
**Multi-Objective and Multi-Variate Global
Sensitivity Analysis of the Soil-Crop Model
XN-CERES in Southwest Germany**

Dissertation to obtain the doctoral degree of Agricultural Sciences
(Dr. sc. agr.)

Faculty of Agricultural Sciences
Institute of Soil Science and Land Evaluation, Biogeophysics
University of Hohenheim

2021

Irene Witte
born in Aurich, Germany

This thesis was accepted as a doctoral dissertation in fulfilment of the requirements for the degree “Doktor der Agrarwissenschaften” by the Faculty of Agricultural Sciences at the University of Hohenheim on 23 March 2020.

Date of examination: 19 January 2021

Examination committee

Supervisor and reviewer: Prof. Dr. Thilo Streck

Co-reviewer: Prof. Dr. Volker Wulfmeyer

Additional examiner: Prof. Dr. Priesack

Vice-dean and chair: Prof. Dr. Andrea Knierim

Table of Contents

Table of Contents	3
List of Figures	7
List of Tables	15
I Introduction	19
1.1 Motivation	19
1.2 Objectives	23
II Material and Methods	27
2.1 Definitions	27
2.2 Site Description	28
2.3 Model Setup	30
2.4 Global Sensitivity Analysis	35
2.4.1 Classification and Subdivison	37
2.4.2 Notation	42
2.4.3 Variance-based Methods	42
2.4.4 Moment-Independent Methods	50
2.4.5 Confidence Intervals	60
2.4.6 CUSUNORO	60
2.5 Data Processing Chain	62
2.5.1 Data Process Chain - Data Preparation	63
2.5.2 Data Process Chain - Data Post-Processing	65

2.5.3	Summary and Analysis Settings	69
2.5.4	Implementation	70
III	Results	73
3.1	Model Output Description	73
3.2	GSA-Comparison	79
3.2.1	Convergence	79
3.2.2	Computing Resources	82
3.2.3	Index Comparison	84
3.3	GSA-Application	87
3.3.1	Model Additivity	88
3.3.2	Parameter Sensitivities	95
3.3.3	Trend Identification	122
3.3.4	Summary	126
IV	Discussion	131
4.1	Crop Group	131
4.1.1	Vegetative Growth	131
4.1.2	Generative Growth	137
4.1.3	Development Stage	139
4.2	Water Group	141
4.2.1	Soil Water Content	141
4.2.2	Matric Potential	143
4.3	Nitrogen Group	145
4.3.1	NO_3^- Content	145
4.3.2	NH_4^+ Content	147
4.4	Flux Group	148
4.5	Comparison of GSA methods	149
V	Conclusion	153
VI	Summary	157
VII	Zusammenfassung	161

References	165
Appendix	179

List of Figures

2.1	Summary of the model configuration and setup for the different submodules: nitrogen, water, crop, and flux with the group-specific target variables and parameters.	31
2.2	Schematic overview of the classification of global sensitivity methods	41
2.3	Scatter plots of the model output y against the two input factors x_1 and x_2 . The red dots represent the conditioned mean of the model output if x_1 (x_2) could be fixed within in sub-range of its original range. The sub-ranges are represented by the dotted vertical lines	46
2.4	Probability density functions for the model output daily transpiration on the x-axis and the corresponding probability on the y-axis for the parameters k_{cmid} and x with one partition.	52
2.5	Probability density functions (pdfs) of the model output daily transpiration (mm) for the parameters k_{cmid} and x for one partition. The area represents the integral of the difference between two shown pdfs.	53

2.6	Cumulative distribution functions of daily actual transpiration unconditional and conditional on parameters <i>kcmid</i> (left) and <i>x</i> (right). Legend entries refer to the maximum value of the specific partition. The vertical bars represent the Kolmogorov-Smirnov and the Kuiper metric, whereby the vertical bar on the left only refers to the Kuiper metric.	57
2.7	CUSUNORO plot for the target variable water content in 30 cm (<i>wc30</i>) and the parameters <i>al1</i> , <i>al2</i> , <i>kcmid</i> and <i>x</i> . Each curve represents one parameter. CUSUNORO curves in the dark gray area and above the red horizontal line have a ω value < 0 . CUSUNORO curves in the light gray area and below the red horizontal line have a ω value > 0 . The red, vertical, dashed line represents the linearity factor 1_n . <i>kcmid</i> has its largest impact at $1_n=0.55$	62
2.8	Schematic overview of the data processing including (field) data pre-processing, data generation and data processing.	64
2.9	Data post-processing: filtering the sensitivity indices by different criteria in a five-stage approach to reduce the numerical noise. Pdf-based importance measures reduced delta ($\hat{\delta}$) and point estimate of delta (δ^{pt}) are given in red, green represents the cdf-based importance measures sensitivity measure relying on the Kuiper metric (β^{kui}) and sensitivity measure relying on the Kolmogorov-Smirnov metric (β^{ks}), and blue stands for the variance based indices from given data and the Pick and Freeze design.	66
3.1	Minimum, maximum, median, mean and prediction intervals of the simulation results for each evaluated date on the x-axis faceted by the individual target variables.	75

3.2	Time series of the model output distributions' skewness and kurtosis over the simulation period for the target variables.	77
3.3	Convergence criteria cvg_{idx} for the values of the different sensitivity indices (colors) and the 15 target variables evaluated at different sample sizes (divided by 10000). The box-plots represent the variance of cvg_{idx} within the time series. The dotted line marks the threshold of 0.5 for convergence (see eq. 2.27).	80
3.4	The 95-quantile of the convergence criteria for the parameter ranks cvg_{rk} plotted as a function of the sample size. Each facet represents one sensitivity index. Error bars indicate standard deviations of cvg_{rk} for the different target variables and dates.	82
3.5	Overview of the needed computing resources as a function of sample size for the moment-independent methods (cdf, δ , MI), the variance-based methods (sobol indices, matrix, VB) and constructing the CUSUNORO curve. File size of model output, total analysis time in days, time to calculate on sensitivity index and the sampling efficiency are markers for computing resources and efficiency and are given on the square root scaled y-axis in their units.	83
3.6	The rank correlations for all sensitivity indice pairs are shown separately for each traget variable and at highest method specific sample size (1000000, 200000).	85
3.7	Stacked $S1^{GD}$ values for the target variables of the crop group plotted against the dates of the cultivation period and the value for $S1^{GD}$ (idxvalue). Each color, marker and line type combination represents one parameter, and colors are target group specific (crop: green, water: blue, nitrogen: red, flux: purple, ini: gray). Area and marker size reflect the importance of each individual parameter.	89

3.8	Stacked $S1^{GD}$ values for target variables of the water group. For labels, see figure 3.7.	91
3.9	Stacked $S1^{GD}$ values for the target variables of the nitrogen group. For labels, see figure 3.7.	92
3.10	Stacked $S1^{sat}$ values for target variables of the flux group. For labels, see figure 3.7.	94
3.11	Parameter ranks of the weighted and averaged sensitivity index values for the target variables of the crop group and for the four different sensitivity indices. Markers and colors represent the parameters.	96
3.12	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable leaf are index (LAI).	98
3.13	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable vegetative biomass ($VegBm$). . .	100
3.14	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable generative biomass ($GenBm$). .	101
3.15	Time series of the sensitivity (idxvalue) for the sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable nitrogen content in the generative biomass ($N - gBm$). .	102
3.16	Time series of the sensitivity (idxvalue) for the sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable development stage (dev).	103
3.17	Parameter ranks of the weighted and averaged sensitivity index values for the target variables of the water group for the four different sensitivity indices. Markers and colors represent the parameters.	104
3.18	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable soil water content in the top soil ($wc30$).	107
3.19	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable soil water content in the sub soil $wc90$	108

3.20	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable matric potential in 15 cm depth (<i>mp15</i>).	109
3.21	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable matric potential in 75 cm depth (<i>mp75</i>).	110
3.22	Parameter ranks of the weighted and averaged sensitivity index values for the target variables of the nitrogen group for the four different sensitivity indices. Markers and colors represent the parameters.	112
3.23	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable nitrate content in the top soil (<i>NO₃30</i>).	114
3.24	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable nitrate content in the sub-soil (<i>NO₃90</i>).	115
3.25	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable ammonium content in the top soil (<i>NH₄30</i>).	116
3.26	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable ammonium content in the sub soil (<i>NH₄90</i>).	117
3.27	Ranking of the target variables of the flux group for the four different sensitivity indices.	119
3.28	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable daily actual Evapotranspiration (<i>aET</i>).	120
3.29	Time series of sensitivity indices first-effect index from the Pick and Freeze design ($S1^{PF}$), total-effect index (ST), δ^{pt} and β^{ks} for the ratio of daily actual transpiration to daily actual evaporation (<i>fTrs</i>).	121

3.30	Values extracted from the CUSUNORO curve for the vegetative biomass (VegBm) and the generative biomass (genBm) for characteristic parameters.	124
3.31	Values extracted from the CUSUNORO curve for the ammonium and water content in the top soil (NH_430 , $wc30$), the matric potential in 15 cm ($mp15$) and the actual evapotranspiration (aET) for the parameter $PHINT$	125
7.1	Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the ratio of daily actual Evapotranspiration to daily potential Evapotranspiration ($faPET$).	180
7.2	Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the plant group for all parameters that have a first-order effect.	181
7.3	Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the water group for all parameters that have a first-order effect.	182
7.4	Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for the plant group's parameter that have a first-order effect.	183
7.5	Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for for the water group's parameter that have a first-order effect.	184
7.6	Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for the nitrogen group's parameter that have a first-order effect.	185
7.7	Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for the ini group's parameter that have a first-order effect.	186

7.8 Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the flux group for all parameters that have a first-order effect. . 187

List of Tables

2.1	Basic soil profile information for the simulation site EC1	29
2.2	Nitrogen (N) and carbon (C) input during the simulation period. Values marked by * are standard values of ExpertN.	29
2.3	Target variables, their abbreviations and their units for the four groups.	32
2.4	Varied parameters ordered by the five groups crop, water, flux, nitrogen and ini with their variation ranges and units.	34
2.5	Varied parameters ordered by the five groups crop, water, flux, nitrogen and ini with their variation ranges and units.	36
2.6	External parameters, their abbreviations (abbr.) and their set value for this work.	70
3.1	Median, maximum (max) and minimum (min) values of kurtosis of the model output distribution over the simulation period for the individual target variables.	79
3.2	Linearity parameter 1_n , its standard deviation, the corresponding parameter value (x_{1n}) and its standard deviation for selected model parameters within the target groups.	123
3.3	Parameter ranks for the weighted and averaged sensitivity indices $S1^{PF}$, ST, $\hat{\delta}$ and β^{ks} . The average is taken over time and target variables. Parameter count gives the number of target variables that the parameter influences.	127

3.4 Comprehensive overview of the results of the global sensitivity analysis for the different target variables. A "+" simply means "has an impact," and "-" indicates "has no impact". An x in parenthesis means that the specific characteristic is true for only one sensitivity index (SI), a "*" means that this is only true during the vegetation period. If there is no "~" in the last column, the ratio of first-effect index (S1) to ST is only poor during the time specified in the time of interaction. 129

List of Abbreviations

δ	- delta
$\hat{\delta}$	- bias reduced delta
δ^{pt}	- point estimate of delta
C	- carbon
cdf	- cumulative density function
cv	- critical value
DPC	- data process chain
ET	- evapotranspiration
GD	- Given Data
GSA	- Global sensitivity analysis
idxvalue-	value of the sensitivity index
LHS	- Latin Hypercube Sample
kde	- kernel density estimator
β^{ks}	- sensitivity measure relying on the Kolmogorov-Smirnov metric
β^{kui}	- sensitivity measure relying on the Kuiper metric
M	- number of partitions
MI	- moment-independent
MvG	- Mualem van Genuchten
N	- nitrogen
NH_4^+	- ammonium
NO_3^-	- nitrate
P&F	- Pick and Freeze design

pdf - probability density function

S1 - first-effect index

$S1^{PF}$ - first-effect index from the Pick and Freeze design

$S1^{GD}$ - first-effect index from Given Data

SI - sensitivity index

SI^{PF} - sensitivity indices from the Pick and Freeze design

SI^{MI} - moment-independent sensitivity indices

SI^{GD} - sensitivity indices from Given Data

SI^{VB} - Variance-based sensitivity indices

SNWC- Soil N and water condition

SR - simulation results

ST - total-effect index

TGV - target variable

TGV_{frc}^{fx} - target variables of the flux group that form a fraction

TGV_{gen}^{pl} - target variables referring to the generative crop organs

TGV_{veg}^{pl} - target variables referring to the vegetative crop organs

TGV_{30}^* - target variables referring to the top soil

TGV_{90}^* - target variables referring to the sub soil

UCSR - uncertainty of the simulation results

VSR - variance of the simulation results

VB - Variance-based

XN - Expert-N

I. Introduction

1.1 Motivation

The changing global climate, agricultural structural change, environmental pollution and a growing world population are some of the major challenges of today's agriculture. Climate change entails higher temperatures and changing rain patterns, which are environmental factors that control crop growth. The structural changes in agriculture have led to the situation that more food has to be produced by fewer farmers in the same area. Hence, each farmer has to manage more acreage while simultaneously increasing its productivity. At the same time, the societal call for more sustainable and environmental-friendly agricultural production has increased. Altogether, today's farmer has to be an efficient manager who has to successfully balance the demand for more food with fewer inputs under changing, uncertain production conditions. Therefore, it is of crucial interest to assess how the changing crop growing conditions will alter agricultural production. Furthermore, farmers need assistance in managing their increasing acreage ecologically and profitably.

Soil-crop models have been widely used to address these challenges. They represent the complex, coupled processes in an agroecosystem in a deterministic manner and have been designed to coherently simulate crop growth, water and nitrogen dynamics in a given environment. Each "dynamic" is represented by its own model, i.e., a soil crop model

is a composition of different sub-models. The typical model outputs are state variables such as soil water content and yield, as well as fluxes such as evapotranspiration (ET), groundwater recharge and nitrogen (N) leaching. While they have traditionally been used to forecast yields, they are nowadays often used to predict the impact of environmental changes on agro-ecosystems. In the context of food security for a growing world population, the Agricultural Model Intercomparison Project (AgMIP) set the goal of predicting how global agricultural production will be affected by climate change (AgMIP (2017); Rosenzweig et al. (2013)). Among others, Asseng et al. (2013) used a set of 30 crop models to predict how the grain yield of wheat, one of the world population's main food sources, will change with rising temperatures in the future (Martre et al. (2015); Asseng et al. (2014)).

In practice, soil-crop models are often used to assess the impacts of agricultural management, e.g., fertilization, irrigation, crop rotation on nitrate leaching, groundwater recharge and quality, or greenhouse gas emissions. This is especially important in regions where water is scarce or where the groundwater is polluted by nitrate or pesticides. In the long run, soil-crop models will be used as a decision support tool to mitigate management practices in simulation studies that decrease the emission of greenhouse gases and groundwater pollution from agricultural production sites (Brilli et al. (2017); Dumont et al. (2015); Wang et al. (2016); Hu et al. (2010); Nolan et al. (2010)).

Lately, soil-crop models have enjoyed increasing popularity in the context of precision farming. They are used to provide data-driven recommendations for site-specific fertilization and to forecast yield (Next-Farming, 2019). Few scientific studies have tested the ability of soil-crop models to reproduce site-specific yields based on field scale varying soil property information, such as humus content and soil texture (Plauborg et al. (2015); Wallor et al. (2018)). In both studies, the soil-crop models' yield prediction is sensitive to varying, site-specific soil properties, but the ability to reproduce measurements differs.

Soil-crop models, and hence all predictions in the above studies, suffer

from uncertainty in the simulation results (SR) due to model structure, model parameters, model input and measurement data (Wallach and Thorburn, 2017). The model structure uncertainty, which is caused by different mathematical descriptions among different soil-crop models can be assessed via the use of a multi-model ensemble, as in Asseng et al. (2013) or Martre et al. (2015). The uncertainty arising from uncertain model inputs, i.e., parameters, initial conditions and forcing, e.g., fertilization and weather data, can be explored and quantified by sensitivity analysis. A sensitivity analysis maps the dependence between the model input and the model output and apportions the model output uncertainty among the sources of uncertainty in the model input (Saltelli et al., 2004) and determines how reliable the derived decisions are (Tarantola et al., 2002). The outcome of a sensitivity analysis is an SI, which quantifies the strength of dependence.

Since soil-crop models often depend on 30 or more parameters, the selection of key drivers of uncertainty and the identification of parameters that can be fixed without any effect on the model output becomes especially important in the context of parameter calibration because it is often infeasible to include all model inputs. Furthermore, sensitivity analysis can provide information about the soil-crop model's structure in regard to equifinality (Beven and Freer (2001); Borgonovo et al. (2017)). Equifinality means that the same state of a model or system can be reached in multiple ways, i.e., many parameter combinations result in the same model output. This leads to an overall poor determinability of the model inputs because they interact with each other. A sensitivity analysis can be either local or global. Local sensitivity methods are mostly derivative-based, and they vary only one parameter at a time (OAT), assuming a linear relation between the model input and model output. Global sensitivity analysis (GSA) has the advantage of being able to consider simultaneous parameter variation, and it does not require any assumptions about the input-output relation. Thus, it is much more suited for the non-linear soil-crop models. In addition, GSA methods can be distinguished by the representation of the model

output uncertainty in the analysis. Variance-based (VB) GSA methods consider the model output variance, whereas moment-independent (MI) methods take into account the whole model output distribution. The latter is of major interest if the model output distribution is distinctive from a normal distribution, e.g., bimodal or skewed.

Although guidelines published by the US EPA (2009) and the European Commission (2009) require conducting a sensitivity analysis before model predictions can be used for decision support, sensitivity analysis has rarely been applied to soil-crop models. A detailed description of the individual studies that conducted a GSA for soil-crop models is given in DeJonge et al. (2012). The studies focus either on yield or above-ground biomass at maturity as the target variable (TGV) of interest and have also been conducted in the context of calibration, i.e., the identification of influential and non-influential parameters. They consider model input variability regarding different climates and soils (Confalonieri et al. (2010b); Richter et al. (2010); Specka et al. (2015); Vanuytrecht et al. (2014); Sexton et al. (2017)), different irrigation management (XING et al. (2017); DeJonge et al. (2012); Liang et al. (2017)) or both Zhao et al. (2014). The crops considered range from maize, rice and wheat to cotton and peanuts. However, the sensitivity analyses used are restricted to the Morris screening (Morris, 1991) and the VB sensitivity methods. Furthermore, Liang et al. (2017), Ruget et al. (2002) and Vanuytrecht et al. (2014) considered, in some way, soil hydraulic and ET parameters in their sensitivity analyses while Ruget et al. (2002) used the response surface method. The other studies, as well as Makowski et al. (2006) and Tan et al. (2017), account for crop-specific parameters only. Only Liang et al. (2017) have considered additional parameters to control the N-transformation in the soil and also considered N-leaching as TGV. Analyzing parameter sensitivities at different points in the vegetation period suggests that they can differ significantly over time (Wang et al. (2013); Tan et al. (2016); Lamboni et al. (2011)).

To our knowledge, no study has additionally considered MI methods

and accounted for crop parameters, soil hydraulic parameters, mineralization and nitrification rates, temporally resolved parameter sensitivities **and** multivariate model outputs at once. In this dissertation, we use GSA to assess the time course of parameter sensitivities for different TGVs that represent the crop, water, nitrogen and flux sub-models of the XN-CERES soil-crop model. Thus, the parameters of all modules are considered, and the cross-module impact is assessed for a deep loess soil profile in Southwest Germany. Furthermore, VB and MI methods are compared, and their suitability for the XN-CERES is evaluated. The results will improve the understanding of the model's behavior and the inherent uncertainty uncertainty regarding model inputs. They will provide a guideline on how to calibrate and use soil-crop models to comprehensively predict the dynamic evolution of an agro-ecosystem. The detailed objectives and research questions are outlined in section 1.2.

1.2 Objectives

This thesis has two major objectives. The first is to increase the understanding of a soil crop model's behavior and structure by applying a GSA (GSA application). The second objective to identify the best-suited GSA method to achieve the first objective. Therefore, different GSA methods are compared (GSA comparison). The following enumeration gives a brief overview of the individual objectives.

1. GSA - application

- 1.1. Factor Prioritization: Identify the key drivers of uncertainty for different TGVs of the XN-CERES.
- 1.2. Factor Fixing: Identify factors that can be fixed at any value with insignificant impact on the TGVs.
- 1.3. Trend Identification: Quantify the direction of the model inputs' impact on the TGVs and qualify their linearity.

- 1.4. Model Interrelations: Discern the presence and assess the magnitude of factors that have a sub-model comprehensive impact.
 - 1.5. Equifinality: Assess the determinability of the model inputs for the overall soil-crop model, the sub-modules and each output variable.
2. GSA - comparison
 - 2.1. Estimation: properties of the individual sensitivity indices in regard to convergence, computer resources, post-processing and stability.
 - 2.2. Identification: Elaborate which GSA methods are suited best for Factor Fixing and Factor Prioritization for the XN-CERESL.

As mentioned in the Introduction, soil-crop models are often used to assess exogenous impacts on the different state-variables of an agro-ecosystem. In most of the outlined application cases, crop models have either been used without calibration or have been calibrated by hand against measured data. In many cases, the calibration process has not been described at all. Hence, often, readers have no knowledge of the uncertainties and whether they were handled correctly.

With the results from the GSA application, we would like to answer four research questions. The **first** is as follows: Can the individual sub-models be calibrated in isolation, or do all parameters have to be considered simultaneously in the calibration process to predict either only one TGV e.g., yield, or to predict two or more TGVs at once? Hence, do the parameters primarily associated with the water model have an impact on the crop growth and the soil N-content? By identifying the key drivers of uncertainty and non-influential parameters for different TGVs of the four models (crop, fluxes, water and N dynamics) this question can be answered directly.

Furthermore, the identification of influential and non-influential param-

eters entails the assessment of the determinability of the parameters, and these is evaluated against the background of the **second** research question: Is it in theory possible to find a cross-module, univocal parameter set? However, a GSA does not allow for conclusions about whether there is a parameter combination that can adequately simulate two TGVs at once, even if the inherent model structure allows for it.

Temporally resolved parameter sensitivities provide information to answer the **third** research question: Are parameters sensitive for the whole simulation period, or can they be restricted to a certain time window? At what time must the modeler have information about the real world state if he or she is interested in specific TGVs? Trend identification allows the assessment of whether two or more sensitive parameters have a contrasting effect on a TGV at the same time. This indicates a competing situation of parameters in the calibration process. For example, two parameters may impact the *LAI* in April. However, one parameter accelerates leaf growth, whereas the other decelerates it. Hence, an increase in parameter one can be compensated by changing parameter two.

The second objective, "GSA comparison" provides an answer to the **fourth** research question: Is variance an appropriate measure to represent the uncertainty of the simulation results (UCSR), or are MI methods more suited? Furthermore, we compare the two methods in regard to the estimation procedure. Are the methods numerically stable? How fast is the convergence of the SIs? What are the computational costs? These objectives and their subordinate targets, placed within the context of the outlined research question, allow for general conclusions about the uncertainty of XN-CERES and its resulting predictive power for simulations of the whole soil-plant system.

II. Material and Methods

2.1 Definitions

In this section, we define the expressions that are frequently used in this dissertation.

Vegetative growth refers to the growth of the crop's vegetative organs, i.e., leaves, stem and roots. The period in which the crop partitions all assimilate into the vegetative organs is referred to as the **vegetative phase**. Enhancement, formation and growth thereby include the increase in size as well as the increase in weight. **Generative growth** refers to the grains and includes the yield quality. Here, yield quality is defined as the N content in the grains. The **generative phase** defines the period in which the assimilates are primarily invested into the grains and their quality. In the XN-CERES, the generative phase starts with BBCH 70, and the vegetative phase includes the period of flowering.

