance criteria. Water (Switzerland) 11, 1–26. https://doi.org/10.3390/w11051062

Wang, J., Li, X., Lu, L., Fang, F., 2013. Parameter sensitivity analysis of crop growth models based on the extended Fourier Amplitude Sensitivity Test method. Environ. Model. Softw. 48, 171–182. https://doi.org/10.1016/j.envsoft.2013.06.007

Wei, T., 2013. A review of sensitivity analysis methods in building energy analysis. Renew. Sustain. Energy Rev. https://doi.org/10.1016/j.rser.2012.12.014

Xing, H. min, Xu, X. gang, Li, Z. hai, Chen, Y. jin, Feng, H. kuan, Yang, G. jun, Chen, Z. xia, 2017. Global sensitivity analysis of the AquaCrop model for winter wheat under different water treatments based on the extended Fourier amplitude sensitivity test. J. Integr. Agric. 16, 2444–2458. https://doi.org/10.1016/S2095-3119(16)61626-X

Zhang, R.-Y., Shan, H.-L., Yang, K., Wang, X.-Y., Cang, X.-Y., Wang, C.-M., Yin, J., Luo, Z.-M., Li, W.-F., Huang, Y.-K., 2021. Identification of brown rust resistance in the field and detection of the Bru1 gene in sugarcane varieties ARTICLE. Crop Breed. Appl. Biotechnol. 21, 32542121. https://doi.org/10.1590/1984

Zhang, Y., Arabi, M., Paustian, K., 2020. Analysis of parameter uncertainty in model simulations of irrigated and rainfed agroecosystems. Environ. Model. Softw. 126. https://doi.org/10.1016/j.envsoft.2020.104642

Zhao, D., Li, Y.R., 2015. Climate Change and Sugarcane Production: Potential Impact and Mitigation Strategies. Int. J. Agron. 2015. https://doi.org/10.1155/2015/547386

Zhao, G., Bryan, B.A., Song, X., 2014. Sensitivity and uncertainty analysis of the APSIM-wheat model: Interactions between cultivar, environmental, and management parameters. Ecol. Modell. 279, 1–11. https://doi.org/10.1016/J.ECOLMODEL.2014.02.003

Zhou, M., Singels, A., Savage, M.J., 2003. Physiological parameters for modelling differences in canopy development between sugarcane cultivars. Proc. South African Sugar Technol. Assoc. 77, 610–621.

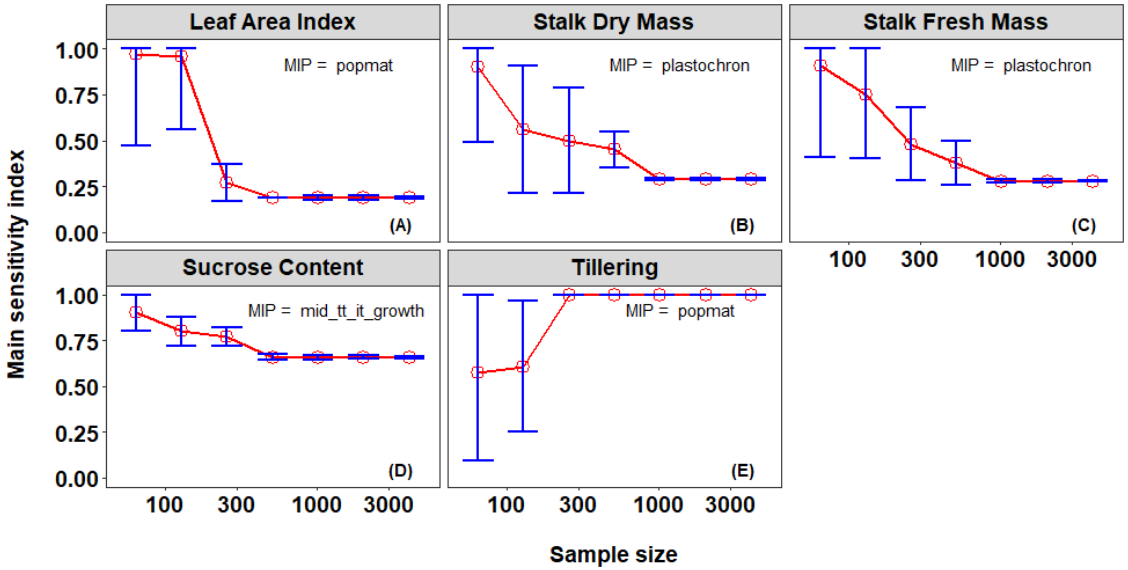
Zhou, M.M., Shoko, M.D., 2011. Seasonal and varietal effects on tiller population development of sugarcane (Saccharum Officinarum L.). South African J. Plant Soil 28, 11–16. https://doi.org/10.1080/02571862.2011.10640007

APPENDICES

APPENDIX A. Maximum (max), minimum (min), and average (avg) values of the soil physical parameters, obtained from the non-parametric regression as a function of depth.