Top soil refers to the first 30 cm of the soil, and **sub soil** refers to the soil depth from 30 to 90 cm. The two soil depth distinctions include a different number of soil layers. Soil parameters include parameters for soil hydraulic and N-transformation. Parameters for the top soil are suffixed with 1, and for the sub soil with 2. If they do not have a number, they refer to both soil depths.

The **vegetation period** refers to the period from February to harvest, i.e., when most of the crop development and growth occurs. The

cultivation period includes the period from sowing to harvest, and the **simulation period** represents the time from simulation start to simulation end.

The terms **sensitivity index**, **uncertainty importance measure** and **sensitivity measure** are used interchangeably.

We classified four **target groups**. Each group is assigned to a number of target variables (TGV), i.e., state-variables that the XN-CERES outputs and assigned to a number of parameters that are varied for the GSA. For example, crop group parameters include all parameters that are assigned to the crop sub-model.

The **XN-CERES** consists of four sub-models for the crop, the water regime, the N dynamic and the fluxes, i.e., the upper boundary flux. The sub-models are all part of the XN-CERES model.

Soil N and water condition (SNWC) is a general term that refers to the amount of crop available N and water in the soil.

Factor and **parameter** are synonyms for a **model input** variable and refer to the inputs of a model that can be varied and are uncertain.

The **model additivity** is defined as the sum of the S1 for each TGV. It is scaled between 1 and 0, whereby 1 means that the model is perfectly additive. According to Borgonovo et al. (2017), the difference between 1 and the model additivity "can be considered as an indicator of the percentage of the model output variation apportioned by interactions." If a parameter is **fixed**, it is excluded from calibration. It has either been calibrated before or set in advance.

2.2 Site Description

In this study, we considered a study site located in the Kraichgau region (48.9°N 8.7°E, 319 m a.s.l.), in Baden-Wuerttemberg, Germany. The mean temperature in the Kraichgau is 9.3°C and annual precipitation is 777 mm. The study site comprises three field trails (stations) that have been maintained and supervised by DFG Research Unit 1695 since 2009. All fields are agriculturally managed and operated by local

farmers (best practice). We considered the field called EC1. The soil developed from loess and is deeply weathered and fertile (*depth* >165 cm) with a high storage capacity of crop-available water. At the EC1

Table 2.1: Basic soil profile information for the simulation site EC1

horizon	lower depth (cm)	texture	organic content	label
A	30	Ut4	1.72	1
B	60	Ut4	0.9	2
C	90	Ut4	0.2	2

station, weather data including precipitation, temperature, humidity and global radiation were measured at a 30-minute frequency. In this study, we only considered the vegetation period of 2011, in which winter wheat was sown on October 19, 2010 and harvested on July 28, 2011. The cultivation period was thus 282 days. Soil profile and management information is given in table 2.1 and table 2.2. For further descriptions of the study sites, the field trails, the field management and the recorded weather data, see Ingwersen et al. (2011) and Wizemann et al. (2015).

Table 2.2: Nitrogen (N) and carbon (C) input during the simulation period. Values marked by * are standard values of ExpertN.

management	date	kg-N	kg-C	kg NH ₄ -N	kg NO ₃ -N	kg- AmidN
fertilization						
NPK	03.11.2010	46	0	23.4	23.4	0
urea	01.04.2011	69	0	0	0	69
calcium- ammonium- nitrate	17.05.2011	54	0	27	27	0
preceding crop						
residues	14.10.2010	10.9	556	-	-	-
root	14.10.2010	30 ¹	1200 ¹	-	-	-

¹ Default values set by the ExpertN software

2.3 Model Setup

Model Configuration A summary of the model setup and configuration is given in figure 2.1. All simulations were set up and run with the agro-ecosystem modeling software Expert-N (XN) 3.1, which offers the possibility of combining a set of different, tested sub-models for water and N dynamics as well as for crop growth and ET. In this work, the CERES crop model (Ritchie and Godwin, 1987), the SOILN model (Johnsson et al., 1987) for N and carbon (C) turnover, the Hydrus1D model (Simunek et al., 1998) for the soil water regime and the FAO-Penman-Monteith (Allen et al., 1998) method to estimate potential crop-specific ET are chosen and combined. We refer to this combination as XN-CERES.

The CERES model simulates crop development based on the concept of thermal time. It simulates vegetative growth, including stem, root and leaf formation and generative growth including yield and its quality. The partitioning of assimilates between the different crop organs is controlled by the phenological stage. The potential biomass growth rate is based on light extraction from radiation (constantly 50%), the light use efficiency, temperature and leaf area index (LAI). The potential biomass growth rate can be reduced by water, N and temperature stress. XN-CERES stages 1-4 only consider vegetative growth and correspond to the BBCH stages 10 to 70. Leaf formation is only considered in stages 1 and 2, and the flower is not represented as a generative organ but assigned to the stem. Generative growth corresponds to the grain-filling phase and maturation and refers to the generative biomass and its quality.

The SOILN model considers three soil organic matter pools: humus, litter and manure. They differ in the assumed stability of their organic matter and vary from stable to easily decomposable. Mineralization is controlled by pool-specific rates, as are nitrification and denitrification. N transport is calculated with the convection dispersion equation. The Hydrus1D model uses the Richards equation (Richards, 1931) to simu-

late 1D water flow. We use the van Genuchten-Mualem parametrization of the hydraulic functions (van Genuchten, 1980). The FAO method calculates potential ET by multiplying a grass reference ET, which accounts for climatic conditions, with crop-specific factors to account for respective crop features. All formulas and XN-specific implementations are documented in Priesack (2006).

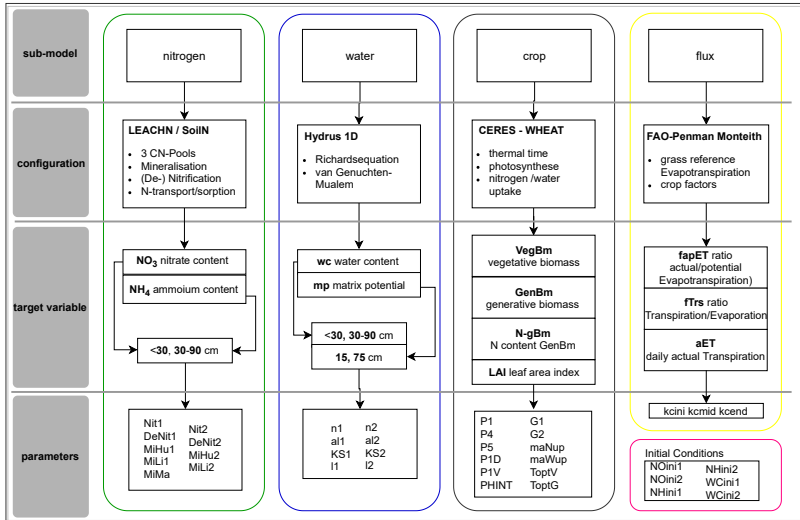


Figure 2.1: Summary of the model configuration and setup for the different submodules: nitrogen, water, crop, and flux with the group-specific target variables and parameters.

Model Initialization We used an atmospheric forcing, i.e., measured weather data (cf. 2.2), and set the lower boundary to free drainage. Initial C and N content for the three soil organic matter pools (humus, litter and manure) were approximated from measurements. As a rule of thumb, to initialize the humus pool’s C and N content, the measured humus content (cf. fig. 2.1) is multiplied by 0.58 and 0.058, respectively. To initialize the litter pools’ C and N content, the known

Table 2.3: Target variables, their abbreviations and their units for the four groups.

Group	Name	Abbreviation	Unit
crop	Development stage	dev	BBCH
	Leaf area index	LAI	m^2m^{-2}
	Vegetative biomass	VegBM	$t ha^{-1}$
	Generative biomass	genBM	$t ha^{-1}$
	Vegetative biomass	N-gBM	%
Water	Water content 0-30 cm	wc30	mm
	Water content 0-90 cm	wc90	mm
	Matrix potential in 15 cm	mp15	$m watercolumn$
	Matric potential in 75 cm	mp75	$m watercolumn$
Nitrogen	Nitrate content 0-30 cm	NO_330	$kg - N ha^{-1}$
	Nitrate content 30-90 cm	NO_390	$kg - N ha^{-1}$
	Ammonium content 0-30 cm	NH_430	$kg - N ha^{-1}$
	Ammonium content 30-90 cm	NN_490	$kg - N ha^{-1}$
Flux	Daily actual evapotranspiration	aET	mm
	Ratio of transpiration to evapotranspiration ¹	fTrs	mm
	Ratio of actual evapotranspiration to potential evapotranspiration ¹	fapET	mm

¹ If not explicitly specified, evaporation, transpiration and evapotranspiration always refer to actual daily values.

mass of the preceding crop residues is partitioned according to its measured C/N ratio (cf. table 2.2). Due to this initialization procedure of the soil organic matter pools, the simulation started October 14, 2010, which was the harvest date of the preceding crop. In order to maximize the simulation period length, the end of the simulation was set to the sowing date of the proceeding crop which, was the August 22, 2011. In total, this produced 312 days of generated model output. Soil initial temperature was set to measured values and was $8.38^\circ C$ in the top soil and $9.67^\circ C$ in the sub soil. Soil management and fertilization were entered into the model as reported by the operating farmers (cf. table 2.2). Discretization of the soil profiles was 5 cm, and the maximum rooting depth of winter wheat was set to 160 cm in accordance with Kutschera and Lichtenegger (1960) and field observations.

Since we are interested in parameter cross-model impact, we look at

a variety of TGVs. One TGV is one state variable, e.g., water content, of the agro-ecosystem that the model outputs. We thus focus on variables that are often measured in field trials or are of interest as inputs in other models. The TGVs are listed in table 2.3 with the assigned group, their abbreviations and unit. We defined four target groups: crop, water, nitrogen and flux, which coincide with the XN sub-models.

Model Parameters In total, we considered 39 parameters in the GSA, including initial conditions for the water and nitrogen content in the soil. An overview of the parameters, their variation range, their assigned group and their units is given in table 2.5. In addition to the groups for the different sub-models introduced above, we defined a fifth group, *ini*, which includes the initial conditions. A brief description of the individual parameters of the five groups is given in table 2.4. For the crop model CERES, we selected 12 parameters: $P1$, $P4$, $P5$, $P1D$, $P1V$, $G1$, $G2$, $mxWup$, $mxNup$, $ToptV$, $ToptG$ and $PHINT$. $P1$, $P4$ and $P5$ [$^{\circ}C\ d$] are thermal temperatures needed to finish the CERES development stages 1, 4 and 5. $P1D$ and $P1V$ are genetic parameters for the photo- and vernalisation sensitivity of winter wheat. They influence the duration of the first stadium as well as the blossom development. Simulated yield is directly affected by $G1$, which determines the number of grains per stem [g^{-1}], and $G2$ represents the maximum grain growth rate [$mg\ d^{-1}$].

For the simulation of the water regime, we used the van Genuchten-Mualem parameterization of the hydraulic functions. To keep the problem simple but still allow for differences in the hydraulic properties of the soil horizons (cf. tbl. 2.1), we divided the soil profile into two horizons.

Table 2.4: Varied parameters ordered by the five groups crop, water, flux, nitrogen and ini with their variation ranges and units.

Group	Name	Description
crop	P1D	genetic parameter for winter wheat's photo-sensitivity
	P1V	genetic parameter for winter wheat's vernalisation sensitivity
	PHINT	phyllochron
	G1	number of grains per stem
	G2	maximum grain growth rate
	P1	thermal temperature from emergence to BBCH 19,
	P4	thermal temperature from BBCH 40 to 69
	P5	thermal temperature from BBCH 70 to 79
	mxWup	daily maximum crop water uptake per cm root length
	mxNup	maximum crop nitrogen uptake per day and hectare
Water	ToptV	optimal temperature during vegetative phase
	ToptG	optimal temperature during generative phase
	n1	representation of the slope of the retention curve
	n2	
	al1	inverse of the air entry point
al2		
KS1	saturated hydraulic conductivity	
Flux	KS2	FAO crop factor for early development stages
	l1	
	l2	tortuosity of the soil pores
	kcini	FAO crop factor for mid development stages
Nitrogen	kcmid	FAO crop factor for end development stages
	kcend	
	Nit1	nitrification rate for the top soil
	DeNit1	denitrification rate for the top soil
	MiHu1	humus mineralization rate for the top soil
	MiLi1	litter mineralization rate for the top soil
	MiMa	manure mineralization rate for the top soil
	Nit2	nitrification rate for the sub soil
	DeNit2	denitrification rate for the sub soil
	MiLi2	humus mineralization rate for the sub soil
MiHu2	humus mineralization rate for the sub soil	
Ini	NHini1	initial soil ammonium (NH_4^+) content in the top soil
	NOini1	initial soil nitrate (NO_3^-) content in the top soil
	NOini2	initial soil NO_3^- content in the sub soil
	NHini2	initial soil NH_4^+ content in the sub soil
	WCini1	initial volumetric soil water content in the top soil
	WCini2	initial volumetric soil water content in the top soil
Dummy	x	no impact on the models

Each horizon was assigned a set of van Genuchten parameters ($n1$, $n2$, $\alpha1$, $\alpha2$, $KS1$, $KS2$, $l1$ and $l2$). Alpha (α) [cm^{-1}] and n [-] can be regarded roughly as shape parameters of the retention curve. Saturated conductivity (KS) [cmd^{-1}] and l [-], which represent the tortuosity of the soil pores, scale the conductivity curve.

In agreement with the van Genuchten parameters, we defined a set of mineralization, nitrification (Nit^*) and denitrification ($Denit^*$) rates for the two horizons ($*1/2$). Mineralization rates were defined for the humus, litter, and manure pool ($MiHu^*$, $MiLi^*$, $MiMa$). Additionally, we considered the crop factors of the FAO approach as uncertain and included $kcini$, $kcmid$ and $kcend$ in the GSA. As mentioned above, we accounted for the initial soil NH_4^+ -, NO_3^- - and water content (NH_4ini^* , NO_3ini^* , $WCini^*$). They were set separately for the two horizons. The parameter ranges were derived from expert knowledge, former studies and literature values. In addition, we set the range of the van Genuchten parameters based on pedotransfer functions and for numerical stability reasons. In the end, we added a dummy parameter x , which has no impact on any model output, to obtain an estimate for numerical noise (cf. 2.4.4.1).

2.4 Global Sensitivity Analysis

At first, the classification and subdivision of GSA is given in section 2.4.1. Section 2.4.2 introduces a general notation. In sections 2.4.3 and 2.4.4, respectively, the mathematical framework, the basic principle, the properties and estimation of the individual sensitivity measures for VB and MI methods are explained in detail. The construction of the curve, a visual GSA method, and the extracted values are described in section 2.4.6. The data process chain (DPC) is explained in section 2.5. Finally, in sections 2.5.3 and 2.5.4, the analysis settings and the frameworks used are summarized.

Table 2.5: Varied parameters ordered by the five groups crop, water, flux, nitrogen and ini with their variation ranges and units.

Group	Name	Lower Boundary	Upper Boundary	Unit
crop	P1D	0.001	0.008	$^{\circ}C^{-1} d^{-1}$
	P1V	25	60	$^{\circ}C^{-1} d^{-1}$
	PHINT	70	150	$^{\circ}C d^{-1}$
	G1	20	4	g^{-1}
	G2	1	4	$g \text{ grain}^{-1} d^{-1}$
	P1	170	400	$^{\circ}C d$
	P4	120	200	$^{\circ}C d$
	P5	400	700	$^{\circ}C d$
	mxWup	0.01	0.1	$cm^3 (cm \text{ root})^{-1} d^{-1}$
	mxNup	0.003	0.027	$kg \text{ ha}^{-1} d^{-1}$
	ToptV	17	29	$^{\circ}C$
ToptG	17	29	$^{\circ}C$	
Water	n1	1.2	1.8	-
	n2	1.2	1.8	-
	al1	0.002	0.03	cm^{-1}
	al2	0.002	0.03	cm^{-1}
	KS1	150	260	$cm d^{-1}$
	KS2	60	150	$cm d^{-1}$
	l1	-1	8	-
l2	-1	8	-s	
Flux	kcini	0.2	0.8	-
	kcmid	0.5	1.25	-
	kcend	0.2	1	-
Nitrogen	Nit1	0.1	1	$kg \text{ ha}^{-1} d^{-1}$
	DeNit1	0.1	1	$kg \text{ ha}^{-1} d^{-1}$
	MiHu1	0.00001	0.0001	d^{-1}
	MiLi1	0.01	0.1	d^{-1}
	MiMa	0.01	0.1	d^{-1}
	Nit2	0.05	0.6	$kg \text{ ha}^{-1} d^{-1}$
	DeNit2	0	0.01	$kg \text{ ha}^{-1} d^{-1}$
	MiLi2	0.01	0.1	d^{-1}
	MiHu2	0.000001	0.00001	d^{-1}
Ini	NHini1	0	2.5	$kg \text{ ha}^{-1}$
	NOini1	0	120	$kg \text{ ha}^{-1}$
	NOini2	0	50	$kg \text{ ha}^{-1}$
	NHini2	0	0.5	$kg \text{ ha}^{-1}$
	WCini1	10	40	$Vol - \%$
WCini2	15	45	$Vol - \%$	
Dummy	x	1	10	-

2.4.1 Classification and Subdivision

A GSA quantifies the input-output dependency over the whole parameter space. It informs the modeler how the uncertainty in the model output can be apportioned among the uncertain model inputs (Saltelli et al., 2004). The result of a GSA is an SI for each model input, which reflects its importance for the total model output uncertainty. Therefore, an SI can also be regarded as an uncertainty importance or sensitivity measure. Saltelli (2002b) has defined three key properties a "good" sensitivity method should fulfill. The fourth property has been added by Borgonovo (2007).

1. The method must be capable of considering every distribution assigned to the model input.
2. The method must be capable of considering simultaneous variation of the model inputs.
3. The method must be model-free.
4. The method must be moment-independent.

According to Borgonovo (2017), these properties can be regarded as prerequisites for a sensitivity method to be global. In particular, the second prerequisite differentiates global sensitivity methods from local sensitivity methods, in which not all model inputs are varied at the same time. For this reason, local methods are often referred to as one-at-a-time (OAT) designs. The third prerequisite states that the method must be independent of a predefined functional relation between the model input and model output (i.e., linear, additive, monotonic). The fourth is discussed in detail in a moment. In the literature, many GSA methods are available. They can be categorized into non-parametric techniques (Helton et al., 2006), screening methods (Morris, 1991), VB methods (Sobol, 1993) and MI approaches (Borgonovo, 2006). In this work, the nonparametric techniques and the screening methods are not

considered because the former do not fulfill prerequisite 3, and the latter do not fully fulfill prerequisite 2. One major distinctive feature of the remaining three categories is how the uncertainty in the model output is represented in the GSA. Whereas VB methods take the model output's variance as a measure of uncertainty, MI methods consider the entire model output distribution. Finally, we define a new category called visual, which frames the CUSUNORO method (Plischke, 2012). CUSUNORO is the curve of the cumulative sum of normalized and reordered model output. It is constructed for each model input. In addition to the representation of the **model output** considered, the categories under consideration - visual, variance based, and moment independent - are distinguished based on the **setting**, the **data input** and the resulting **uncertainty importance measure**. This is explained in the following sections.

Settings Saltelli et al. (2004) proposed the following settings in the context of GSA:

1. In the **Factor Prioritization** setting, one is interested in finding the model inputs that, if fixed, lead to the greatest reduction in model output uncertainty.
2. In the **Factor Fixing** setting, the aim is to identify the model inputs that can be fixed at any location within their range without having an influence on the model output uncertainty.
3. In the **Variance Cutting** setting, the output uncertainty is reduced to a given threshold.
4. In the **Factor Mapping** setting, values of the model input are determined that lead to model realizations in a specified range of the model output space.

In this study, the aim is to identify the model inputs that are either key drivers or negligible for the model output uncertainty. Therefore, the

GSA is conducted in the sense of the Factor Prioritization and Factor Fixing settings.

Data Input Two distinctions are made regarding the GSA data input: Given Data (GD) or specific sampling scheme. The GD approach (Plischke et al., 2013) is based on the principle that the model output generated by any sampling algorithm (i.e., Random Monte Carlo, Latin Hypercube, ...) can be used to calculate the desired SI. In contrast, a method that depends on a specific sampling design requires the model output to be generated by a predefined, method-specific sampling algorithm (Cukier et al., 1978; Saltelli, 2002a; Pianosi and Wagener, 2015). The advantage of the GD principle is the possibility of reusing the generated model output for different purposes (i.e., model calibration) and to calculate different SIs from one sample (Borgonovo et al., 2016).

Model Output As mentioned above the different GSA categories account differently for uncertainty in the model output. VB methods expect the model output variance to be a sufficient measure of the underlying uncertainty (Saltelli, 2002b), whereas the CUSUNORO method uses the cumulative sum of the normalized and reordered model output. However, MI methods take the entire model output distribution into account, which can either be cumulative density functions (cdfs) or probability density functions (pdfs). Independent of the chosen distribution function, they do not rely on a moment that describes the uncertainty in the model output. Therefore, they are better suited in cases where the distribution of model output is different from a normal distribution (i.e., multi-modal, skewed).

Uncertainty Importance Measure This paragraph gives a short overview and description of the SIs returned by the different GSA methods. A detailed explanation of the individual SIs and their properties can be found in section 2.4. VB methods, independent of the data input, provide the S1, which measures the **direct** impact of a model input

on the model output’s variance. Hence, how much the model output variance could be reduced if one could fix that model input to its true value. Furthermore, the sum of the S1 for all factors gives information about the additivity and, to some extent, the determinability of the model. If a specific sampling design is used in addition to the S1, the second-effect index and the **total-effect index** (ST) can be assessed. Higher-order effects or the ST give the contribution of a model input together with other model inputs on the model output’s variance, i.e., its interaction. In the literature, they are also referred to as Sobol indices or first-/second-order index and totals.

In statistics, many different metrics exist to measure the difference between two distributions. For the MI methods, we focused on three uncertainty importance measures: 1. **delta** (δ) (Borgonovo, 2007), 2. $\beta^{k_{ui}}$ and 3. β^{k_s} . The first considers the difference between the integrals of two pdfs. Numbers two and three measure the distance between two cdfs using the Kuiper metric (Kuiper, 1960) and the Kolmogorov-Smirnov metric, respectively.

Finally, we define two new measures that also offer valuable insights into the model structure instead of measuring sensitivity. From the CUSUNORO plots, we extract information about the **linearity** 1_n and the **direction** ω of the model input’s impact on the model output. Loosely speaking, ω assesses if the model output increases or decreases along with the factor. 1_n indicates whether the factors’ impact degree is unequivocal over its entire factor range or higher in a sub-range. This is also called trend-identification.

In this work, GSA methods were applied to a soil-crop model in two settings: Factor Prioritization and Factor Fixing. We tested and compared MI and VB methods with regard to their suitability for the two settings addressed. Furthermore, we compared the estimation of SIs from GD with the classical estimation of the Sobol Indices in terms of agreement, convergence and computational costs. Trend-identification methods were used to further specify ”what happens in the model.”

In figure 2.2, the GSA methods used in this work are summarized in accordance with the above described classification.

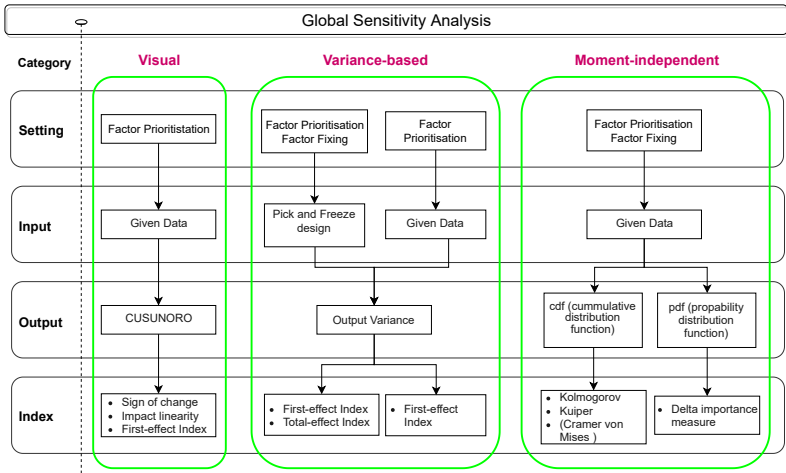


Figure 2.2: Schematic overview of the classification of global sensitivity methods

2.4.2 Notation

In a general formulation, the model output y is a multivariate function f of input vector X :

$$y = f(X). \quad (2.1)$$

The input vector is defined as $X = (x_1, x_2, x_k, \dots, x_z)$, where z is the total number of model inputs, and x_k refers to one parameter of the parameter vector. The index i is used to indicate that a model input x_k is conditioned on a value i within its parameter range, which is written as $x_k = x_{ki}$. N denotes the number of samples chosen by the modeler, and M denotes the number of partitions. Sub-ranges from the initial range of x_k are labeled as c_m^k , where $c_m^k \cap c_{m'}^k = \emptyset$ with $m = 1, 2 \dots M$ and $m \neq m'$. The pdf and the cdf are referred to as $f_Y(y)$ and as F_Y , respectively. Estimates of functions are labeled with a circumflex (*hat*).

2.4.3 Variance-based Methods

Brief History The application of VB global sensitivity methods can be traced to Cukier et al. (1978). They developed the so-called FAST method (Fourier Amplitude Sensitivity Test). Later, Hora and Iman (1986) introduced the uncertainty importance, which is defined as the expected reduction in the model output variance if a factor could be fixed at some point in its uncertainty range. Four years later, for the first time, Iman and Hora (1990) defined S1 with respect to the log-transformed model output. Only Sobol (1993) provided the basic concept and proof of the Monte Carlo based estimation of the VB sensitivity measures. It relies on the statistical framework of High Dimensional Model Representation (HDMR), also called functional ANOVA expansion, which has been firmly established by Efron and Stein (1981). The next paragraph provides only a summary of the mathematical intuition of the HDMR and variance decomposition. For more mathematical details on HDMR and the derivation, with proof, of the Sobol indices,

the reader is referred to the cited literature above or to Saltelli et al. (2008) and Borgonovo (2017).

Mathematical Framework The decomposition of $f(X)$ is an "expansion into effect functions of increasing dimension" (Sobol, 1993) if 1. $f(X)$ can be written as

$$f(X) = f_0 + \sum_{k=1}^z f_{x_k}(x_k) + \sum_{k<j} f_{x_k, x_j}(x_k, x_j) + \dots + f_{x_1, x_2, \dots, x_z}(x_1, x_2, \dots, x_z), \quad (2.2)$$

2. is integrable, and 3. the effect functions with regard to their own variables are zero (Sobol, 1993; Borgonovo, 2017). Then, the zero-degree term f_0 is the expectation of the model output $\mathbb{E}(y)$ and is assessed by integrating $f(X)$ with respect to all model inputs. The effect function of the first dimension, i.e., the first-order term, $f_{x_k}(x_k)$ is the mean effect of x_k on the model output, if it varies alone. It is assessed by integrating f_{x_k} with respect to all model inputs but x_k and subtracting the zero-degree term. To continue, the second-order term $f_{x_k, x_j}(x_k, x_j)$ with $x \neq j$ represents the joint, average effect of x_k and x_j on the model output. Hence, $\mathbb{E}(y|(x_k, x_j))$ is obtained by integrating $f(X)$ with respect to all model inputs but (x_k, x_j) and subtracting f_0 as well as the two first-order terms $f_{x_k}(x_k)$ and $f_{x_j}(x_j)$. This is repeated until the effect function of the highest dimension that can be assessed by taking the difference between the zero-degree term and the sum of all effect functions of dimension $< z$. Put simply, the difference between the spanned space of the model output with respect to all model inputs minus the spanned space of the model output with respect to all model inputs but the targeted one gives the contribution of that specific model input to the overall model output space. If $f(X)$ is measurable and square integrable, equation (2.2) can be rewritten as a decomposition of variance:

$$\mathbb{V}(y) = \sum_{k=1}^z Vx_k + \sum_{k<j} Vx_{k,j} + \dots + Vx_{1,2,\dots,z}. \quad (2.3)$$

The variance decomposition has 2^z terms, and the different terms can be understood as the contribution of the target model input(s) to the model output's variance. Hence, Vx_k is the contribution of the individual model input x_k to the total variance of the model output. This is already the interpretation of S1. Indeed, normalizing Vx_k by the total variance $\mathbb{V}(y)$ of the model output yields the first-order effect or the first-effect index $S1x_k$.

$$S1x_k = \frac{Vx_k}{\mathbb{V}(y)}. \quad (2.4)$$

In accordance with S1, the indices of higher dimensions can be written as

$$S2x_{k,j} = \frac{Vx_{k,j}}{\mathbb{V}(y)}, \quad SZx_{k,j,\dots,z} = \frac{Vx_{k,j,\dots,z}}{\mathbb{V}(y)} \quad (2.5)$$

and the total-effect index of the model input factor x_k is then the sum of the individual indices:

$$STx_k = S1x_k + S2x_{k,j} + \dots + SZx_{k,j,\dots,z} \quad (2.6)$$

Basic Principle To better understand the meaning of $S1$ it is worth going one step back and considering the scatter plots in figure 2.3. The model output y is plotted against two model inputs, x_1 and x_2 . Each black dot is a realization of the model output generated by a unique, random combination of x_1 and x_2 . In a first step, the initial range of the two factors is partitioned in, for this example, eight bins M , each of which encloses an equal number of realization points. The enclosed realization points are the model response of x_k being held within a sub-range c_m^k of its initial range. In this case, M ranges from zero to seven, and the bins are represented by the dotted vertical lines. This can be expressed as y being conditioned on x_{kc_m} . The second step is to calculate the mean of the model realizations within each of the eight bins $\mathbb{E}(y|x_{kc_m})$. Hence, $\mathbb{E}(y|x_{kc_m})$ is the average model response if x_k

could be fixed within c_m of its original range. $\mathbb{E}(y|x1_{c_m})$ and $\mathbb{E}(y|x2_{c_m})$ are represented by the red dots. The second step is repeated for all c_m . The course of the red dots in figure 2.3 shows the tendency of $\mathbb{E}(y|x1_{c_m})$ to increase in number, with $x1$ increasing in number. In contrast, the red dots for $x2$ are close to the horizontal line. In a third step, this tendency can be captured by computing the variance over the means of the conditional model output: $\mathbb{V}[\mathbb{E}(y|x_{kc_m})]$. For $x1$, $\mathbb{V}[\mathbb{E}(y|x_{kc_m})]$ is clearly different from zero, whereas it is nearly zero for $x2$. The last step is to normalize $\mathbb{V}[\mathbb{E}(y|x_{kc_m})]$ by the unconditioned variance of the model output $\mathbb{V}(y)$, which is similar to equation 2.4. Indeed, in accordance with Sobol (1993), when we replace the bins with defined fixations i , we can also write the first-effect index as:

$$S1x_k = \frac{\mathbb{V}[\mathbb{E}(y|x_k = x_{ki})]}{\mathbb{V}(y)}. \quad (2.7)$$

The variance Vx_k that factor x_k contributes to the total variance $\mathbb{V}(y)$ is calculated by taking the variance over the expectations of the differently conditioned model outputs. Therefore, the S1 values can also be interpreted as the amount of variance reduction in the model output - on average - if x_k could be fixed at its true value x_{ki} . In our example, $S1_{x1} = 0.87$ and $S1_{x2} = 0.002$. The variance in the model output could be reduced by 87% if $x1$ could be fixed to its true value and only by about 0.2% if $x2$ could be determined. Despite the intuition of the meaning of the S1, the principle of conditioned and unconditioned model output is valid for all GSA methods.

Sobol Indices Unfortunately, the first-effect index defined in equation 2.7 only provides information about the direct impact of the model input on the model output's variance. Although S1 might be 0, the model input may still influence the model output's variance by interacting with other model inputs. Recall equation 2.3, with its effect functions of increasing dimensions. Only the first-order term is covered by S1. However, to truly identify a model input as unimportant,

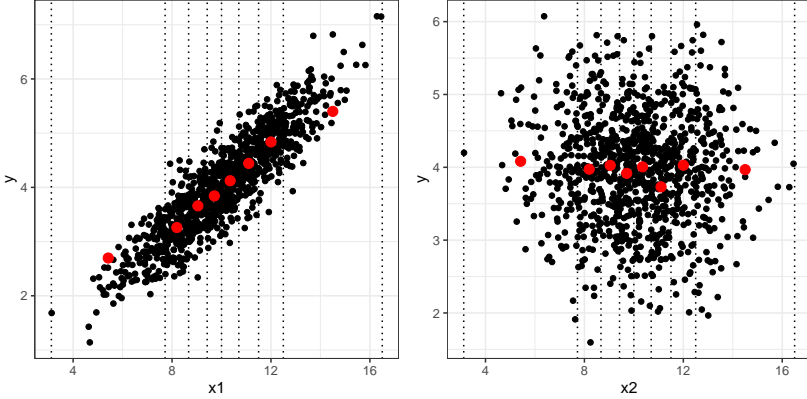


Figure 2.3: Scatter plots of the model output y against the two input factors x_1 and x_2 . The red dots represent the conditioned mean of the model output if x_1 (x_2) could be fixed within in sub-range of its original range. The sub-ranges are represented by the dotted vertical lines

a sensitivity measure taking the effect functions of all dimensions into account is required. In theory, it is possible to assess the sensitivity measures of all dimensions, but in practice, a brute force Monte Carlo assessment of the functions of higher orders is computationally not affordable. Among others, Saltelli et al. (2004) have declared that reporting S1 and ST is sufficient to describe the importance of an input factor for the model output's variance. Consider the fact that the unconditional variance can be written in terms of conditional variances decomposed into a main effect and a residual term. If we condition on all factors but one, i.e., $x_{\sim k}$ the total variance $V(y)$ can be written as:

$$\mathbb{V}(y) = \mathbb{V}[\mathbb{E}[Y|x_{\sim k}]] + \mathbb{E}[\mathbb{V}[Y|x_{\sim k}]]. \quad (2.8)$$

The main term is directly recognized as the first-order term and, and in this case, gives the direct contribution of all factors but x_k to the model output's variance. The residual term is the variance leftover if one could fix all values but x_k to their true values i . By transforming

equation 2.8, we can define the total-effect index (ST) as

$$STx_k = 1 - \frac{\mathbb{V}x_{\sim k} [\mathbb{E}x_k [Y|x_{\sim k}]]}{\mathbb{V}(y)}. \quad (2.9)$$

The reader may keep this in mind as the total-effect is the remaining variance - on average - if all factors but x_k could be fixed to their true values. In a Monte Carlo approach, ST could be calculated by fixing all parameters at a value i and then vary x_k over its whole parameter range. Since the true values are unknown, this has to be repeated for all values of i and hence all resulting combinations. At each fixing combination, the mean of the model output and afterwards its variance is calculated. This provides the direct effect of all parameters but x_k . Subtracting this from unity finally gives the variance that is left due to the variation of x_k . Note that ST includes $S1x_k$. A brute force Monte Carlo estimation of the set of S1s and STs for all parameters is beyond the scope of the available computational resources; for an explanation, see Saltelli et al. (2008). Fortunately, an estimation procedure that solves the issue of computational costs exists and is presented in section 2.4.3.1.

Properties We have defined two indices that characterize the sensitivities of the model inputs. In the following enumeration, we outline their properties.

1. The sum of all S1s is less or equal to 1, $\sum_{k=1,\dots,z} S1x_k \leq 1$
2. The sum of all STs is greater or equal to 1, $\sum_{k=1,\dots,z} STx_k \geq 1$
3. STx_k is always greater or equal to $S1x_k$, $STx_k \geq S1x_k$
4. If $STx_k = 0$, the factor x_k has no impact on the model output at all.
5. The additivity of the model is given by $1 - \sum_{k=1,\dots,z} S1x_k$.
6. $STx_k - S1x_k$ gives the involvement of x_k in interactions.

These properties are only valid under the assumption that all model inputs are independent of each other. If the reader is interested in treating correlations among the model inputs, he or she is referred to Xu and Gertner (2007) or Mara and Tarantola (2012).

2.4.3.1 Estimation: Pick and Freeze Design

Until today, the calculation of the Sobol indices, i.e., S1 and ST, has implied the use of a special sampling scheme. The Pick and Freeze design (P&F) developed by Saltelli (2002a), based on the works of Homma and Saltelli (1996) and Sobol (1993), is the most popular Monte Carlo based approach. The Sobol indices can be assessed by the computational cost of $z(N + 2)$ model evaluation instead of N^2 in a brute force approach. The idea is to generate two independent samples, A and B , of size $N \cdot z$.

$$A = \begin{pmatrix} a_1^{(1)} & a_2^{(1)} & \dots & a_k^{(1)} & \dots & a_z^{(1)} \\ a_1^{(2)} & a_2^{(2)} & \dots & a_k^{(2)} & \dots & a_z^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ a_1^{(N-1)} & a_2^{(N-1)} & \dots & a_k^{(N-1)} & \dots & a_z^{(N-1)} \\ a_1^{(N)} & a_2^{(N)} & \dots & a_k^{(N)} & \dots & a_z^{(N)} \end{pmatrix}$$

$$B = \begin{pmatrix} b_1^{(1)} & b_2^{(1)} & \dots & b_k^{(1)} & \dots & b_z^{(1)} \\ b_1^{(2)} & b_2^{(2)} & \dots & b_k^{(2)} & \dots & b_z^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_1^{(N-1)} & b_2^{(N-1)} & \dots & b_k^{(N-1)} & \dots & b_z^{(N-1)} \\ b_1^{(N)} & b_2^{(N)} & \dots & b_k^{(N)} & \dots & b_z^{(N)} \end{pmatrix}$$

Each row is one parameter combination, and each column holds the values of one factor. A third matrix C with size $z(N \cdot z)$ is generated

by taking all columns from matrix B but the k^{th} column from matrix A iteratively for all input factors z . Thus, for each model input, a matrix Cx_k of size $N \cdot z$ is generated.

$$Cx_k = \begin{pmatrix} b_1^{(1)} & b_2^{(1)} & \dots & a_k^{(1)} & \dots & b_z^{(1)} \\ b_1^{(2)} & b_2^{(2)} & \dots & a_k^{(2)} & \dots & b_z^{(2)} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_1^{(N-1)} & b_2^{(N-1)} & \dots & a_k^{(N-1)} & \dots & b_z^{(N-1)} \\ b_1^{(N)} & b_2^{(N)} & \dots & a_k^{(N)} & \dots & b_z^{(N)} \end{pmatrix}$$

With the generated model output of the three matrices y_A , y_B and y_{Cx_k} for each model input, the $S1^{PF}$ for each model input x_k can be calculated as:

$$S1x_k^{PF} = \frac{\frac{1}{N} \sum_{j=1}^N y_A^j \cdot y_{Cx_k}^j - \frac{1}{N^2} \sum_{j=1}^N y_A^j \sum_{j=1}^N y_B^j}{\frac{1}{N} \sum_{j=1}^N (y_A^j)^2 - f_0^2} \quad (2.10)$$

and ST as:

$$STx_k = 1 - \frac{\frac{1}{N} \sum_{j=1}^N y_B^j \cdot y_{C_i}^j - f_0^2}{\frac{1}{N} \sum_{j=1}^N (y_A^j)^2 - f_0^2} \quad (2.11)$$

with

$$f_0^2 = \frac{1}{N} \left(\sum_{j=1}^N y_A^j \right)^2 \quad (2.12)$$

and \cdot representing the scalar product of two vectors. From A and C , it becomes clear that the scalar product of the corresponding model output y_A and y_{C_i} is an approximation of the setting in which all factors are varied except x_k . Hence, it is an approximation of $S1$ as given in equation 2.7. If the factor x_k has an impact on the model output, high (or low) values in y_A are more often multiplied with high (or low)

values in y_{Cx_k} . Thus, the resulting scalar product for the factor x_k is high. However, if the values multiplied with each other are random, the scalar product is smaller. The same idea can be transferred to the formulation of ST in equation 2.11. Comparing Cx_k and B , the only existing difference is the k^{th} column, i.e., all factors are fixed but x_k , which is the idea of ST. Therefore, ST is high if the scalar product of y_B and y_{Cx_k} increases. Hence, if high values are multiplied by high values and vice versa. However, in the case of missing simulation runs due to program crashes, the missing samples have to be filtered in each matrix (matrix matching) to ensure that values with the same index are always multiplied.

2.4.3.2 Estimation: Given Data

Adopted from Plischke et al. (2013) and Strong et al. (2012), the estimation of the S1 from GD $S1x_k^{GD}$ is given by:

$$S1x_k^{GD} = \frac{\sum_{cm=1}^M n_m (\hat{y}_{cm} - \hat{y})^2}{\sum_{j=1}^N (y_j - \hat{y})^2}. \quad (2.13)$$

Equation 2.13 represents the idea of the partitioning scatter plots, which is explained in detail in section 2.4.3. The numerator is the squared difference between the model output's mean \hat{y}_{cm} within each partition cm and the overall model output mean \hat{y} . The denominator is the total variance of the model output. This is repeated for each partition and summed up. Hence, $S1x_k^{GD}$ relates a local variance (conditioned) to the global variance (unconditioned) over the range of all local (conditioned) values of x_k . The impact of the choice of M is discussed in section 2.4.4.1.

2.4.4 Moment-Independent Methods

In principle, MI GSA methods also rely on the idea of conditioning the model output on different x_{ki} (cf. section 2.4.3). Unlike the VB meth-

ods, MI methods directly relate the conditioned to the unconditioned model output without depending on a summarizing metric. Hence, these methods directly consider the model output and do not impose any requirements on the shape of the model output distribution. This is important if mean and variance are insufficient metrics to describe the model output’s distribution, e.g., if it is different from a normal distribution. MI methods use either the pdf or the cdf to represent the model output. The mathematical framework, the basic principle and the estimation of the different uncertainty importance measures are introduced in section 2.4.4.1 and section 2.4.4.2.

2.4.4.1 Pdf-based Method

Mathematical Framework Borgonovo (2007) introduced the MI sensitivity measure δ , which is given by

$$\delta x_k = \frac{1}{2} \mathbb{E}x_k \left[\int |f_Y(y) - f_{Y|x_k=x_i}(y)| dy \right]. \quad (2.14)$$

δ measures the statistical dependence between the model output y and the model input x_k by taking the expectation ($\mathbb{E}x_k$) over the integrals of the absolute differences between the conditional model output’s pdf $f_{Y|x_k=x_i}(y)$ and the unconditional model output’s pdf $f_Y(y)$ for all values of i .

Basic Principle The idea behind Borgonovo’s δ is built on the concept of dividing the factor input space into partitions, like the VBs methods (compare section 2.4.3), and estimating the conditional model output for each partition. Subsequently, it is set in relation to the unconditional model output. This is illustrated in figure 2.4 with actual simulation data. In figure 2.4, the whole range of simulated values (x-axis) for the aET in [mm] on July 5, 2011 is plotted against its frequency of occurrence (y-axis). The left panel shows the pdfs for the unconditional simulated aET ($f_Y(y)$, black) and the pdf of aET con-

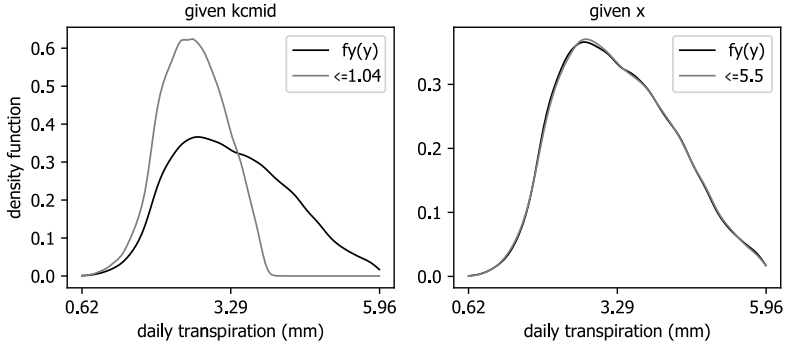


Figure 2.4: Probability density functions for the model output daily transpiration on the x-axis and the corresponding probability on the y-axis for the parameters $kcmid$ and x with one partition.

ditioned on $kcmid$ being less than or equal to 1.04 ($f_{Y|kcmid \leq 1.04}(y)$, gray). The same is plotted in the right panel for the dummy parameter x . For simplicity, we divide the input space of $kcmid$ and x only once. Therefore, we have two partitions, of which only one is shown. The optimal number of partitions M is subject to ongoing research and is briefly discussed in the paragraph "Partitioning" in this section. One notices that the parameter x does not affect the shape of $f_Y(y)$ when it is held below 5.5, whereas the probability that aET is around 3 mm is much higher if $kcmid \leq 1.04$ is compared to the unconditional pdf. Furthermore, one can observe that restricting $kcmid$ to a maximum of 1.04 does not allow for aET rates higher than 4 mm. Hence, $kcmid$ has an impact on the aET and x does not, which is of course not surprising. The δ -importance measure quantifies this intuition by integrating the absolute differences between the two pdfs. This is demonstrated in figure 2.5 (gray area). The δ -importance measure can be interpreted as the reduction in uncertainty if the modeler could restrict its belief about $kcmid$. In our example, δ_{kcmid} is 0.4 and δ_x 0.002. Note that the δ -importance measure does not allow any quantitative statement about the reduction, as in VB methods. A higher number means a higher re-

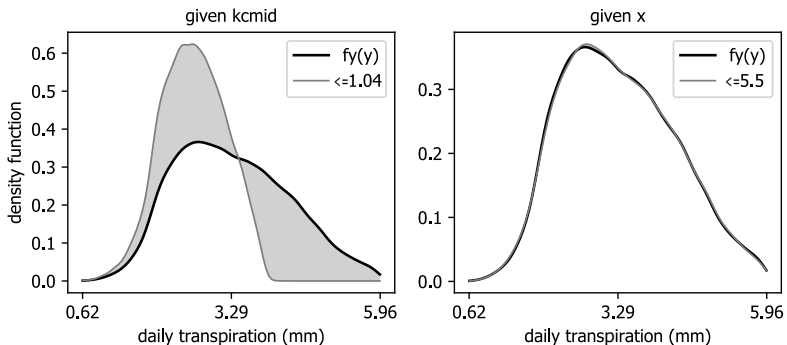


Figure 2.5: Probability density functions (pdfs) of the model output daily transpiration (mm) for the parameters k_{cmid} and x for one partition. The area represents the integral of the difference between two shown pdfs.

duction in uncertainty or a higher factor sensitivity, but its strength is relative to the other factors and gives neither a percentage in relation to the unconditional model output nor a separation between interaction and main effect. Nevertheless, the δ -importance measure has some advantageous properties, which are discussed in the next paragraph.

Properties The properties regarding the joint sensitivities of two or more factors are not of concern in our case but are listed for completeness. It is noteworthy that joint effects can be assessed by the δ -importance measure without assuming model input independence. The proofs for the following properties can be found in Borgonovo (2006) and Borgonovo (2017).

1. Each δx_k is normalized between 0 and 1.
2. δx_k of 0 implies statistical independence of y and x_k .
3. The joint sensitivity of all model inputs is 1.
4. The δ -importance measure is transformation- and scale-invariant.