DEPTH (cm)	WWp ($m^3\ m^{-3}$)			FCp ($m^3\ m^{-3}$)			STp ($m^3\ m^{-3}$)			Ksat ($cm\ h^{-1}$)			Pclay ($g\ g^{-1}$)			Psilt (g g)			Psand (g g)		
	min	avg	max	min	avg	max	min	avg	max	min	avg	max	min	avg	max	min	avg	max	min	avg	max
Ferralic Arenosol																					
5	0.048	0.071	0.093	0.162	0.192	0.222	0.407	0.426	0.445	9.183	10.311	11.439	0.094	0.109	0.124	0.073	0.130	0.188	0.708	0.761	0.813
15	0.048	0.069	0.091	0.159	0.188	0.217	0.403	0.422	0.441	8.977	10.072	11.167	0.094	0.109	0.124	0.073	0.129	0.185	0.711	0.762	0.812
25	0.047	0.068	0.089	0.156	0.185	0.213	0.399	0.418	0.436	8.772	9.833	10.895	0.095	0.109	0.124	0.074	0.128	0.182	0.714	0.763	0.812
35	0.047	0.067	0.087	0.153	0.181	0.208	0.396	0.413	0.431	8.567	9.595	10.622	0.096	0.110	0.124	0.074	0.126	0.179	0.716	0.764	0.811
45	0.046	0.066	0.085	0.151	0.177	0.203	0.392	0.409	0.426	8.361	9.356	10.350	0.096	0.110	0.124	0.074	0.125	0.176	0.719	0.765	0.811
50	0.046	0.065	0.084	0.149	0.175	0.201	0.390	0.407	0.424	8.259	9.236	10.214	0.097	0.110	0.124	0.075	0.124	0.174	0.720	0.765	0.811
60	0.045	0.064	0.083	0.146	0.171	0.196	0.387	0.403	0.419	8.053	8.998	9.942	0.098	0.110	0.123	0.075	0.123	0.171	0.723	0.766	0.810
70	0.045	0.063	0.081	0.143	0.168	0.192	0.383	0.399	0.414	7.848	8.759	9.670	0.098	0.111	0.123	0.075	0.122	0.168	0.725	0.767	0.810
Haplic Acrisol																					
5	0.030	0.064	0.097	0.162	0.206	0.250	0.415	0.434	0.452	8.215	10.856	13.498	0.112	0.166	0.219	0.092	0.134	0.176	0.623	0.701	0.778
15	0.055	0.088	0.120	0.177	0.220	0.264	0.407	0.424	0.442	6.717	9.317	11.918	0.134	0.186	0.239	0.091	0.132	0.174	0.605	0.682	0.758
25	0.079	0.111	0.144	0.192	0.235	0.277	0.398	0.415	0.433	5.219	7.778	10.337	0.155	0.207	0.258	0.090	0.131	0.172	0.587	0.663	0.738
30	0.091	0.123	0.155	0.200	0.242	0.284	0.393	0.411	0.428	4.471	7.009	9.547	0.165	0.217	0.268	0.090	0.130	0.171	0.579	0.653	0.728
35	0.103	0.135	0.167	0.207	0.249	0.291	0.389	0.406	0.423	3.722	6.239	8.757	0.176	0.227	0.278	0.089	0.130	0.170	0.570	0.644	0.718
45	0.128	0.159	0.190	0.222	0.263	0.304	0.380	0.397	0.414	2.224	4.701	7.177	0.197	0.247	0.297	0.089	0.128	0.168	0.552	0.625	0.697
50	0.140	0.171	0.202	0.229	0.270	0.311	0.375	0.392	0.409	1.475	3.931	6.387	0.208	0.257	0.307	0.088	0.127	0.167	0.543	0.615	0.687
55	0.152	0.183	0.213	0.237	0.277	0.318	0.371	0.388	0.404	0.726	3.162	5.597	0.218	0.268	0.317	0.088	0.127	0.166	0.534	0.606	0.677
Haplic Ferrosol																					
5	0.196	0.211	0.226	0.350	0.366	0.382	0.473	0.494	0.515	0.468	0.570	0.672	0.468	0.547	0.627	0.148	0.193	0.237	0.215	0.260	0.305
15	0.197	0.212	0.226	0.349	0.365	0.380	0.472	0.493	0.513	0.468	0.569	0.670	0.473	0.551	0.630	0.149	0.193	0.237	0.211	0.255	0.300
30	0.199	0.213	0.228	0.347	0.363	0.378	0.470	0.491	0.511	0.469	0.568	0.667	0.480	0.557	0.634	0.151	0.194	0.238	0.205	0.249	0.293
45	0.200	0.215	0.229	0.346	0.361	0.376	0.469	0.488	0.508	0.469	0.566	0.663	0.487	0.562	0.638	0.153	0.195	0.238	0.199	0.242	0.285
60	0.202	0.216	0.230	0.344	0.359	0.374	0.467	0.486	0.506	0.470	0.565	0.660	0.494	0.568	0.643	0.154	0.196	0.238	0.193	0.236	0.278
75	0.204	0.218	0.231	0.343	0.357	0.372	0.465	0.484	0.503	0.470	0.564	0.657	0.501	0.574	0.647	0.156	0.197	0.238	0.188	0.229	0.271
90	0.205	0.219	0.233	0.341	0.355	0.370	0.463	0.482	0.501	0.470	0.562	0.654	0.508	0.579	0.651	0.158	0.198	0.238	0.182	0.223	0.263
120	0.209	0.222	0.235	0.338	0.352	0.366	0.460	0.478	0.496	0.471	0.560	0.648	0.522	0.591	0.660	0.161	0.200	0.239	0.170	0.209	0.249
Humic Plinthosol																					
5	0.254	0.296	0.338	0.411	0.441	0.471	0.497	0.538	0.580	0.241	0.360	0.479	0.468	0.498	0.527	0.268	0.323	0.378	0.141	0.179	0.217
15	0.247	0.290	0.333	0.404	0.434	0.465	0.492	0.535	0.577	0.257	0.378	0.500	0.474	0.505	0.535	0.258	0.314	0.370	0.142	0.181	0.220
30	0.237	0.281	0.326	0.393	0.424	0.455	0.485	0.529	0.573	0.280	0.405	0.530	0.484	0.515	0.546	0.243	0.301	0.358	0.144	0.184	0.224
45	0.227	0.273	0.318	0.382	0.414	0.446	0.478	0.523	0.568	0.303	0.432	0.561	0.493	0.525	0.557	0.228	0.287	0.347	0.146	0.188	0.229
60	0.217	0.264	0.311	0.370	0.404	0.437	0.471	0.518	0.564	0.327	0.459	0.592	0.503	0.536	0.569	0.212	0.274	0.335	0.148	0.191	0.233
75	0.207	0.256	0.304	0.359	0.393	0.427	0.464	0.512	0.560	0.350	0.486	0.622	0.512	0.546	0.580	0.197	0.260	0.323	0.150	0.194	0.238
90	0.197	0.247	0.297	0.348	0.383	0.418	0.457	0.506	0.555	0.373	0.513	0.653	0.522	0.556	0.591	0.182	0.246	0.311	0.152	0.197	0.242
120	0.177	0.230	0.282	0.326	0.363	0.400	0.443	0.495	0.547	0.419	0.567	0.714	0.541	0.577	0.614	0.151	0.219	0.287	0.156	0.204	0.251