5. The joint importance $\delta x_{k,j}$ equals δx_k if y is independent from x_j .

Estimation Having the basic principle and the properties in mind, the estimation of the δ -importance measure is introduced. Since we rely on the GD principle (Plischke et al., 2013), we discuss how to assess the pdfs and how to partition the input space. Other methods to estimate the δ -importance measure are described in Borgonovo (2007), who used histogram binning; Castaings et al. (2012), who used kernel-densities in an improved double-loop sampling design; Ratto et al. (2009), who used truncated Edgeworth series, and Wei et al. (2014), who rewrote δ in a copula form.

Density Estimation In this study, we assess the form of the unknown pdf of the model output (f_Y) empirically via a kernel density estimator (kde) ($\hat{f}(x)$) (Parzen, 1962) with a gaussian kernel ($K(x)$) given by:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K \frac{x - x_i}{h} \quad (2.15)$$

and

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right). \quad (2.16)$$

The bandwidth (h) is one of three parameters that need to be externally set for the estimation of the δ -importance measure. It strongly influences the estimated pdf and is more important than the choice of the kernel itself. If the bandwidth is too high, the kde oversmooths the underlying true distribution, whereas a bandwidth that is too small leads to "odd data artifacts." To estimate the bandwidth, different approximation procedures are available. We used Silverman's rule (Silverman, 1998), which accounts for the total number of data points and the dimension (number of quadrature points). Quadrature points are the second external parameter and were set, in our case, to 110. Silverman's rule is easy and cheap in terms of computing time, but, as a

rule of thumb, it is less accurate. Nevertheless, the choice of bandwidth calculation and the chosen number of quadrature points are common settings in this context. In any case, using kdes introduces a numerical error (noise) in the estimation of δ . The numerical noise is directly visible in figure 2.4. The conditional and unconditional pdfs for x differ, although we know that it is a dummy parameter with absolutely no impact. For a detailed discussion about bandwidth selection and its impact on the kdes, we redirect the reader to Sheather (2004).

Partitioning The third external parameter with a strong impact on the absolute value of the estimated δ is the number of partitions M . The number of partitions is the number of parts (bins/partitions) (compare section 2.4.3) into which the set range of the input factor is cut. For each partition c_m , a conditional pdf is estimated that reflects the probability of the model output if the input factor is restricted to the selected sub-range. If the number of partitions is too small, the impact of the input factor is imprecise because the conditional pdf refers to a large part of the input factor’s range. The estimator of δ is biased from below. In contrast, if the number of partitions is high, the number of data points within each partition is too small for the kde, and the estimated δ is upward biased (Borgonovo et al., 2016). In the example of section 2.4.4.1, an increase of the number of partitions from two to ten to 48 changes δ_{kcmid} from 0.4 to 0.42 to 0.46 and δ_x from 0.002 over 0.009 to 0.018. δ increases with an increasing number of partitions. Plischke et al. (2013) have proved that, in theory, with an increasing number of partitions and sample size, the δ -importance measure converges to its true value. However, in practice, one is limited by computational time and thus the number of partitions and simulation runs. Therefore, the number of partitions is an additional source of numerical noise in the estimation process of the δ -importance measure. In agreement with other implementations of MI methods, we use an equal partitioning of the factor range of size 48.

2.4.4.2 Cdf-based Methods

Instead of pdfs, one can use cdfs to describe the model output distribution. This relieves the modeler from finding a suitable bandwidth and of an adequate number of quadrature points. Measuring the sensitivity of an input factor by using cdfs is comparable to pdf-based methods. This section highlights the differences between MI methods and is structured as section 2.4.4.1.

Mathematical Framework When cdfs are used to represent the model output distribution, one can choose between different uncertainty importance measures. We focus on two uncertainty importance measures that are based on two well established distance measures, i.e., the Kolmogorov-Smirnov and the β^{kui} per metric. Independent of the metric (d) Baucells and Borgonovo (2013) defined the importance measure β for any input factor x_k as

$$\beta x_k = \mathbb{E} \left[d \{ F_Y(y), F_{Y|x_k=x_{ki}}(y) \} \right] \quad (2.17)$$

with $F_Y(y)$ representing the unconditional cdf of the model output y and $F_{Y|x_{ki}}(y)$ the conditional cdf, and \mathbb{E} is the expectation of d for all values of i . In case of the Kolmogorov-Smirnov metric, d is substituted by

$$d_{ks} = \sup_y |F_Y(y) - F_{Y|x_k=x_{ki}}(y)| \quad (2.18)$$

with sup representing the supremum. Analogously, for the Kuiper metric, we obtain

$$d_{kui} = \left[\sup_y \{ F_Y(y) - F_{Y|x_k=x_{ki}}(y) \} + \sup_y \{ F_{Y|x_k=x_{ki}}(y) - F_Y(y) \} \right]. \quad (2.19)$$

The Kolmogorov-Smirnov metric is insensitive to deviations that occur in the tails of a distribution (Mason and Schuenemeyer, 1983), which is improved by the Kuiper metric. For further information about the

Kolmogorov-Smirnov- and the Kuiper metric, we refer the reader to Anderson and Darling (1952) and Crnkovic and Drachman (1996).

Basic Principle β^{kui} and β^{ks} relate the unconditional and conditional model output (cf. sections 2.4.3 and 2.4.4.1). Again, the factor range is subdivided, and the conditional model output cdf is estimated for each sub-range, i.e., partition. Figure 2.6 shows the same model output as figure 2.4 and 2.5. This time, aET on the y-axis is plotted against the cumulative probability on the x-axis. The conditional cdfs are plotted for three partitions labeled with their upper boundary. The vertical bars represent the Kolmogorov-Smirnov and Kuiper metrics, respectively (cf. equation 2.18 and 2.19.) It can be seen that $kcmid$ has an impact on aET , whereas x does not. In our example,

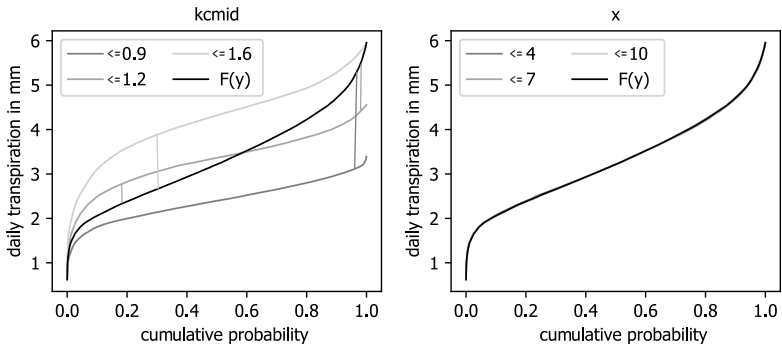


Figure 2.6: Cumulative distribution functions of daily actual transpiration unconditional and conditional on parameters $kcmid$ (left) and x (right). Legend entries refer to the maximum value of the specific partition. The vertical bars represent the Kolmogorov-Smirnov and the Kuiper metric, whereby the vertical bar on the left only refers to the Kuiper metric.

the Kuiper metric differs from the Kolmogorov-Smirnov metric only if $kcmid$ is between 0.9 and 1.2 . For the other two partitions, one of the two summands in equation 2.19 is always zero. Nevertheless, the resulting β^{kui} and β^{ks} , which are the averaged metrics over all parti-

tions, differ. The two importance measures β^{kui} and β^{ks} are 0.41 and 0.36 for *kcmid* and 0.009 and 0.007 for *x*, respectively. Although the reduction in numerical noise by switching from pdfs to cdfs is invisible in the graphs, it still appears in the numbers. Coping with this error is the topic of section 2.4.4.3.

Properties MI importance measures based on cdfs have the same properties as the δ -importance measure (cf. section 2.4.4.1). They are normalized between 0 and 1, are scale- and transformation-invariant and nullity implies that the input factor and the model output are independent.

Estimation To estimate β^{kui} and β^{ks} , the empirical conditional and unconditional cdfs have to be constructed. The empirical unconditional cdf ($\hat{F}_Y(y)$) is given by

$$\hat{F}_Y(y) = \frac{1}{N} \sum_{i=1}^N H(y - y_i), \quad H(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1, & \text{if } x \geq 0, \end{cases} \quad (2.20)$$

where $H(x)$ is the Heaviside function. The empirical conditional cdf is then written as

$$\hat{F}_{Y|x_k \in c_m^i}(y) = \frac{1}{n_m^i} \sum_{i: x_i \in C_m^i} H(y - y_i) \quad (2.21)$$

for all partitions c_m . For the implementation, the model output is sorted increasingly, and the corresponding cumulative probabilities are calculated by dividing the $[0, 1]$ interval by the number of data points in the model output. Again, the accuracy of the estimation of the conditional cdf and hence the accuracy of the calculated values of β^{kui} and β^{ks} depends on the number of partitions. For an explanation, see section 2.4.4.1. The influence of partitioning schemes on the estimation of uncertainty importance measures has been discussed by Plischke (2012).

2.4.4.3 Critical Values

In section 2.4.4.1 and section 2.4.4.2, we saw that the MI methods are afflicted with a numerical error that results from the choice of the number of partitions, quadrature points and, in the case of pdf-based methods, the bandwidth. In order to still use the moment-independent sensitivity indices (SI^{MI}) for the Factor Fixing setting, we have to consider this error in the analysis. For β^{ks} , we can use the Kolmogorov-Smirnov Test, which tests if two variates follow the same distribution. The conditional cdf is significantly different from the unconditional cdf if

$$d_{ks} \geq K_\alpha \sqrt{\frac{1}{N} + \frac{1}{N_m}} \quad (2.22)$$

where the d_{ks} is $\max_y \left| \hat{F}_Y(y) - \hat{F}_Y|_{x_k \in c_m}(y) \right|$, K_α is the upper boundary of the Kolmogorov distribution, and N and N_m are the number of data points of each empirical cdf. Since the aim is to reduce the numerical noise in the estimates for SI^{MI} , we have to select an adequate value for K_α . To set K_α , we can exploit our knowledge that the dummy parameter x is non-influential in setting a critical value. It is calculated from

$$K_{x_k}^{ks} = \max_{m=1,2,\dots,M} \left(\frac{d_{ks}}{\sqrt{\frac{1}{N} + \frac{1}{N_m}}} \right). \quad (2.23)$$

Hence, the critical value is the maximum value of the Kolmogorov-Smirnov distance over all evaluated partitions c_m . Substituting d_{ks} with the estimates of equation 2.19 yields the empirical critical value $K_{x_k}^{kui}$ for β^{kui} . For the δ -importance measure, the critical value can be assessed by multiplying the fraction of the right side of the inequality with 0.5, which takes into account that the δ importance measure is based on integration. Based on the critical value, one can calculate the threshold (K_{crit}), above which model input and model output cannot

be considered independent. Furthermore, one can calculate the probability α and the size of the corresponding confidence interval at which the conditional model output distribution would be classified as not significantly different from the unconditional distribution of the model output. To assess the threshold, the critical values $K_{x_k}^*$ are inserted in equation 2.22. In addition, a bias filter proposed by Plischke et al. (2013) based on the work of Efron and Tibshirani (1993) is used to reduce the numerical noise in the estimates of δ . The bias-reduced estimator, $\hat{\delta}$, is given by

$$\hat{\delta} = 2 \mathbb{E}[\delta] - \delta_{pt} \quad (2.24)$$

where δ_{pt} is the estimate for each bootstrap sample (cf. section 2.4.5) and $\mathbb{E}[\delta]$ is the expectation over all bootstrap replicates.

2.4.5 Confidence Intervals

Bootstrapping with replacement is used to assess confidence intervals for the Sobol indices (cf. section 2.4.3), the SI^{MI} values and $\hat{\delta}$. For the sensitivity indices from the Pick and Freeze design (SI^{PF}), the percentile method is used. Here, the endpoints for the 95% interval are 2.5% and 97.5%, respectively. Confidence intervals for SI^{MI} , however, are estimated with the moment method, which relies on large sample theory and assumes a symmetric 95% interval. Both methods are adopted from Archer et al. (1997).

2.4.6 CUSUNORO

CUSUNORO plots (CUmulative SUms of NOrmalized Reordered Output) (Plischke, 2012) visualize the dependency of the model output on the model input. Furthermore, the impact direction ω of a certain model parameter x_k on the model output can be identified as well as the linearity 1_n of the model parameters' influence. Loosely speaking, ω assesses if the increase of a given parameter affects the model output

positively or negatively. The 1_n value of the impact reveals whether the parameter affects the model output equally over its entire range or is important only in a sub-range. The points of the CUSUNORO-curve $c_{x_k}(i)$ for each parameter x_k are calculated as follows:

$$c_{x_k}(i) = \frac{1}{n \cdot \sqrt{V(Y)}} \sum_{i=1}^j (Y_i - \bar{Y}). \quad (2.25)$$

To construct the values of the CUSUNORO plot, the model output mean \bar{Y} is subtracted from each model output realization $Y(i)$ and divided by $1/n \cdot \sqrt{V(Y)}$. This gives the normalized model output Y_{norm} , which is reordered by the increasing parameter x_k . This procedure is repeated separately for each model input. Plotting the cumulative sums of the normalized reordered output on the y-axis against the empirical cumulative distribution functions (ecdf) of the model input on the x-axis gives the CUSUNORO plot. The start and end points of the CUSUNORO curve are 0 by definition. Since visual inspection becomes infeasible when the model setup is larger (regarding the TGVs and the time resolution), we extract the x-y-coordinates of the extrema from the CUSUNORO-curve. Hence, we save the information about the degree of the parameter's impact (y-value of the extrema), the direction of the impact (if it is a maximum or minimum) and the linearity of the impact (x-value). If the x-value of the extremum is at 0.5, the impact is linear. If the extremum is reached at values ≤ 0.5 , the parameter's impact is larger if the parameter value is small and vice versa. An example is given in figure 2.7 for the TGV *wc30* at July 5, 2011. The CUSUNORO curves are constructed for the parameters *al1*, *al2*, *kcmid* and *x*. Curves above the red horizontal line, in the light gray area mean that the corresponding parameters have a negative impact on *wc30*, i.e., the soil water content decreases as *kcmid* or *al1* increases in value. *al2* has a positive effect on *wc30*. The dashed, vertical red line at $1_n = 0.55$ indicates that *kcmid* impacts *wc30* more strongly with higher values.

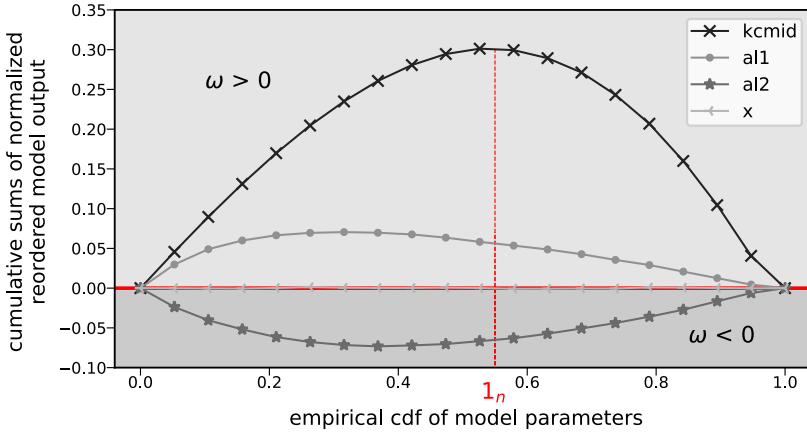


Figure 2.7: CUSUNORO plot for the target variable water content in 30 cm (wc_{30}) and the parameters $al1$, $al2$, $kcmid$ and x . Each curve represents one parameter. CUSUNORO curves in the dark gray area and above the red horizontal line have a ω value < 0 . CUSUNORO curves in the light gray area and below the red horizontal line have a ω value > 0 . The red, vertical, dashed line represents the linearity factor 1_n . $kcmid$ has its largest impact at $1_n=0.55$.

2.5 Data Processing Chain

The DPC describes the data flow from raw field measurements to the aggregated set of the different SIs. The DPC is divided into two parts: data preparation and data post-processing. In data preparation, data refers to field measurements and data generated by simulations, whereas in data post-processing, data always refers to the set of already gained SIs. The organization of this section is as follows: first, an overview of the different DPC stages is given. In section 2.5.1, DPC with its stages (field) data pre-processing (1), data evaluation (3) and data generation (2) are described. Data cleaning (4) and data analysis (5) are explained in more detail in sections 2.5.2.1 and 2.5.2. Exoge-

nous analysis settings are given in section 2.5.3, and the frameworks used are described in section 2.5.4.

Overview From the raw field data to averaged, weighted SIs, a DPC had to be established. It is divided into the following five steps:

1. Data process chain - Data preparation
 - 1.1. Data pre-processing: preparation of the raw field data (cf. section 2.2), implementation of an XN-Linux version, automatic model and project setup
 - 1.2. Data generation: sample creation and simulation runs in a high-performance computing environment
 - 1.3. Data evaluation: selection of analysis dates and variables, computation of different SIs in a high-performance computing environment
2. Data process chain - Data post-processing
 - 2.1. Data-cleaning: filtering of SIs
 - 2.2. Data-analysis: weighting, averaging, evaluating and plotting

2.5.1 Data Process Chain - Data Preparation

Data preparation is outlined in figure 2.8. First, the XN software code base had to be ported to a Linux platform as the prerequisite for running on the available High Performance Computing Clusters. This is needed for the generation of Monte Carlo simulations. Secondly, the XN projects had to be set up which is done automatically with a software module written in the python language. The purpose of **data pre-processing** was to prepare raw files such that they can be automatically imported into the XN. At this step, the model configuration had also been determined. With the XN-projects and configuration ready, the data pre-processing was finished.

The next step was data generation (cf. section 2.5.1.1). It includes

the generation of optimized Latin Hypercube Samples (LHSs) and the generation of model output by running the XN for all parameter combinations. The simulations were executed with mpirun. Two independent LHSs (McKay et al., 1979) were generated, one for the sensitivity indices from Given Data (SI^{GD}) and one following the scheme of the P&F (cf. section 2.4.3.1). Subsequently, the raw model output was evaluated, i.e., it was thinned out to every sixth day of the model output, and the SIs were calculated for the targeted model outputs. This was also done in parallel on four nodes with 28 cores each and refers to the third step, data evaluation. The time-series of the six considered SIs and each target variable was finally transferred back to a personal computer, where the data post-processing (see section 2.5.2) was performed.

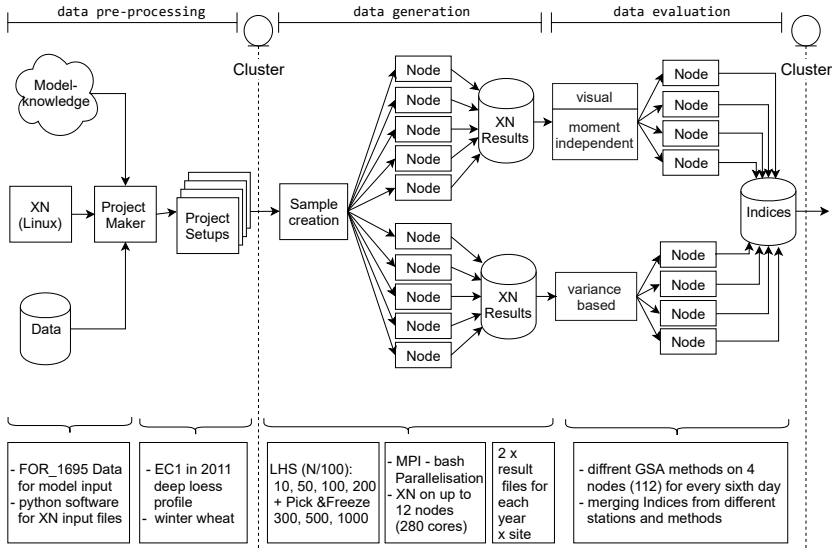


Figure 2.8: Schematic overview of the data processing including (field) data pre-processing, data generation and data processing.

2.5.1.1 Data Generation

The parameter combinations with which the XN is run are always directly generated on the High Performance Computing cluster by an optimized LHS. It has been shown to adequately sample high-dimensional parameter spaces by combining the advantages of stratified and random Monte Carlo sampling techniques. The parameter space is divided into n disjoint intervals of equal probability. From each interval, one parameter value is selected at random and combined with other parameter values without replacement (Helton and Davis, 2002, 2003). For the optimization of the LHSs, we used heuristic Simulated Annealing to improve the uniform distribution of the sampling points in parameter space. It is adapted from the Metropolis-Hastings algorithm (Metropolis et al., 1953) and inspired by the idea of a cooling metal that aims to converge to an energetically low state. Thus, the temperature describes the time-dependent acceptance probability of moving from the current state to the new state. We configured the algorithm with a geometric profile for the temperature and used the C^2 metric (Jin et al., 2005) as the space-filling optimization criteria. For mathematical descriptions of Simulated Annealing and the comparison of different space-filling criteria in the context of the design of experiments, see Damblin et al. (2013) and the literature cited within.

2.5.2 Data Process Chain - Data Post-Processing

2.5.2.1 Data Cleaning

Data post-processing consists of data filtering and data evaluation. Hereafter, the term data refers to the SIs. Figure 2.9 provides an overview of the five-stage approach used to clean the calculated SIs of numerical noise. This five-stage approach has been iteratively developed while conducting the GSA and the evaluation of the SIs. Step 1 is only applied to the δ -importance measure because it is the only SI that is affected by this error. In practice, δ values can be larger

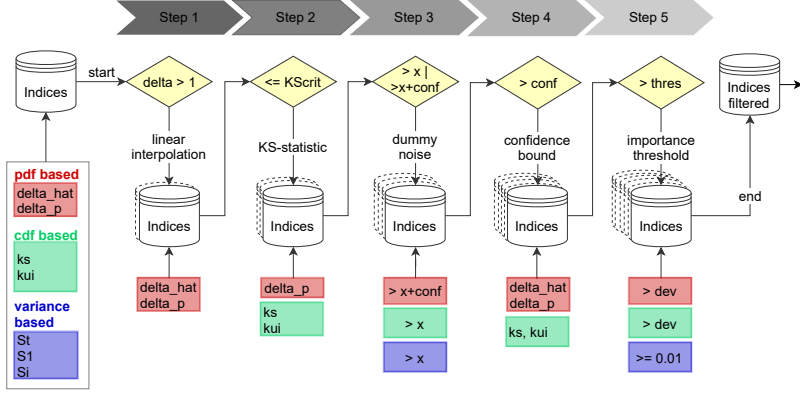


Figure 2.9: Data post-processing: filtering the sensitivity indices by different criteria in a five-stage approach to reduce the numerical noise. Pdf-based importance measures $\hat{\delta}$ and δ^{pt} are given in red, green represents the cdf-based importance measures β^{kui} and β^{ks} , and blue stands for the variance based indices from given data and the Pick and Freeze design.

than 1; the reason for this is discussed in chapter IV. Step 2 filters out all parameters where β^{ks} , β^{kui} and δ_{p} have smaller values than the calculated K_{crit} (cf. section 2.4.4.3). In step 3, the value of the sensitivity index (idxvalue) of the dummy parameter x is subtracted from each idxvalue for all SIs because its idxvalue is considered the minimum numerical noise within the SI estimation process. In the case of pdf-based SIs, the confidence interval of x is added to x 's value. In step 4, all SI^{MI} that, after subtracting x , are smaller than their own upper confidence bound are set to 0 and classified as non-influential. Step 5 is applied to all SIs. All parameters that are still nonzero after passing steps 1 to 4 are finally compared to a threshold. Variance-based sensitivity indices (SI^{VB}) have to be greater than 0.01, denoting an influence of 1% on the model output's variance. This is a common threshold used in the literature. For SI^{MI} , it is assumed that the KS-statistic is sufficient to distinguish influential from non-influential model inputs. However, we exploit our knowledge of the parameter

dependencies of the target variable *dev*. Here, we definitely know from the equations which parameters are influential and which are not. The idxvalues of the non-influential parameters are not "allowed" to be different from 0. The maximum idxvalue of non-influential parameters that nevertheless shows an influence on the TGV *dev* serves as a last threshold. The idxvalues of all SI^{MI} for all TGVs have to be greater than this threshold to be classified as an influential parameter.