APPENDIX B. Evolution of sensitivity index of the most important parameter (MIP) with increasing sample size for variables leaf area index (A), mass stalk dry (B) and fresh (C), sucrose content (D), and tillering (E) for bare; red line is the average of the 10 simulations for each sample size and in blue in blue we have the max and min S_i of each sample size



APPENDIX C. Relative value of the model output variance (σ^2) explained individually by each parameter, and the variance sum ($\sum \sigma^2$) of the parameters; we only considered the parameters that presented S_i >0.05 and sample size of 2049 in treatment Bare and GCTB.

Variable	Bare				GCTB			
	Parameters	σ^2	Rank	$\sum \sigma^2$	Parameters	σ^2	Rank	$\sum \sigma^2$
SDM	*plastochron*	29.0%	1°	29.0%	*plastochron*	31.6%	1°	31.6%
	max_it_dw	17.2%	2°	46.2%	*max_it_dw*	14.0%	2°	45.5%
	n_lf_it_form	10.6%	3°	56.8%	*n_lf_when_stk_emerg*	12.7%	3°	58.3%
	eff	10.5%	4°	67.3%	*n_lf_it_form*	10.4%	4°	68.7%
	popmat	8.9%	5°	76.2%	*eff*	8.0%	5°	76.7%
	n_lf_when_stk_emerg	5.4%	6°	81.6%	*popmat*	5.5%	6°	82.2%
SFM	*plastochron*	27.9%	1°	27.9%	*plastochron*	30.4%	1°	30.4%
	mid_tt_it_growth	21.1%	2°	49.0%	*max_it_dw*	15.9%	2°	46.3%
	max_it_dw	18.9%	3°	67.9%	*mid_tt_it_growth*	15.9%	3°	62.2%
	end_tt_it_growth	12.7%	4°	80.6%	*n_lf_when_stk_emerg*	8.7%	4°	70.9%
	eff	9.0%	5°	89.6%	*end_tt_it_growth*	8.2%	5°	79.1%
	n_lf_it_form	7.1%	6°	96.8%	*eff*	7.2%	6°	86.2%
POL	*mid_tt_it_growth*	65.8%	1°	65.8%	*n_lf_it_form*	7.1%	7°	93.3%
	end_tt_it_growth	20.1%	2°	85.9%	*mid_tt_it_growth*	64.2%	1°	64.2%
TIL	*popmat*	99.8%	1°	99.8%	*end_tt_it_growth*	16.7%	2°	80.9%
LAI	*popmat*	18.8%	1°	18.8%	*popmat*	99.8%	1°	99.8%
	mla	15.1%	2°	33.9%	*popmat*	19.8%	1°	19.8%
					mla	14.8%	2°	34.6%