2.5.2.2 Data Analysis

Weighting, Averaging, Ranking and Correlating To further summarize the already thinned out weekly resolved time series of SIs, they are averaged to identify the parameters that have, on average, the largest impact on XN-CERES. Further, the SIs are weighted to obtain the most influential parameters because a parameter can have a high SI while the overall uncertainty in the model output is low. Hence, a parameter with a low SI in a situation of a high model output uncertainty can be much more important to the result of a simulation. How to address multivariate and multidimensional (time and space) model output is subject to ongoing research. A few studies exist that deal with the calculation of SIs in multivariate and/or multidimensional cases (Lamboni et al., 2011; Marrel et al., 2016; Gamboa et al., 2014). However, the suggested methods are all related only to SI^{VB} and use the concept of Principle Component Analysis. Since our case is both multidimensional (time series) and multivariate (target variables), and we additionally want to compare MI uncertainty measures with VB uncertainty measures, we decided to apply a simple approach. For each target variable and SI, the weighted SI of each parameter (SIx_k^w) is calculated by:

$$SIx_k^w = \frac{\sum_{d=1}^D (SI_x^d * \mathbb{V}[y_d])}{\sum_{d=1}^D \mathbb{V}[y_d]}, \quad (2.26)$$

where d is the time window chosen as the average interval, D is the total number of time intervals and $\mathbb{V}[y_d]$ is the variance of the model output

in the interval d . Afterward, the weighted SIs are averaged according to set time interval d . The weighted and averaged SIs are sorted in decreasing order. Hence, the parameter with the highest SI receives the lowest rank. SI with the same value are assigned the average rank. Therefore, half-ranks are possible. Furthermore, all zero values, i.e., all non-influential parameters, are assigned the same rank. However, parameters with a low rank have the largest impact, and vice versa. As a last step, we calculate pairwise correlations of the parameter ranks over all dates for all SIs to analyze if they prioritize and exclude the same parameters. The resulting rank correlation is a useful statistic for the second objective, i.e., GSA-comparison. We chose the Pearson correlation coefficient (Pearson, 1895) instead of the often-used Spearman rank correlation coefficient (Spearman, 1904) because we are also interested in correctly identifying the non-influential parameters.

2.5.2.3 Index Convergence

Convergence of the $idxvalue$ and the assigned parameter ranks provides information about the numerical stability of the estimated SIs. Knowing about the uncertainty in the $idxvalues$ and the ranking is an essential part of the uncertainty analysis and GSA. We check the convergence of the $idxvalue$ as well as of their ranks. The calculation of $idxvalue$ convergence (cvg_{idx}) is based on Sarrazin et al. (2016) and given by

$$cvg_{idx} = \max_{k=1, \dots, z} (SIx_k^{ub} - SIx_k^{lb}), \quad (2.27)$$

with ub representing the upper bound of the bootstrap confidence interval and lb the lower bound, respectively. If $cvg_{idx} \leq 0.5$, Sarrazin et al. (2016) have suggested to assume index convergence. cvg_{idx} is calculated for each parameter, LHS size and TGV. Since ranking and saving all SIs from each bootstrap sample is not reasonable in our case due to the high amount of data and time needed for the ranking, the method of Sarrazin et al. (2016) to check rank convergence was adapted. Thus, we compared the ranks originating from two different LHS sizes

to calculate the ranking convergence ($conv_{rank}$), which is given by:

$$cvgr_k = \sum_{k=1}^z |Rx_k^{lhs1} - Rx_k^{lhs2}| * \frac{\max_{k=1}^z SIx_k}{\sum_{k=1}^z SIx_k^{lhs}}, \quad (2.28)$$

where $|Rx_k^{lhs1} - Rx_k^{lhs2}|$ is the difference between the two ranks of a certain parameter Rx_k originating from two LHS of different size (lhs1, lhs2). The deviation is weighted by the ratio of the overall maximum of the idxvalue of the compared sample sizes to the sum of the maximum idxvalue of all parameters of the two compared samples $lhs1$ and $lhs2$ separately for each date and target variable. Larger deviations in the top-ranked parameters are considered more important than in the lower ranks. The 0.95 quantile of ($cvgr_k$) is taken as a convergence criterion. If it is below 1, the aberration between the ranks of two LHS sizes is, on average, less than one rank. In our case, we reference all samples sizes ($lhs1$) to the maximum sample size $lhs2$, and the 95-quantile is calculated considering the set of $cvgr_k$ of all TGVs and dates.

2.5.3 Summary and Analysis Settings

In chapter II, we defined six SIs: three VB sensitivity measures and three MI sensitivity measures. For simplicity, these SIs are further distinguished by the model input design and the model output representation. All SIs that are calculated from GD, δ , $\hat{\delta}$, β^{ks} , β^{kui} and first-effect index from Given Data ($S1^{GD}$) are summarized as SI^{GD} , whereas SI^{PF} refers to $S1^{PF}$ and ST. Further, a distinction is made between SI^{VB} and SI^{MI} . The SI^{VB} can be further split based on the model input. The introduced SI^{MI} include δ^{pt} and $\hat{\delta}$ estimated from pdfs and β^{kui} and β^{ks} from cdfs. For all sensitivity measures, we provided the mathematical framework, the basic principle, their properties and their estimation procedures. We have also discussed the extensive DPC.

In addition to the theoretical framework, some external parameters have to be set as exogenous in advance. These parameters are listed in

table 2.6. Furthermore, descriptive statistics such as the four moments of a distribution, mean, median, skewness and excess kurtosis as well as maximum, minimum and the 5-, 25-, 75- and 95- quantiles for each TGV and date are recorded. In this study, kurtosis always refers to excess kurtosis.

Table 2.6: External parameters, their abbreviations (abbr.) and their set value for this work.

external parameter	abbreviation	value
number of partitions	M	48
evaluated days	D	52
quadrature points	-	110
bootstrap sample size	-	$10^2, 100^1$
LHS sizes	N	10000, 50000, 100000, 200000, 500000 ¹ , 1000000 ¹

¹ Sample sizes only for SI^{PF}

² Sample sizes only for SI^{GD}

2.5.4 Implementation

All MI methods and data pre- and post processing were implemented in python version 3.5. The δ estimation routines were partly adopted from Herman and Usher (2017). The implementation of the cdf-based importance measures and the implementation of the method were inspired by matlab codes or personal messages of the researcher Dr. Plischke from TU Claustahl. For the calculation of SI^{PF} , the open source C software package OpenTURNS (Baudin et al., 2015) was used. The implementation was performed in collaboration with the researcher Pamphile Roy from CERFACS Toulouse. For the generation and optimization of the LHSs, the OpenTURNS classes LHSexperiment and SpaceFillingC2 were used. All simulations and calculations were done on the bwUniCluster. For hardware specifications, see Haefner and Hartmut (2019). The routine mpirun of the OpenMPI Library (1.8.7-intel-14.0) (Gabriel et al., 2004) was used for the parallelization of the simulations and SI calculations. The whole process of data management - copying, concatenating and splitting files on the cluster itself - was organized

into various bash routines, whereas the XN-project generation and manipulation were done with a python script. On the bwUniCluster itself, python version 3.4 was used.

III. Results

This chapter is subdivided into three major sections. Section 3.1 describes the model output and its distribution of the different TGVs. Section 3.2 shows the relevant results regarding objective 2, GSA-Comparison (cf. 1.2), i.e., the use of computer resources (3.2.2), the convergence of the individual SIs (3.2.1) and their comparison in regard to the parameter ranking (3.2.3). Section 3.3 presents the results concerning objective 1 (cf. 1.2). The additivity of the model is shown in section 3.3.1, and the actual parameter sensitivities are given in section 3.3.2. The results are subdivided by the target groups, and within each group, they are expanded by the individual TGVs. A general description of the figures' arrangement is given in section 3.1. Most of the time, the TGV *dev*, i.e., the crops' development stage, is not indicated because it is used in the filtering of the DPC. The *idxvalue* and *ranks* always refer to the weighted SIs. The results for $\hat{\delta}$ are not shown because at high sample sizes the numerical noise could not be filtered. The results of a GSA strongly depend on the set parameter ranges (Shin et al., 2013). Hence, all results of this study are conditioned on the set parameter ranges.

3.1 Model Output Description

To understand the range of the simulation results for the different TGVs and to ensure that simulated values within the LHS are within realistic

bounds, descriptive statistics of the SR are shown in figure 3.1. The unit and label of the y-axis and/or x-axis refer to the particular facet name, i.e., the sub-plot label. The TGVs are always arranged in the same way; the crop group in the first row is followed by the water, nitrogen and flux groups. The red dots refer to the simulation mean and the black dots refer to the median. The gray lines represent the minimum and maximum simulation result and the gray areas are the 50% and the 95% prediction intervals, respectively. The simulation results are plotted against the evaluated dates to account for their variation over the simulation period. For the units of the y-axis, see table 2.3.

In general, diverging mean and median and the position of the prediction intervals in relation to the minimum and maximum simulated values give information about the general tendency and the alignment of the model output within its simulation range. From figure 3.1 it can be seen that, in each simulation, the crop emerges and develops since the minimum for *dev*, *LAI* and *VegBm* is greater than 0. However, the crop does not reach the generative phase in all simulations. This can be seen from the fact that the minimum for *genBm* and *N-gBm*, which is 0, and a minimum for *dev* of about 65. TGVs of the plant group are fairly well centered with regard to mean, median and prediction intervals.

The last statement also holds for the soil water content. Whereas the maximum soil water content is quite stable, the center and the prediction intervals show a slight tendency towards higher simulated soil water contents at the simulation start and end. The picture for the matric potential is totally different. The simulation range during the vegetation period and at the simulation start covers pF-values from 0 to 4, but half of the SRs are below pF 3 for *mp15* and around pF 3 for *mp75*. High matric potentials above 3.3 are rare, and very high matric potentials above pF 3.8 are very rare (2.5% of the simulations). In all, *mp75* is simulated more often a bit drier than *mp15*. During the winter, the SR have a much smaller range between pF 0 and 1.

Concerning the center and the alignment, the simulated soil NO_3^- and

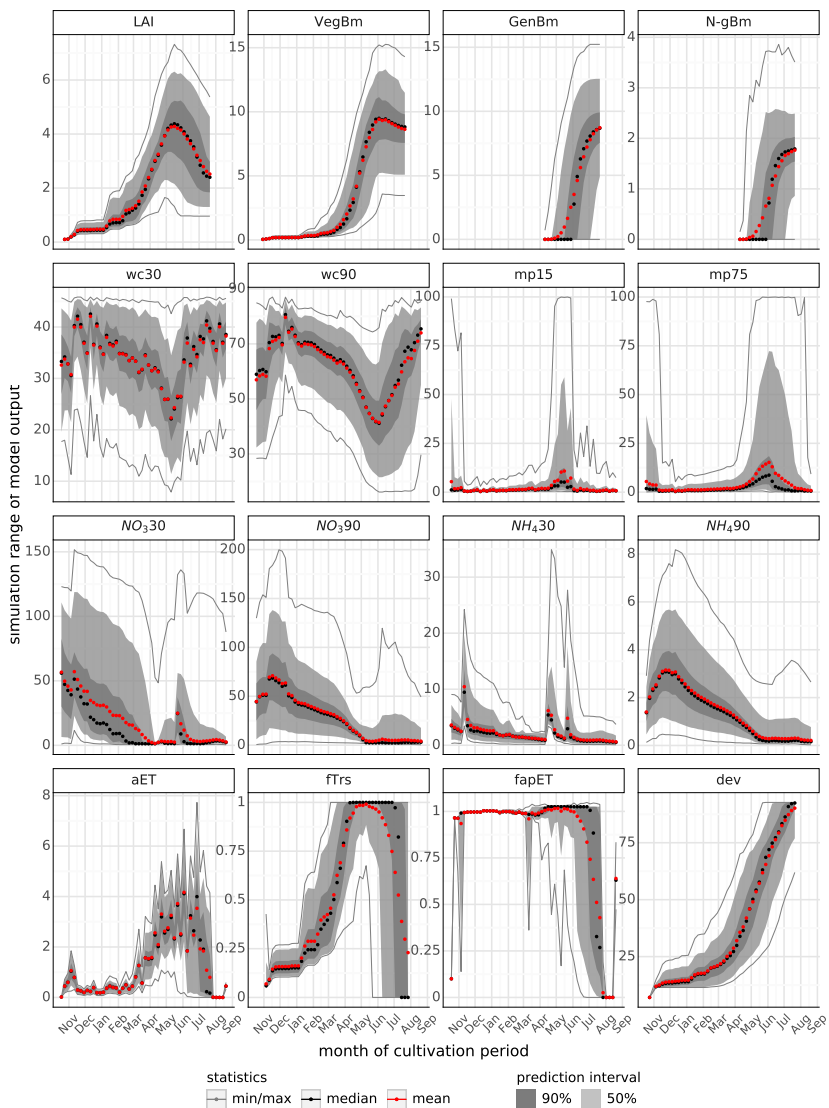


Figure 3.1: Minimum, maximum, median, mean and prediction intervals of the simulation results for each evaluated date on the x-axis faceted by the individual target variables.

NH_4^+ contents in the sub soil behave similar to the soil water content. However, in the upper soil, they show a clear tendency towards lower values. Nevertheless, all TGVs of the nitrogen group show the same time pattern. After a short increase in November or December, they constantly decrease until April. At that point, the range of simulated soil NH_4^+ and NO_3^- contents is, according to the 95%-quantile, small, which means that all simulations approach the same state. From April on, one can observe one to two additional peaks where the simulated range and soil NH_4^+ and NO_3^- contents increase again. These increases coincide with the fertilization dates (cf. table 2.2). In general, this pattern is more distinct for TGV_{30}^N . The simulated NO_3^- content in the soil ranges from 0 to 200 kg N per hectare, and the NH_4^+ content from 0 to 40 or from 0 to 8 kg N per hectare for the top and sub soil, respectively.

Last, the SRs of the TGVs of the flux group are the most "special" ones. Whereas the target variables of the flux group that form a fraction (TGV_{frc}^{fx}), $fTrs$ and $fapET$ show no or almost no variation in the SRs, and aET has the most time-varying SRs. Only at the end of the cultivation period is the total range of possible values for TGV_{frc}^{fx} covered by the simulations. However, the reverse could also be true. Regarding $fTrs$, the 2.5-quantile does not diverge from 1 in April and May, although the minimum value is lower (0.7). Hence, a very small fraction of the sample leads to another SR than at least 97.5% of the parameter combinations. $fapET$ shows no variation from December to March, and only little variation until June in 97.5% of the simulations. In the end, it is noteworthy that the TGVs of the plant and flux group show the same development of the prediction intervals. They start quite small and expand in the cultivation period. Hence, the variation in the simulation results is close to 0 when the crop has not yet started or starts to grow and increases with progressing crop development.

In figure 3.2, the temporal evolution of skewness and excess kurtosis for the TGVs' model output distribution is shown. The individual TGVs again form the facets, i.e., the sub-plots. The black lines comprise the

range from -1 to 1. Skewness and kurtosis are the third and fourth moment of any distribution, and they are measures of the symmetry and shape, respectively, of the underlying distribution. Additionally,

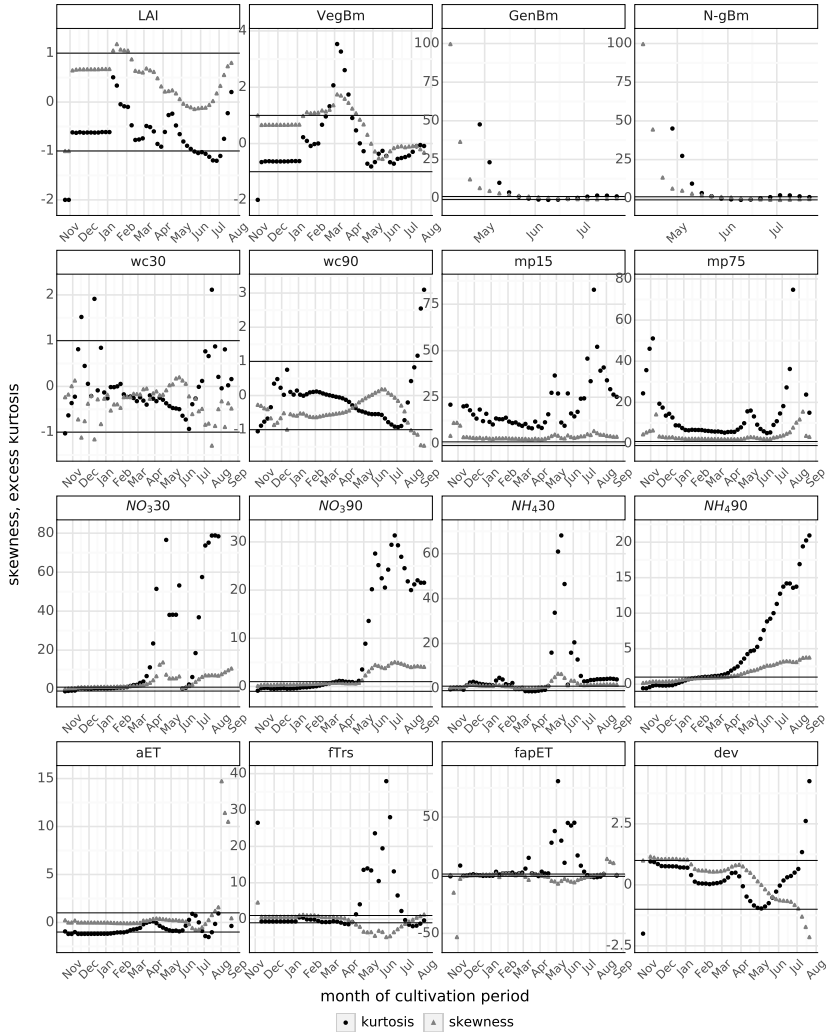


Figure 3.2: Time series of the model output distributions' skewness and kurtosis over the simulation period for the target variables.

table 3.1 gives the median, minimum and maximum values for excess kurtosis because, in the case of the fourth moment, the extreme values are too high for graphical representation. Distributions with negative excess kurtosis are platykurtic (broad), and with positive kurtosis, they are leptokurtic (slender). A negative skew means that the distribution's mass is centered to the right, whereas a positive skew refers to distributions where the mass is centered to the left. The values are always referenced with regard to a normal distribution.

The TGVs *wc30* and *wc90* as well as *dev*, *LAI*, *VegBm* and *aET* are inconspicuous in regard to skewness and kurtosis. Their values are mostly between 1 and -1. In addition, their distributions only slightly tends to be "broad," with negative medians and small minimum and maximum values for kurtosis. However, there are two exceptions. First, the distribution of *VegBm* is more slender and left-skewed in March. Second, *dev* evolves to a right-skewed and leptokurtic distribution towards the end of the vegetation period. For the other three TGVs, the maximum and minimum values occur mostly at the beginning and/or the end of the simulation period.

In conjunction with *genBm*, *N-gBm* shows a wide range of possible distribution appearances regarding kurtosis and skewness, whereby especially the high maximum kurtosis values in April and May are striking. Hence, at the outset of the yield formation, the distributions are highly centered and peak at low yields and low N contents. Over time, their distributions converge towards the shape and symmetry of a normal distribution.

TGVs of the nitrogen group tend towards leptokurtic, left-skewed distributions, i.e., to smaller N contents in the soil. Especially for TGV_{90}^N , a clear increase over time for both kurtosis and skewness is observable, whereas TGV_{30}^N have a more dynamic pattern over time. Repeatedly, kurtosis and skewness increase and decrease from February onwards.

TGVs of the flux group have a kurtosis larger than 1 from April to June, which is accompanied by right-skewed distributions. Hence, most of the water is transpired and not evaporated, and most of the time, there is

no reduction in ET.

Finally, independent of the soil depth, the distribution of the matrix potentials are highly centered on low potentials, i.e., they are highly leptokurtic - on average and maximally - and highly skewed to the left. Note that TGVs of the nitrogen and plant group can be bimodal during the vegetation period, which is not reflected in any descriptive statistic.

Table 3.1: Median, maximum (max) and minimum (min) values of kurtosis of the model output distribution over the simulation period for the individual target variables.

target variable	median	min	max	target variable	median	min	max
crop				Water			
dev	0.3	-2.0	4.25	wc30	-0.22	-1.03	2.11
LAI	-0.62	-2.0	0.51	wc90	-0.07	-1.05	3.10
VegBm	-0.36	-2.0	3.53	mp15	16.17	8.28	222.32
GenBm	1.29	-1.19	9333	mp75	9.10	5.08	334.91
N-gBm	1.63	-1.0	9333				
Flux				Nitrogen			
aET	-0.87	-1.52	297.2	NO_330	2.77	-1.13	268.3
fTrs	-0.57	-1.94	37.93	NO_390	0.93	-0.87	31.34
fapET	1.63	-1.79	3315	NH_430	2.46	-1.28	68.14
				NN_490	1.47	-0.58	20.94

3.2 GSA-Comparison

3.2.1 Convergence

In section 2.5.2.2, we defined two criteria to check whether the SIs converged. In figure 3.3, the convergence criteria for the idx values defined in equation 2.27 are shown. cvg_{idx} on the log-scaled y-axis is plotted against the sample size on the x-axis. Further, the cvg_{idx} for each TGV is shown in its own facet, and the boxplots represent the variance of cvg_{idx} within the time series. The order of the TGVs is known. The colors represent the seven different SIs, and the dotted line is the convergence threshold suggested by Sarrazin et al. (2016). One can observe that the values of the SI^{PF} hardly converge, independent of the

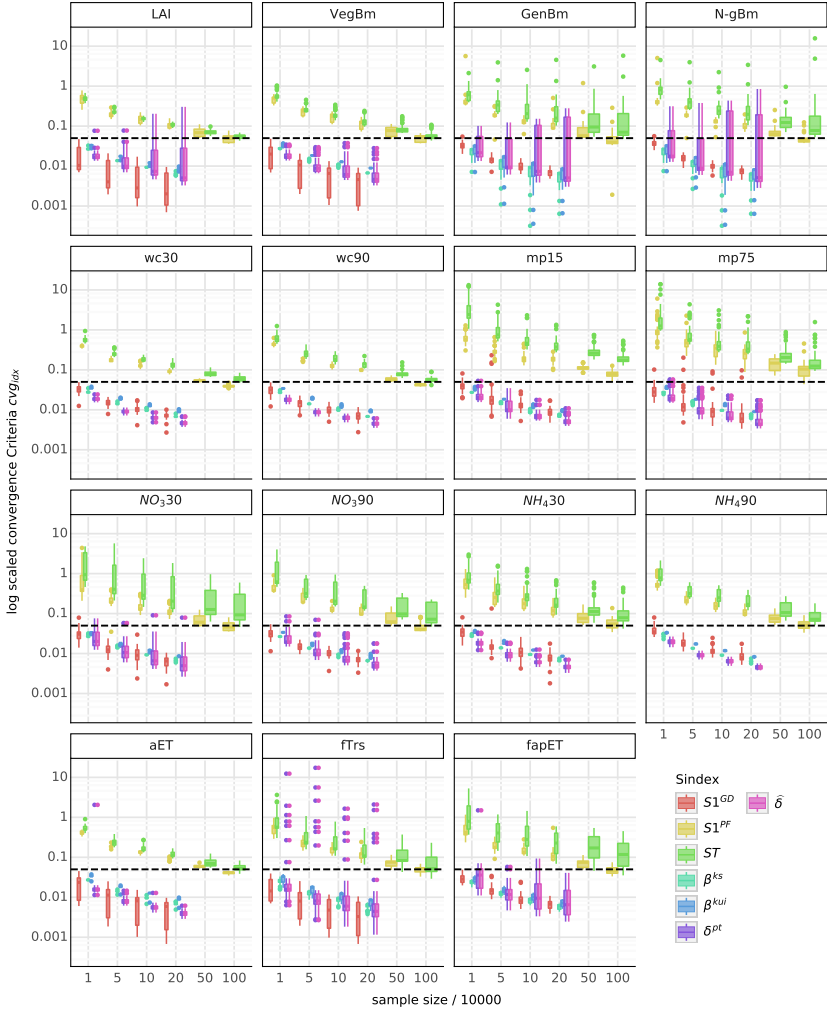


Figure 3.3: Convergence criteria cvg_{idx} for the values of the different sensitivity indices (colors) and the 15 target variables evaluated at different sample sizes (divided by 10000). The box-plots represent the variance of cvg_{idx} within the time series. The dotted line marks the threshold of 0.5 for convergence (see eq. 2.27).

TGV. With a sample size of 1000000, the SI^{PF} can almost reach the desired threshold. Furthermore, ST converges more slowly than $S1^{PF}$. In contrast, the SI^{MI} and $S1^{GD}$ undercut the threshold right from the smallest sample size. Common among all SIs, time variability for the TGVs $fTrs$, NO_330 , $genBm$ and $N - gBm$ is higher compared to the other TGVs. Comparing the SI^{GD} , β^{ks} and β^{kui} have the lowest time variability, whereas $S1^{GD}$ and δ alternate with the highest time variability. Finally, the $\hat{\delta}$ show higher time variability for the TGVs that are non-existing or do not vary during the whole simulation period (cf. 3.1). Interestingly, the temporal variability of $GenBm$ and $N - gBm$ increases with sample size, and in the case of δ , it does so to such an extent that it crosses the desired threshold. We would like to note that the variability of $cv_{g_{idx}}$ mostly arises from large confidence intervals at the onset of the yield formation.

In section 2.5.2.3, we additionally defined a second convergence criteria that considers the ranking ($cv_{g_{rk}}$) of the SIs instead of their idxvalue. Figure 3.4 shows the 95-quantile of $cv_{g_{rk}}$ with its standard deviation across the time series and TGVs. For the first time, facets represent the different SIs with the SI^{VB} in the top row and the SI^{MI} in the bottom row. Although the SI^{PF} did not converge in value, the parameter ranks between the different LHS sizes are consistent in that the required threshold is satisfied. Analogously with value convergence, rank convergence improves with increasing sample size for all SIs (except $S1^{GD}$), but with a damping effect. What is neither accounted for by $cv_{g_{rk}}$ nor by $cv_{g_{idx}}$ is the fact that, with an increasing sample size, the number of parameters that become sensitive increases for the SI^{MI} . In contrast to SI^{PF} , where top-ranked parameters change ranks at a lower sample size, highly sensitive parameter ranks are stable for all SI^{MI} .

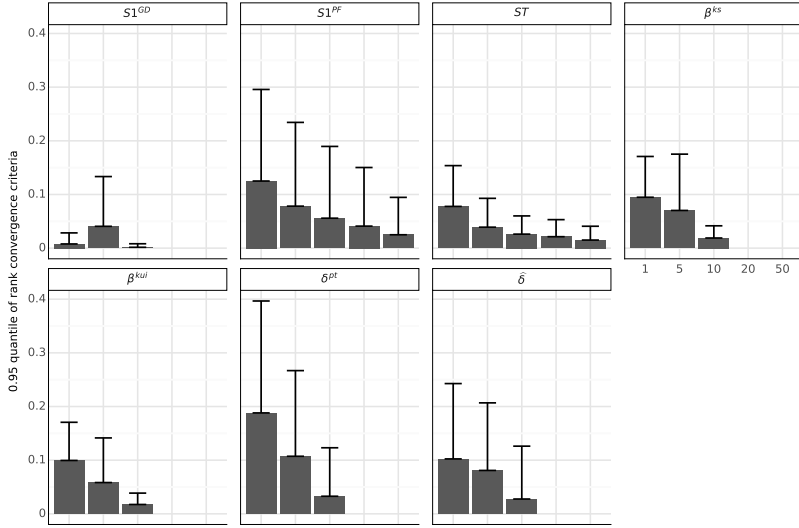


Figure 3.4: The 95-quantile of the convergence criteria for the parameter ranks $cv g_{rk}$ plotted as a function of the sample size. Each facet represents one sensitivity index. Error bars indicate standard deviations of $cv g_{rk}$ for the different target variables and dates.

3.2.2 Computing Resources

This section discusses the necessary computing resources and their use efficiency to conduct a GSA for XN-CERES with the set objectives. As representatives for the computing resources, we use the evaluated amount of data (file size) in GB; the sampling efficiency, which is the ratio of set sample size and actually evaluated simulation runs; average analysis time in minutes which is the time needed to calculate one SI; and the total analysis time in days, which is the time needed to calculate the set of indices for the different TGVs and dates presented here. Figure 3.5 shows the specified criteria for each respective method or SI. Analysis times are given for the SIs, β^{ks} and β^{kwi} (cdf), δ and SI^{PF} (sobol), for the matching of the three matrices (matrix, compare 2.4.3.1), and for the CUSUNORO method. The last item includes

the time needed to construct the CUSUNORO curve and to extract the desired values. For the sampling efficiency, a distinction between VB and MI methods is sufficient where the former refers to SI^{PF} and the latter to SI^{GD} . The file size is independent of the method, and

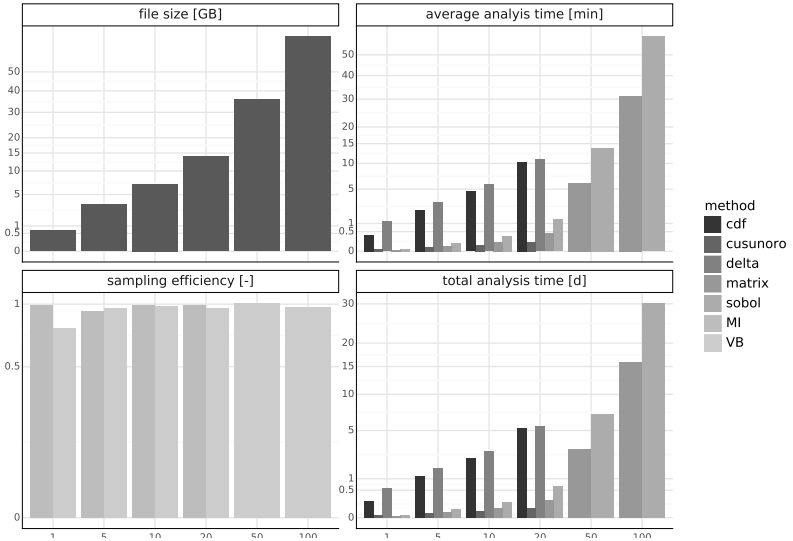


Figure 3.5: Overview of the needed computing resources as a function of sample size for the moment-independent methods (cdf, δ , MI), the variance-based methods (sobol indices, matrix, VB) and constructing the CUSUNORO curve. File size of model output, total analysis time in days, time to calculate on sensitivity index and the sampling efficiency are markers for computing resources and efficiency and are given on the square root scaled y-axis in their units.

therefore no distinction is made. However, each method separately requires the amount of data shown to be processed. Hence, to calculate all SI for a sample size of 200000, 52 GB (4x14 GB) must be evaluated, which would take about 12 days on the bwUniCluster. Considering also the time it takes to run the XN, the whole DPC takes 79 days of computing time. Keep in mind that, for the SI^{PF} , the times of the matrix matching and the index calculation must be summed up. Calculating SI^{PF} is much cheaper regarding analysis time, but slow

value convergence requires higher sample sizes than the MI methods. Therefore, this advantage is canceled out. Furthermore, the sampling efficiency (70%-86%) is less in comparison to MI methods (97%) due to the model crashes and the loss of samples resulting from the matrix matching. The XN-CERES crashes if the numerics become insoluble, e.g., in the case of divisions by 0 or mass balance errors due to low infiltration capacity of the soil in the model.

3.2.3 Index Comparison

Objective 2.2 is to assess how appropriate the different GSA methods are to identify sensitive parameters for the Factor Fixing and Factor Prioritization settings and how they compare. Therefore, the rank correlation (cf. 2.5.2.2) of the different SIs is given in figure 3.6. The rank correlation is written in the fields; the darker the red, the lower the rank correlation. Therefore, red squares directly indicate disagreement between SIs, whereas white squares indicate agreement. From figure 3.6, six key questions about the methods and the model can be answered by visual inspection.

1. How do the SIs estimated from GD and the P&F compare?
2. Is our model affected by interaction effects?
3. Is the model outputs' variance an appropriate measure of uncertainty?
4. Is there a difference in measuring the sensitivity with the Kolmogorov-Smirnov or the Kuiper metrics?
5. Does the kind of bias correction of δ lead to different parameter rankings?
6. Does it make a difference if the pdf or the cdf is considered to represent the model's output distributions?

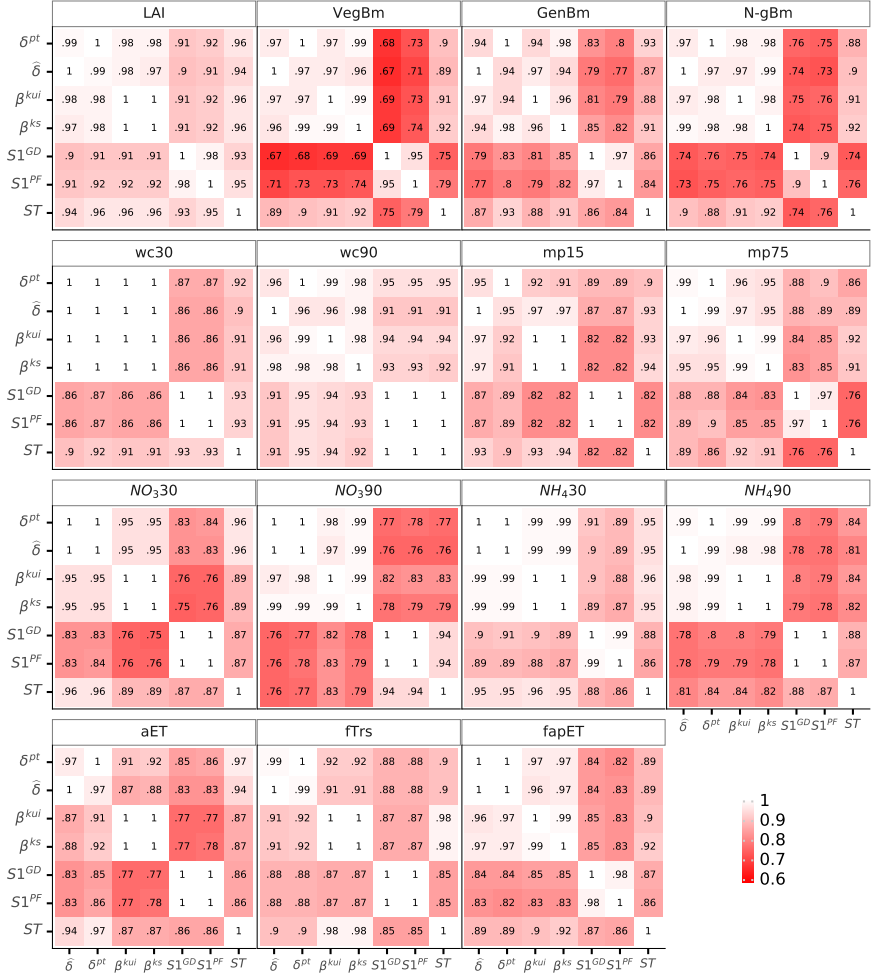


Figure 3.6: The rank correlations for all sensitivity indice pairs are shown separately for each target variable and at highest method specific sample size (1000000, 200000).

The first question can be answered by inspecting the four-field domain spanned by SI^{GD} and SI^{PF} in the lower right corner. For most of the TGVs, it is dyed white, which means perfect rank correlation. However,

there is a disagreement in all crop group TGVs, although to different extent. For *mp75* and *fapET*, slight differences between the two S1s can be observed.

The second question can be inferred in two ways. A first indicator is the rank correlation of the S1s and ST, which is given in the fields A5-6, B7 (A refers to columns and B to rows). An interaction is present if the rank correlation is low, i.e., the defined squares are red. Otherwise, the parameter's impact is only direct, and the idxvalues of S1s and ST are equal. As one can see, interactions are present in each TGV except *wc90*. However, the degree of interaction is different. It is highest for *N - gBm*, *VegBm* and *mp75*, and lowest for *wc30*, *NO₃90* and *LAI*, whereas the remaining TGVs are in between but below 0.9. A second indicator for the presence of interactions in the model can, but does not have to be, the correlation of S1s and SI^{MI} . Lower correlations may, but do not have to, indicate that the model is not only controlled by first-order effects. It could also mean that the variance is not an appropriate representation for the UCSR. The answer to question three explains how to identify if low correlations result from interactions or are due to an incorrect representation of the UCSR by variance.

To answer this question, it is valid to consider the rank correlations of SI^{VB} and SI^{MI} filling the fields A5-7, B5-7. Three cases can be distinguished. In the simplest case, all twelve fields are red, which means that the variance is not an appropriate representation of the UCSR. This is the case for all TGVs. In the second case, the correlation between ST and SI^{MI} is higher than the one between S1s and SI^{MI} , which is a clear indicator that there are parameters that only have an interaction effect and are hence only detected by either ST or SI^{MI} . This is not the case for *NO₃90*, *wc30* and *wc90* and *fapET*, which means that here the difference in the ranking is only caused by the different model output representations. In the last case, the differences between SI^{MI} and SI^{VB} are caused by both interaction effects and the different model output representations. This is especially the case if the correlation between SI^{MI} and S1s is considerably different from

the one between ST and S1s. This is the case for $mp75$, NO_330 and aET . In summary, we can say that all TGVs from the crop group are affected more by interaction effects than by different representations of the model output. Soil water contents and NO_390 are hardly affected by interaction effects but by the different representations of the model output. All other TGVs are affected by both characteristics.

The fourth question is answered by focusing on the fields A3-4, B3-4, which are consistently white or close to white. Hence, it does not matter if the Kuiper or the Kolmogorov-Smirnov metric is used to measure the distance between the conditional and unconditional cdfs.

The choice of the bias correction method for δ can affect the resulting parameter ranking. This can be inferred from the correlations between $\hat{\delta}$ and δ^{pt} covering A1-2, B1-2. The fifth question can be answered in the affirmative, although the differences are low. It is worth noting that the choice of numerical noise filtering mostly affects the number of parameters identified as sensitive at the edge of detection. Hence, the critical value is more restrictive in its threshold function than the filtering scheme (cf. fig. 2.9). However, it is too restrictive for smaller sample sizes.

Question six can also be affirmed because the correlations of cdf- based and pdf-based measures are distinctive, although to a different extent. Therefore, we keep one SI from each model output representation. Given the almost perfect correlation of β^{ks} and β^{kui} , we henceforth consider β^{ks} . We chose δ^{pt} due to its better restriction properties at higher sample sizes. Furthermore, their convergence behavior is similar.

3.3 GSA-Application

In this section, the results regarding Objective 1 GSA-application are presented. First in section 3.3.1, the additivity of the XN-CERES model is presented. In section 3.3.2, the parameter sensitivities over time and the time-independent parameter ranking are shown. They are divided into the four target groups, crop, water, nitrogen, and flux.

Section 3.3.3 covers trend identification for selected parameters. Based on the results from section 3.2, only parameter sensitivities and ranks for β^{ks} , δ^{pt} , and SI^{PF} are considered in section 3.3.2. For the model additivity, we chose $S1^{GD}$ because its value converged (cf. section 3.3), and the drift in parameter ranking between $S1^{GD}$ and $S1^{PF}$ is negligible.

3.3.1 Model Additivity

The additivity of the sub-models is shown in figures 3.7 to 3.10, each concerning the target group specific TGVs. In all figures, the $idxvalue$ of $S1^{GD}$ is plotted against the month of cultivation period. Axes are free, and therefore the scales should be observed. Each combination of color, line type and marker represents one parameter. Parameters associated with the same target group have colors from the same color map, where green is the crop group, blue the water group, purple the flux group, red the nitrogen group and gray the ini group. Therefore, it can directly be seen if there are model interrelations. Furthermore, the $idxvalues$ are stacked, which means that one can directly read the additivity of the soil-crop model at the upper most line. Additionally, the marker size and the size of the spanned area for each parameter give the direct contribution of this specific parameter to the TGV output's variance. Finally, the number of parameters that do have a direct impact on the TGV can be inferred by counting the parameters listed in the legend. For the model additivity, we chose $S1^{GD}$ because its $idxvalue$ converged (cf. fig. 3.3). Within each paragraph, we first present commonalities and differences across the specific group TGVs and then proceed with the individual TGV.

Crop First, it is notable that most of the direct impact on the crop group TGVs is associated with crop group parameters. Second, the crop model for the target variables referring to the vegetative crop organs is near additive, with 80% to 100% of direct explainable variance

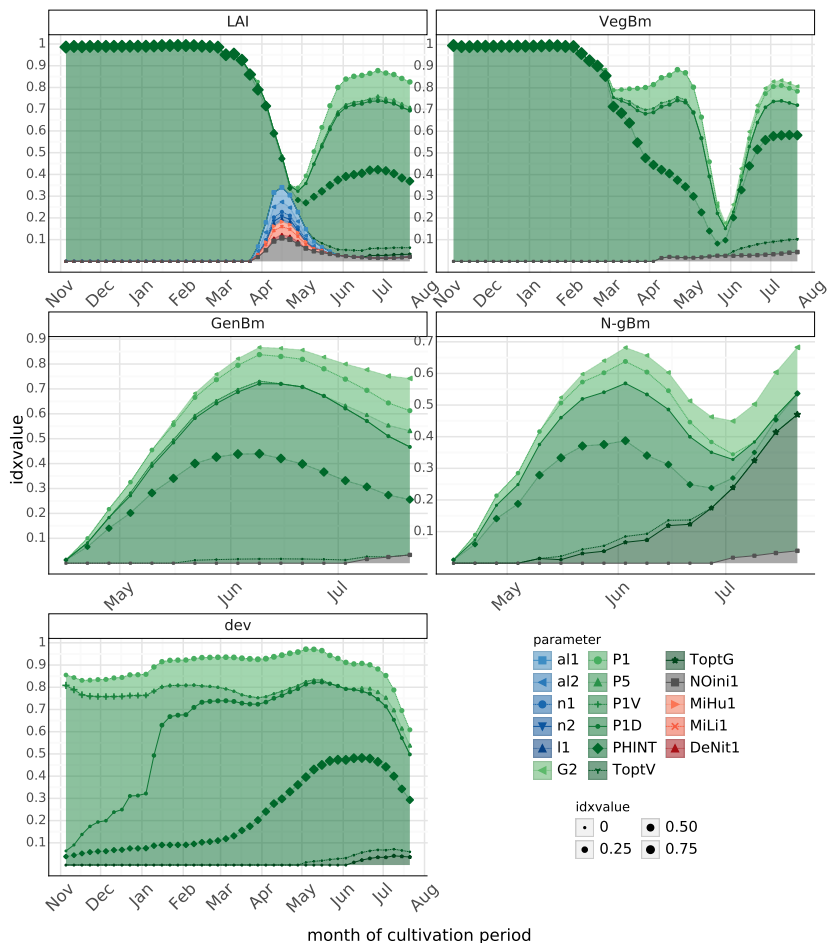


Figure 3.7: Stacked $S1^{GD}$ values for the target variables of the crop group plotted against the dates of the cultivation period and the value for $S1^{GD}$ (idxvalue). Each color, marker and line type combination represents one parameter, and colors are target group specific (crop: green, water: blue, nitrogen: red, flux: purple, ini: gray). Area and marker size reflect the importance of each individual parameter.

of the simulation results (VSR). Third, $mxWup$, $mxNup$, $G2$ and $P4$ do not have any direct impact on any TGV of the plant group.

For *LAI* and *VegBm*, the parameter *PHINT* is the most important parameter until March; it explains 100% of the TGVs' output variance. Around April or the end of May, respectively, the model additivity suddenly drops to only 20%-30%, followed by a steep increase. At the same time, both TGVs start to depend directly on *NOini1*. For *LAI*, mineralization rates and Mualem van Genuchten (MvG) parameters emerge shortly. Afterward, in addition to *PHINT*, *P1D* and *P1* become important, whereas *P1V*, *ToptG*, *ToptV* and *G2* play only a minor role after the drop. For *VegBm*, one observes a first drop in March, where *PHINT* is continuously replaced by *P1D* and *P1*, which then make up 10%-40% of the direct explainable variance in *VegBm*'s model output. The model additivity for TGVs referring to the generative crop organs increases slowly starting in mid-April, which coincides with the moment the model additivity of *VegBm* drops for the first time. Again, *PHINT* and *P1D* are the most important parameters. For *GenBm*, the model reaches a maximum additivity of about 90% in mid-June. After the peak, *PHINT* and *P1* lose their dominant role, and the parameters *G2*, *P1*, *P5* and *NOini1* co-determine *GenBm*'s output variance. The nitrogen content in the generative biomass is increasingly dominated by *ToptG*, explaining up to 50% of *N - gBm*'s model output variance. The duration of stadium 1, i.e., (*P1*) and *P1D*, are not unimportant, nor is the maximum grain growth rate *G2* and *NOini1*. In total, a model additivity of only 70% is reached. Finally, the model additivity for the crop development stages is constantly high, with only less than 90% at the start and end of the cultivation period. *P1V* causes most of *dev*'s variance in the beginning but is relieved mainly by *P1D*, *P1* and *PHINT*.

Water The MvG parameters are the most important parameters for the TGVs of the water group. Furthermore, *kcmid* has a high, direct impact during the vegetation period on both the matric potential and the soil water content. In total, 19 parameters from four groups have an impact on the TGVs of the water group. Regarding the time course

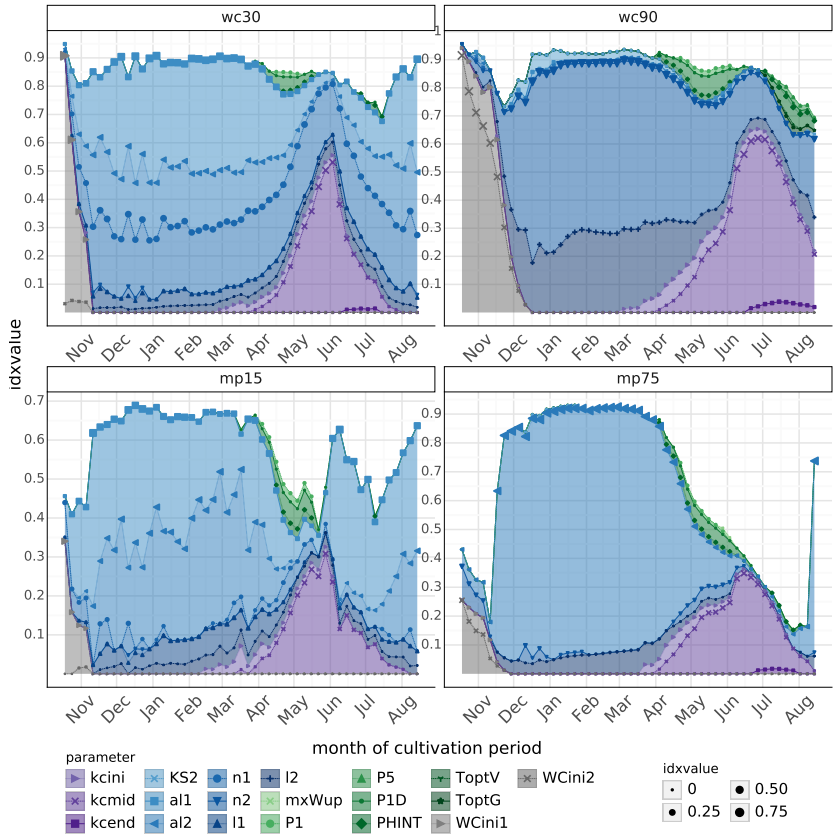


Figure 3.8: Stacked $S1^{GD}$ values for target variables of the water group. For labels, see figure 3.7.

of the model additivity, matric potential and soil water content differ. The model for the soil water content is constantly has a constantly high model additivity with 85%-95% of direct explainable VSR. Initial soil water content is the major source of uncertainty for slightly more than one month. During winter, this role is resumed by $al1$, $al2$ and $n1$ for the upper soil and $l2$ and $n2$ for the sub soil. During the vegetation period, $kcmid$ explains up to 60% of the variance. Crop group parameters affect $wc30$ most during April, whereas $wc90$ is more affected in

July.

The matric potential is similar in winter, although the model additivity for $mp15$ never reaches more than 0.75. Especially during the vegetation period, the matric potential is predominantly controlled by interaction effects. The matric potential is also impacted more by $al1$, $al2$, $l1$ and $l2$ than by $n1$ and $n2$.

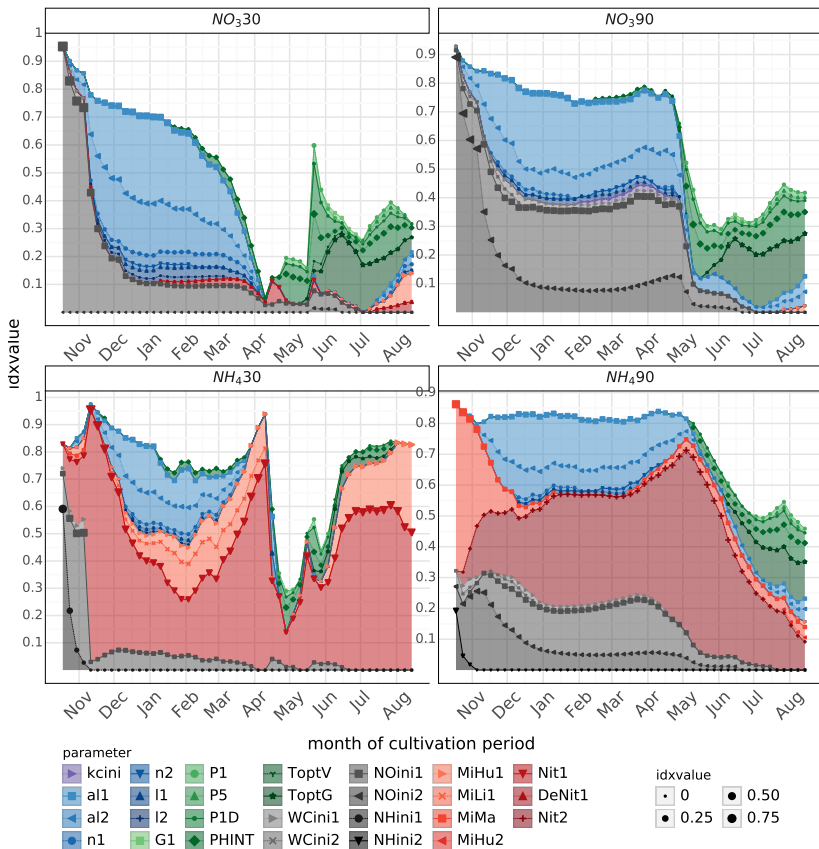


Figure 3.9: Stacked $S1^{GD}$ values for the target variables of the nitrogen group. For labels, see figure 3.7.

Nitrogen In total, 27 parameters from all groups have a direct impact on the N content in the soil. All TGVs of the nitrogen group have in common that, until April, both N-forms are directly affected by *NOini1*, *al1*, *al2*, *n1* and *l2*, which are then replaced by crop parameters, mainly by *ToptG*. Furthermore, the direct impact of the initial N content is reduced to a minimum until mid-May. The flux parameter *kcini* impacts the soil N content.

Comparing the NO_3^- contents with the NH_4^+ contents independent of the soil depth, one can conclude that the NH_4^+ content mostly depends on nitrification rates, whereas the NO_3^- content is more dependent on the MvG parameters. Nonetheless, MvG parameters cannot be neglected for the NH_4^+ content. Crop parameters are much more important for the NO_3^- content from May onwards. On average, they explain 35% of the VSR. In contrast, regarding the NH_4^+ content, crop parameters explain only 5% with a maximum of about 20% of the VSR. The NH_4^+ contents mainly depend on mineralization rates and *NHini1* and *NHini2*, respectively. The NH_4^+ and NO_3^- content in the subsoil have in common that the idxvalues of $S1^{PF}$ constantly add up to about 75% until May. Afterwards, both show a decrease in the sum of their idxvalues. In the case of NH_490 , it is slow but constant, and in case of NO_390 it is more rapid. Both TGVs are not only impacted by *NOini1*, but also by *NOini2*.

For NO_330 , we directly observe a decrease in the model additivity, which reaches its minimum in April. During this time, about 60% of the direct explainable VSR is caused by the MvG parameters, mainly *al*. The initial NO_3^- content reduces its contribution to the VSR from 100% to 10% within two months. Furthermore, we can identify small increases in the model additivity at the time of fertilization, for NO_330 especially on the last fertilization date. In contrast, for NH_430 , the first and second fertilization dates have a larger impact on the model additivity and the parameter sensitivities. The increase of the model additivity is also in line with the dynamics of the model output (cf. figure 3.1). For NH_430 , the moment *Nit1*'s importance decreases again

coincides with the dates of fertilization. Mineralization rates gain importance after fertilization.

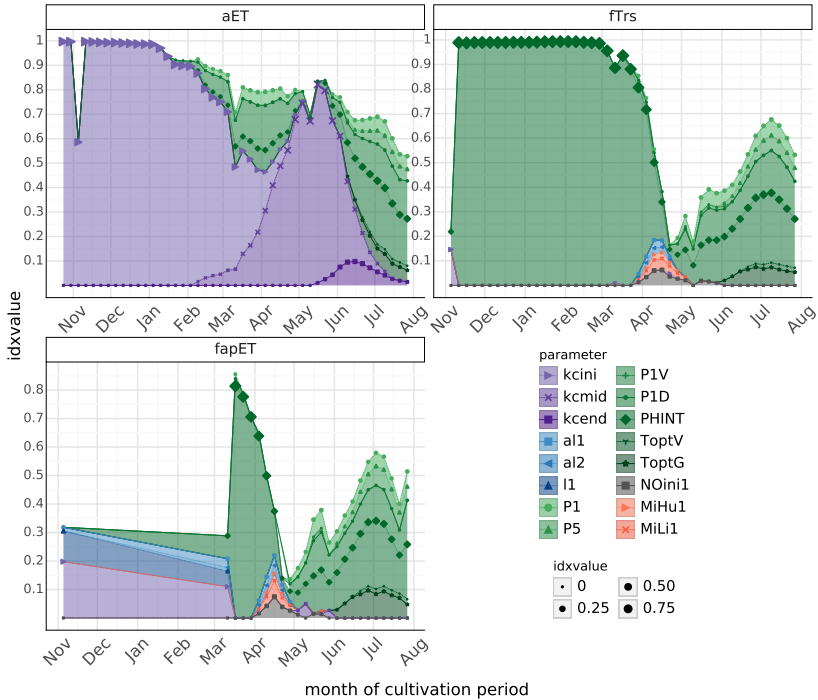


Figure 3.10: Stacked $S1^{sat}$ values for target variables of the flux group. For labels, see figure 3.7.

Flux The model additivity for the TGVs of the flux group is shown in figure 3.10. Since there are few similarities, the individual TGVs are discussed individually. The model for the daily actual evapotranspiration aET starts as perfect additive. Here, $kcini$ is the dominant parameter. As soon as the vegetation period starts, crop parameters become important, and the model additivity decreases to 0.5. $kcini$ is slowly replaced by $kcmid$, whereas at the end of the cultivation period, crop parameters have in particular a direct effect on aET .

For the ratio of transpiration to evaporation, $fTrs$, the model is perfectly additive and dominated by *PHINT* until March. Furthermore, $fTrs$ is impacted by the same parameters as the TGV *LAI*. Their course of the models' additivity and the important parameters are comparable (cf. figure 3.7).

Last, the ratio of actual to potential ET ($fapET$) shows the most diverse behavior. Disregarding the peak in March, the direct explainable VSR hardly reaches 50%. Between November and March, no parameter shows a direct impact (no markers means no values). Figure 3.1 shows that there is simply no variation in the model output during this time. Furthermore, the course of the model additivity reflects the width of the 50% and 95% prediction interval. Crop parameters and *kcmid* exhibit the highest impact on the ratio of actual to potential ET in the model, whereas the parameters from the other groups play a minor, temporary role around April.

3.3.2 Parameter Sensitivities

The topic of this section is the parameter sensitivities and the parameter ranking for the selected SIs (cf. 3.3). Section 3.3.2.1 shows the results for the crop group, sections 3.3.2.2 and 3.3.2.3 for the water and nitrogen groups respectively, and section 3.3.2.4 for the flux group. Within each group, the time-independent parameter ranking for each TGV is shown as well as the time series of the parameter sensitivities. We start with the former because it provides a more general impression, and we use the latter to relate individual findings. Since the figures are organized by a generic pattern structure, a full description of the time-independent figures and the time-dependent figures is found once at figure 3.11 and 3.12. Although some of the information in the two figure types is repetitive, the different visualizations help make different aspects easier to perceive.

The discussion of the figures is oriented towards a generic pattern. We elaborate commonalities and differences between the individual TGVs

before we present the TGV specific peculiarities in detail. Furthermore, we label each distinct statement with a number at the beginning.

3.3.2.1 Crop Group

Time-independent Parameter Ranking

Figure 3.11 shows the time-independent parameter ranks of the averaged and weighted idxvalues (cf. eq. 2.26), faceted by the group specific TGVs. The chosen time interval d denotes the whole simulation period.

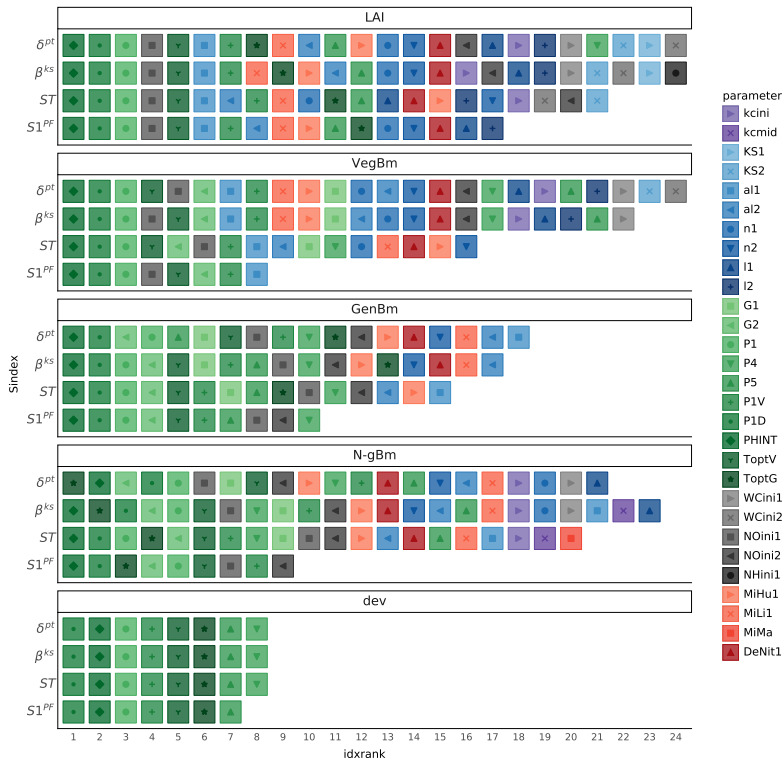


Figure 3.11: Parameter ranks of the weighted and averaged sensitivity index values for the target variables of the crop group and for the four different sensitivity indices. Markers and colors represent the parameters.

Unique combinations of colors and markers represent the parameters and are the same as in section 3.3.1. Ranks are plotted against the four SIs. Again, colors make the parameter group that has the major impact, or play a minor role for the specific TGV, directly visible. Furthermore, it can be directly compared whether the individual SIs result in the same parameter ranking and, if not, what the differences are.

The color-coding shows that (1) all SIs indicate that TGVs of the crop group are dominated by parameters from the crop group, and (2) that the initial NO_3^- content in the top soil is the first foreign group parameter. (3) The TGVs are impacted by up to 24 parameters. Furthermore, we can see that, in general, (4) MI methods identify more parameters as sensitive. Although these parameters are always ranked lowest (except $P4$ and $WCini2$ for $VegBm$ and LAI), this is an indication that (5) the UCSR can not be fully represented by the variance. This means that ST and SI^{MI} agree on the parameters that are most important within the common set of ranked parameters, although the order within the common set can be different.

However, (6) for $VegBm$, $GenBm$ and $N - gBm$, about half of the ranked parameters do not have a direct impact on the UCSR. Furthermore, (7) for the target variables referring to the vegetative crop organs (TGV_{veg}^{pl} s) parameters from the water and nitrogen group are ranked higher, whereas the generative crop parts are more sensitive to further crop group parameters. In addition, (8) al is the most important MvG parameter for TGV_{veg}^{pl} s independent of soil depth. Last, (9) the MvG parameters for the sub soil are ranked higher for the target variables referring to the generative crop organs (TGV_{gen}^{pl} s). Except for LAI , parameters from the water and nitrogen group impact the crop model only through interactions. This confirms the findings from figure 3.6: The crop group TGVs are not only controlled by direct impacts. We note that, for the TGV of the crop group, the disagreement between the $S1^{PF}$ and $S1^{GD}$ results either from different ranking order among the lowest-ranked parameters or from additionally identified parameters at the lower end of the ranking by one of the two SIs.

Individual TGVs are discussed in association with the time-dependent figures 3.12 for the *LAI*, figure 3.13 for *VegBm*, 3.14 for *GegBm*, 3.15 for *N - gBm* and 3.16 for *dev* in the following paragraph.

Time-dependent Parameter Sensitivities

Figure 3.12 shows the time series of the parameter sensitivities for *LAI*. The facets contain the parameters, which are identified as sensitive by at least two of the four selected SIs or by ST. The number next to the parameter name in the facet reflects the average rank and standard deviation taken over the four SIs. Further, the average rank deter-

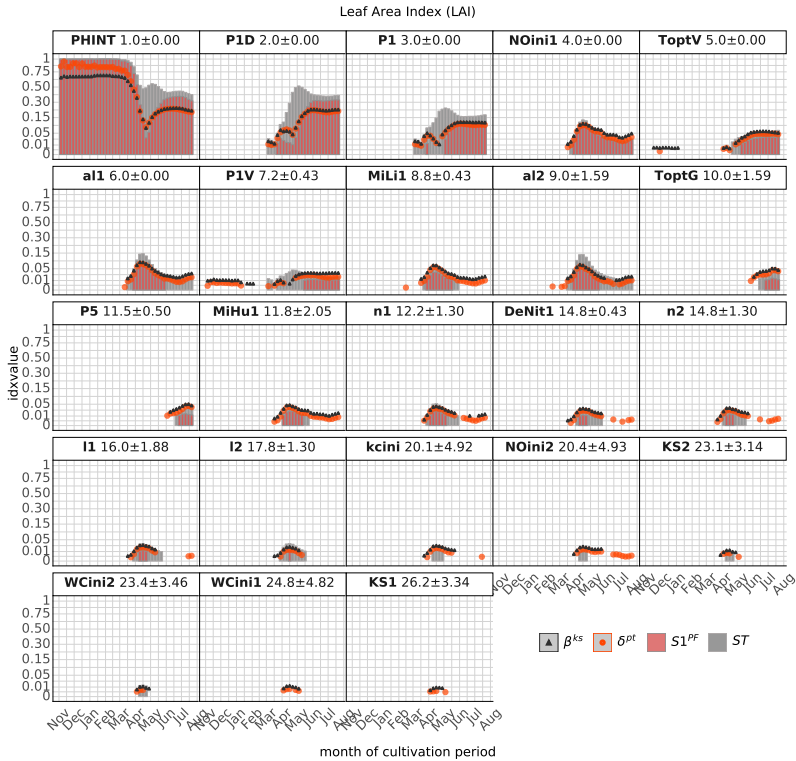


Figure 3.12: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable leaf are index (*LAI*).

mines the order of appearance. The idxvalues are plotted against the dates. The y-axis is square root transformed. The idxvalues shown correspond to the weighted SIs, with a time window of six days (cf. eq. 2.26). The colors represent the four SIs, i.e., ST, $S1^{PF}$, δ^{pt} and β^{ks} . The title indicates the corresponding TGV. It can be seen that (1) until April, the whole UCSR of LAI is caused by $PHINT$, and (2) according to the SI^{MI} , also slightly by $P1V$ and $ToptV$. (3) The drop in the model additivity in April (cf. fig. 3.7) arises, according to SI^{VB} , from the parameter $PHINT$ interacting with other parameters. (4) However, SI^{MI} keep following the time course of $S1^{PF}$. (5) At the same time, parameters from other groups, such as MvGs parameters, mineralization and denitrification rates, as well as the $NOini1$, become important. Interestingly, this happens when the simulated LAI starts to decrease (cf. 3.1). Two further pieces of information are noteworthy. Lower-ranked parameters often coincide with short periods of sensitivity. SI^{MI} , especially δ^{pt} , identifies parameter as sensitive for longer time periods than SI^{VB} .

In 3.13, the parameter sensitivities of $VegBm$ are shown. Until February, the parameter sensitivities for $VegBm$ are comparable to those for LAI . When the dominance of $PHINT$ starts to vanish, one can observe that, besides $P1D$, $P1$ and $NOini1$, al , mineralization and denitrification rates turn from insensitive to sensitive. This is also comparable to LAI , even if the impact is not direct. During the decrease in model additivity around May (cf. 3.3.1), the parameters $PHINT$, $P1D$, and $P1$ remain the most influential, but their contribution to the VSR turns from direct to interacting. Again, this happens when the simulated $VegBm$ is no longer increasing (cf. fig. 3.1). Likewise LAI , SI^{MI} tend to follow the time course of the $S1^{PF}$ and not of ST. This means that a drop in the direct impact of a parameter leads to a smaller idxvalue of β^{ks} and $\hat{\delta}$. Regardless, ST and SI^{MI} agree quite well on parameters with no first-order effect, although they are identified earlier by SI^{MI} . Still, SI^{MI} include parameters in the ranking from April to June that are not sensitive according to SI^{VB} .

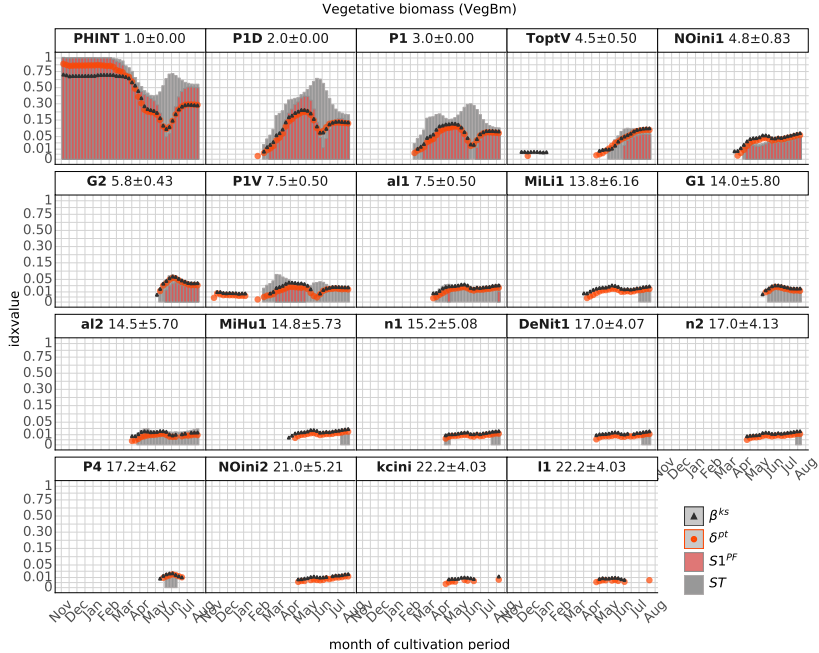


Figure 3.13: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable vegetative biomass ($VegBm$).

The two lowest-ranked parameters by the SI^{MI} (cf. fig. 3.11) are not shown because they are only an "insular-phenomenon" during the simulation period, with $idxvalue$ smaller than 0.01.

Figures 3.14 and 3.15 show the time series of the parameter sensitivities for $GenBM$ and $N - gBM$. First, the commonalities of the two TGVs are discussed simultaneously. Afterwards, TGV specific findings are presented separately. As seen in figure 3.7, (1) in the beginning of the yield formation, the VSR is mainly caused by parameters interacting with each other. (2) Even the determinability of the key parameters $PHINT$, $P1$ and $P1D$ is limited here due to the poor ratio of $S1^{PF}$ to ST . (3) Over time, the two SI^{PF} approach in value. (4) However, the SI^{MI} for these parameters can be discerned. Whereas β^{ks} mostly follows the course of $S1^{PF}$, δ^{pt} also shows high sensitivities to these pa-

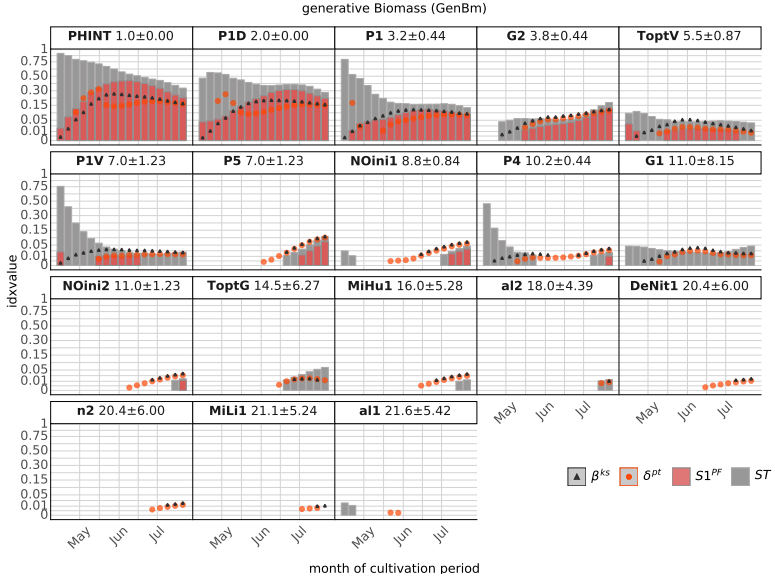


Figure 3.14: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable generative biomass ($GenBm$).

rameters in the beginning of yield formation. (5) In contrast, the two SI^{MI} correspond well for all other parameters, whereby their influence increases slowly. (6) $P1V$ and $G1$ only impact the two TGVs by interactions. (7) Parameters from the nitrogen and ini groups emerge later in the yield formation process, and the SI^{MI} identify these parameters earlier than SI^{PF} . (8) Later in the season, SI^{MI} additionally tend to identify more parameters as sensitive, but mostly with idxvalue values below 0.01 and infrequent occurrence.

Regarding the differences between the two TGV_{gen}^{pl} s, first, $N - gBm$ depends on more parameters according to SI^{MI} . Second, $ToptG$ is much more important for its UCSR, and the ratio of $S1^{GD}$ and ST for the parameters $PHINT$, $P1$ and $P1D$ is again increasing, i.e., worsen towards the end of the vegetation period and is never above 0.5. Hence, their determinability is questionable. Finally, $P4$ has its only direct impact on $GenBm$, but only once over the whole simulation period. The

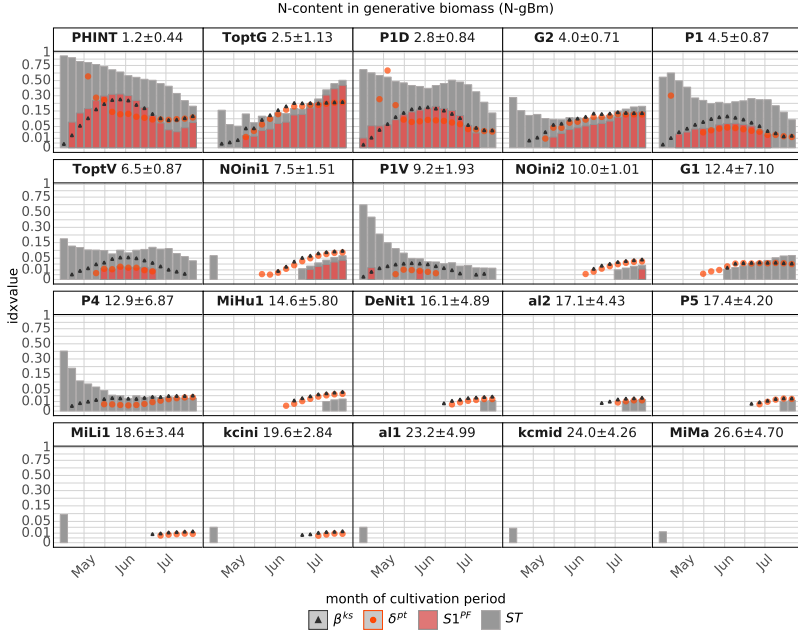


Figure 3.15: Time series of the sensitivity (idxvalue) for the sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable nitrogen content in the generative biomass ($N - gBm$).

parameter sensitivities for dev are shown in figure 3.16. As was seen before, its UCSR is mainly affected by first-order effects and only by eight parameters. Since dev is used as a filter in the DPC (cf. 2.5.2), this is not surprising. The missing 5% to 10% of direct explainable VSRs at the cultivation period start originate from interaction effects (cf. 3.7). Due to the high ratio of $S1^{PF}$ to ST in this case, it is negligible in regard to parameter determinability. As was observed before, SI^{MI} follow the course of $S1^{PF}$. $P4$ is the parameter that explains the lower model additivity at the cultivation end and is the only parameter with no direct impact on UCSR. However, MI methods rate it as more influential than VB methods.

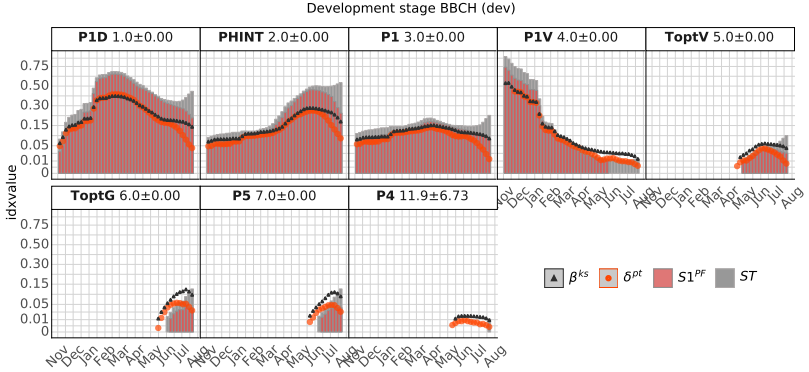


Figure 3.16: Time series of the sensitivity ($idxvalue$) for the sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable development stage ($dev.$)

Summary

First, we have seen that the UCSR of TGVs of the crop group are dominated by parameters from their own group with $PHINT$, $P1$ and $P1D$ comprehensively as the most influential parameters. Except for LAI , $NOini1$ is the only foreign group parameter that has a first-order effect on crop group TGVs - $S1$ comprehensively. Second, SI^{MI} follows the course of the first-order effect, if present, for the TGV_{veg}^{pl} , whereas for TGV_{gen}^{pl} δ^{pt} is comparable to ST in the beginning. Third, both methods agree on the parameter ranking of the most influential parameters as well as on a common set. Fourth, we have seen that, in general, SI^{MI} and ST can identify parameters that have no first-order effect. The SI^{MI} seem to be more sensitive for low influential parameters. Fifth, $P4$ is a parameter with nearly no direct impact on any TGV of the crop group.

3.3.2.2 Water Group

Time-independent Parameter Ranking

Figure 3.17 shows the weighted, averaged parameter ranking for the TGVs *wc30*, *wc90*, *mp15* and *mp75*. In total, 37 parameters have an influence on the soil water regime, whereby 19 of them do not have a first-order effect. All crop parameters except *mxNup*, all water parameters, all initial conditions, *kcini*, *kcmid* and *kcend* as well as nitrification, denitrification and mineralization rates are sensitive. Only

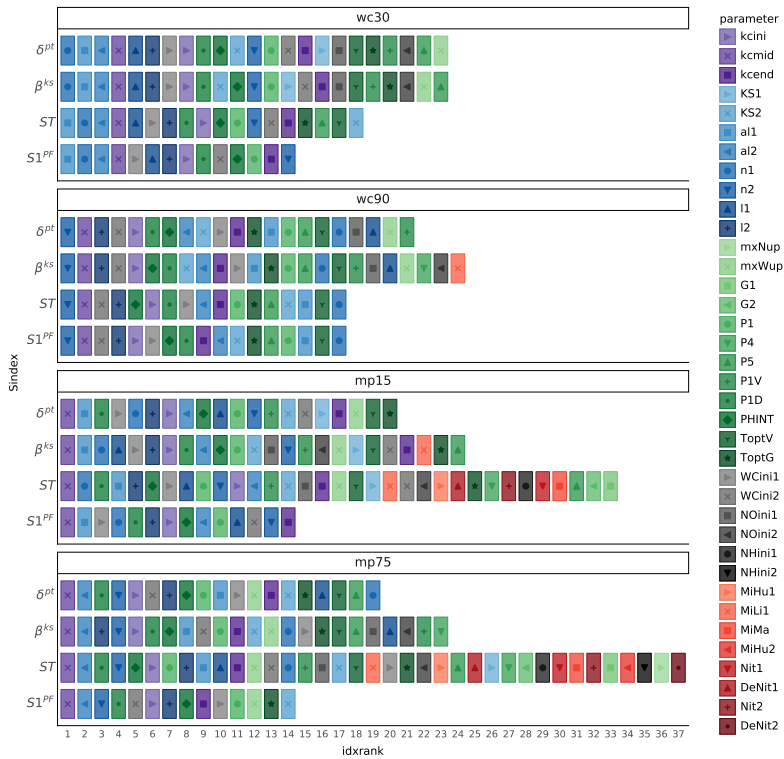


Figure 3.17: Parameter ranks of the weighted and averaged sensitivity index values for the target variables of the water group for the four different sensitivity indices. Markers and colors represent the parameters.

MiLi2 seems to be unimportant for the soil water regime. However, this statement is quickly relativized because nitrification and denitrification rates and the initial NH_4^+ content are only ranked by ST and not by any other SI. In contrast, for the soil water content, SI^{MI} reveal the most parameters as sensitive.

For *wc30*, (1) the nine highest-ranked parameters form a common set across SIs. However, (2) ST prioritizes more parameters from the crop group, whereas the SI^{MI} prioritize *l2*, *KS2* and *KS1*. Hence, there is a disagreement between methods if the water content in the sub soil or the crop growth has a larger impact on *wc30*. (3) Furthermore, compared to the other three TGVs of the water group, *wc30* depends more strongly on parameters from its own group, whereas crop parameters are ranked lower. (4) The agreement between SIs for *wc90* is quite good, as seen in 3.6 and the ranked parameters form a common set. (5) Again, SI^{MI} ranks *KS1* higher than the other methods. (5) Furthermore, *wc90* depends more on *n2* and *l2* than on *al*, which is the most important parameter for all other TGVs of the water group.

The picture for the matric potential looks completely different. (1) By far, ST identifies the most parameters as sensitive. (2) Additionally, the order of the ranks between $S1^{PF}$ and ST does not originate from a common set. Hence, parameters with a first-order effect are ranked higher in ST's ranking. This indicates that, for these parameters, the $S1^{PF}$ to ST ratio is poor. For example, this is the case for *PHINT*, *P1D* and *P1*. However, if a parameter appears much lower in the ST ranking than in the $S1^{PF}$ ranking, there must be parameters that only have a total effect that is even higher than the first-order effect of other parameters. This is the case, for example, for *al1* and *kcmid*. (3) MI methods are in between, although for the matric potential, β^{ks} identifies more parameters than the δ^{pt} . They compare better with the color pattern of $S1^{PF}$ for *mp15*. For *mp75*, the opposite is true. (4) Additionally, the two SI^{MI} show a difference in the order of the parameter ranks.

Time-dependent Parameter Sensitivities

The time series of the parameter sensitivities for *wc30* in figure 3.18 shows five interesting features. First, the decrease in importance of the MvG parameters, especially *al*, in May is mostly compensated by *kcmid* and slightly by *P1D*, *PHINT* and *P1*. Second, the only month where interactions are considerable is July (cf. fig. 3.8). At the same time, we can observe that crop group parameters become sensitive or re-increase their sensitivity. Furthermore, the impact is, or becomes to, interacting. Nevertheless, the individual idxvalues are small. Third, the initial water content's impact only lasts for about four to six weeks. Fourth, the SI^{MI} follow the time course of $S1^{PF}$, as seen before in figure 3.13 for *VegBm*. Fifth, *KS1* as well as *KS2* have, surprisingly, no impact on *wc30*'s variance but on its distribution in the whole simulation period. Finally, we note that the additional parameters identifies as sensitive by the SI^{MI} all appear between April and June but with only small idxvalues.

The temporally resolved parameter sensitivities for *wc90* are in some relations similar those of *wc30* and are presented in figure 3.19. (1) SI^{MI} follow the time track of $S1^{PF}$. (2) The decrease in importance of the MvG parameters coincide, on the one hand, with the onset of the same crop parameter sensitivities. On the other hand, *kcmid* takes over as the most important parameter. Thus, they have a comparable quantitative and qualitative effect on the simulated *wc90* and on the simulated *wc30*. (3) Regarding timing, the first two described patterns for *wc30* are shifted to June. (4) In contrast to *wc30*, the crop parameters retain, or show, a first-order effect. (4) Furthermore, the initial water content impacts *wc90* for eight to ten weeks, and (5) *KS2* is now declared sensitive by all SI. Three features can be observed for the first time from the sequence of the parameter sensitivities for *mp15* (figure 3.20). (1) As was seen in figure 3.17, *mp15* is highly affected by parameter interactions over the whole simulation period. This time, even the most influential parameters *n1*, *al1*, *kcmid*, *l2*, *l1* and *al2* most of the time have a greater total effect than a first-order effect, which

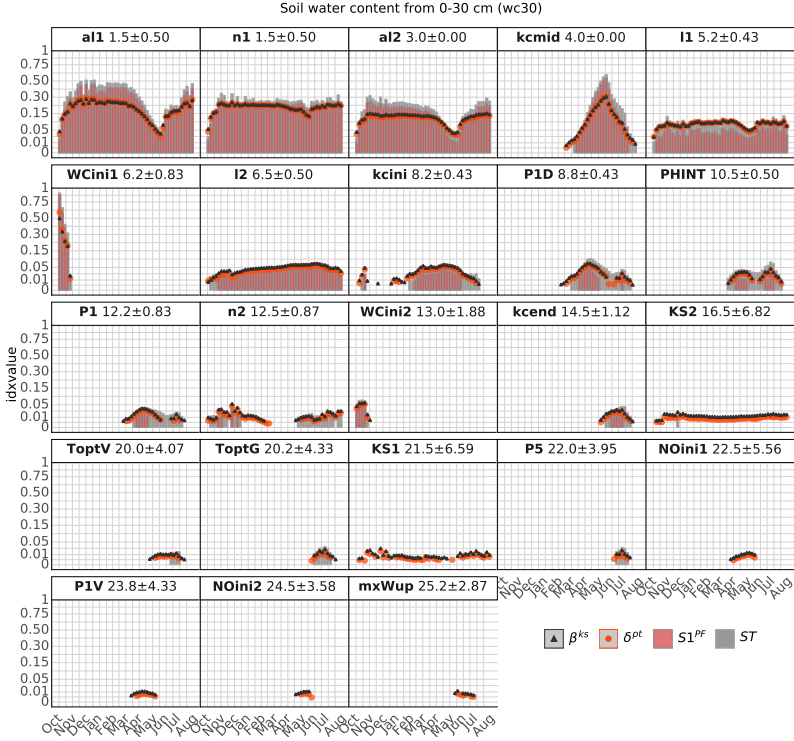


Figure 3.18: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable soil water content in the top soil ($wc30$).

is apparent from the ratio of $S1^{PF}$ to ST and mostly < 0.5 . (2) This time, SI^{VB} s identify nine more parameters as sensitive compared to the SI^{MI} . The four hidden parameters have only once an impact and an $idxvalue$ slightly above 0.01. (3) This time, β^{ks} is more similar to the chronological course of ST , whereas δ^{pt} follows the track of $S1^{PF}$. We would like to note three more aspects. (1) By taking the example of $mp15$, we would like to draw attention to the effect of weighting SIs. At the first glance, it might be striking that, for example, $kcmid$ and $P1D$ are high-ranked parameters, although their impact is clearly time-restricted. However, their highest sensitivity coincides with the time

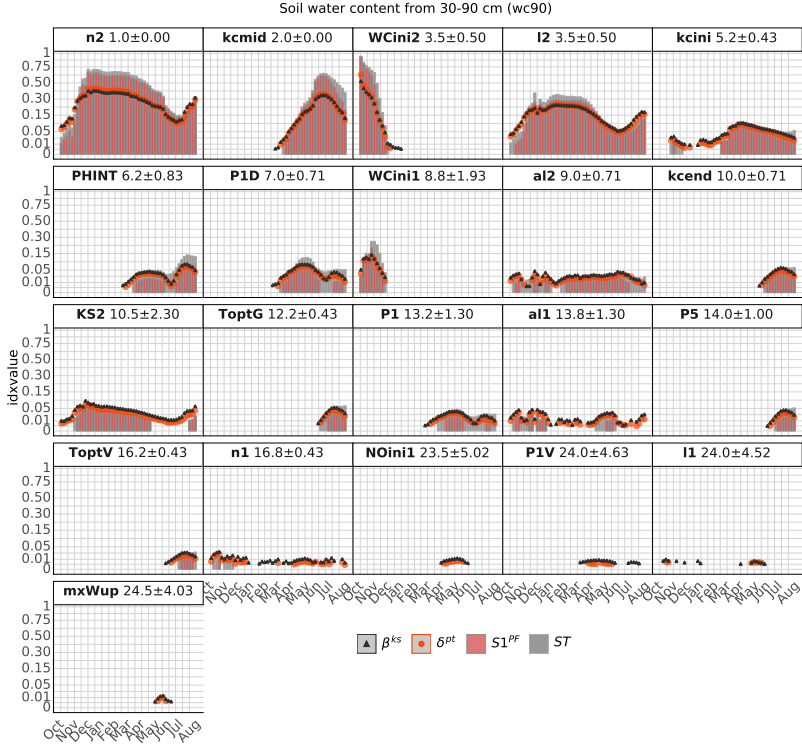


Figure 3.19: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable soil water content in the sub soil $wc90$.

of the highest UCSR in the model output (cf. fig. 3.1), and therefore, they are much more important for the uncertainty of $mp15$. (2) The initial water content again loses its impact within four weeks. (3) Low-ranked parameters and parameters that are only sensitive according to ST occur rarely and have low $idxvalue$. The chronological sequence of the parameter sensitivities for $mp75$ is given in figure 3.21. This time, we start with the low-ranked parameters at the bottom of the figure. (1) From April to August, VSR of $mp75$ is affected by 14 parameters only according to ST , whereby only twelve are shown in order to save space. The missing two parameters are only sensitive at one evaluated

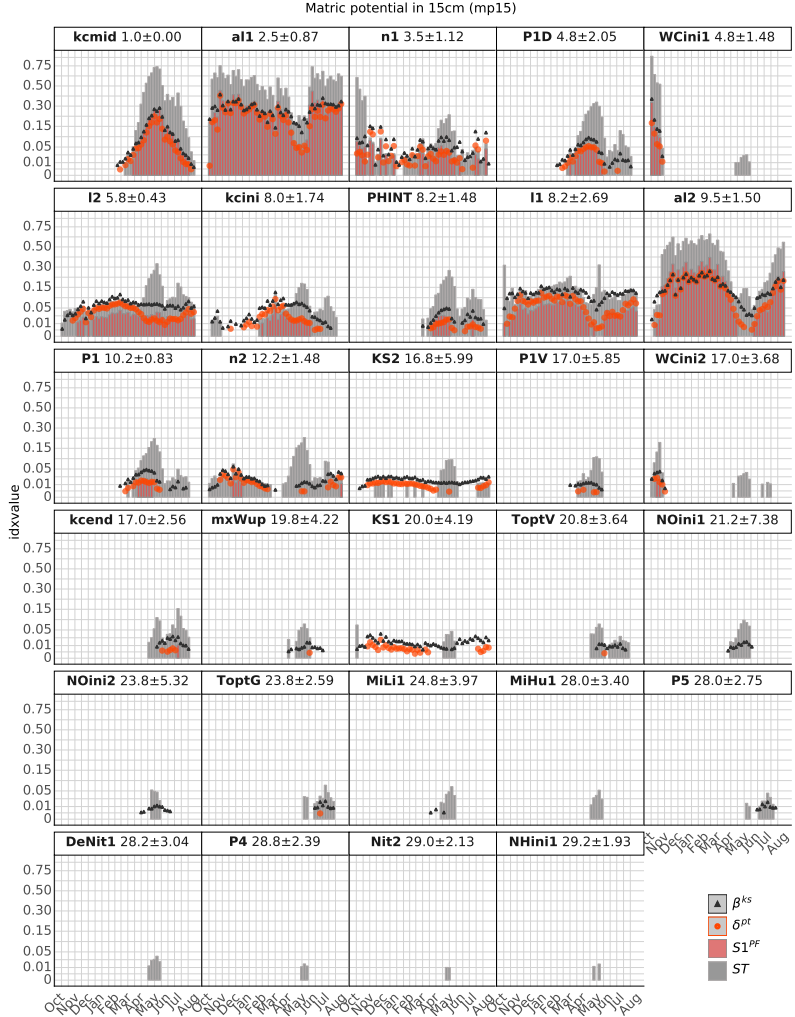


Figure 3.20: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable matric potential in 15 cm depth ($mp15$).

date, and their idxvalue is at the lower sensitivity threshold. The ST 's idxvalue of some of these parameters can make up to 15% of the model output's variance, which is in contrast to $mp15$. The rest of the param-

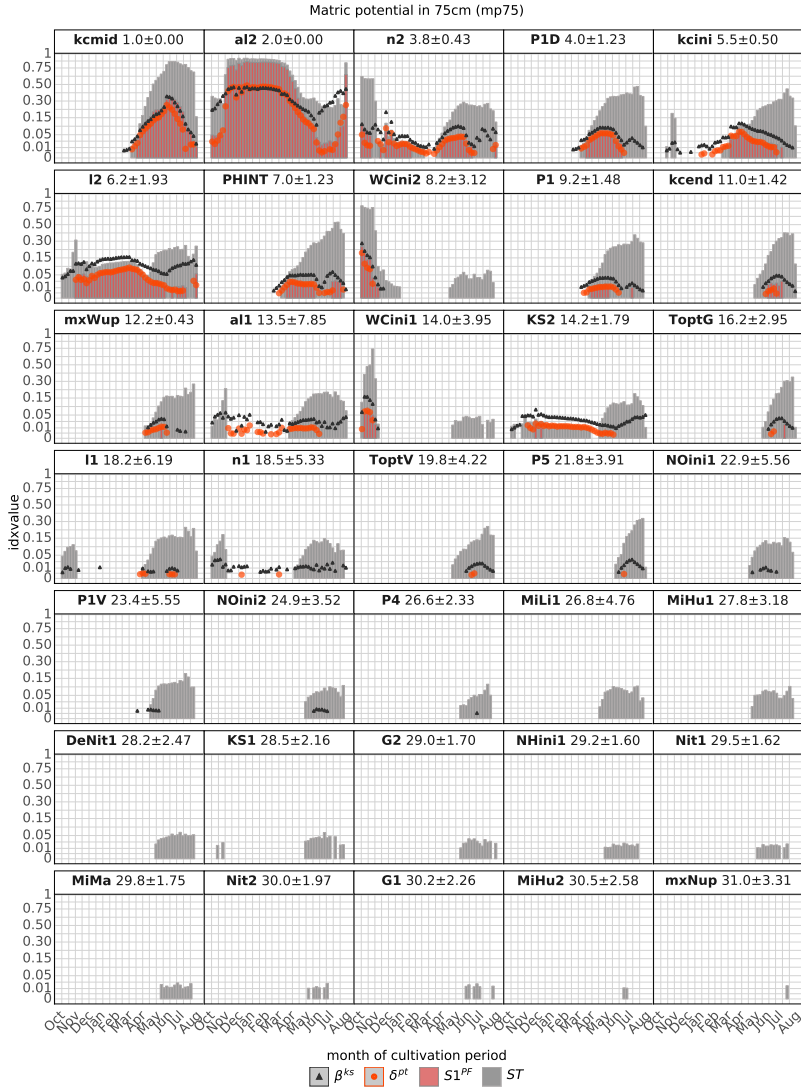


Figure 3.21: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable matric potential in 75 cm depth ($mp75$).

eter sensitivities is similar to *mp15*. (1) Most parameters have a poor $S1^{PF}$ to ST ratio, (2) the initial water content loses its impact within four to eight weeks and (3) *kcmid* is the most influential parameter, even though its sensitivity is temporary. Its increase in importance coincides with the sensitivity decline of *n2* and *al2*. (4) Although β^{ks} is again more in line with ST and its time course, it is still conservative with regard to the parameters that start in row five.

Summary We observed that, for the soil water content, SI^{MI} identify more parameters as sensitive with a low degree of influence, whereas for the matric potential, the ST considers up to 14 additional parameters as sensitive with a share of up to 15% of the VSR. Furthermore, the interesting time of the model simulations ranges from April to June, where the number of parameters that have an impact on the TGVs of the water group is highest. Initial water content is perceptible in the model between four to ten weeks and crop parameters thoroughly have a direct impact on the soil water regime as well as the flux group parameters. In particular, the importance of *kcmid* is should be emphasized here. For the matric potential, especially during the period of the highest UCSR, most parameters only have an interaction effect. However, the ratio of $S1^{PF}$ and ST is poor over the whole simulation period, even for the key parameters. The agreement between SIs is very high for the soil water content.

3.3.2.3 Nitrogen Group

Time-independent Parameter Ranking

Figure 3.22 shows the parameter ranks for the TGVs of the nitrogen group. In total, 35 parameters are identified as sensitive by at least one of the four SIs. Only two commonalities among the TGVs of the nitrogen group can be found, 1. *NOini1*, and for the TGV_{90}^N , additionally *NOini2*, as well as *al1* and *al2* are top-ranked parameters. 2. β^{ks} always identifies the most parameters as sensitive, and the number does not significantly change among TGVs of the nitrogen group. For the

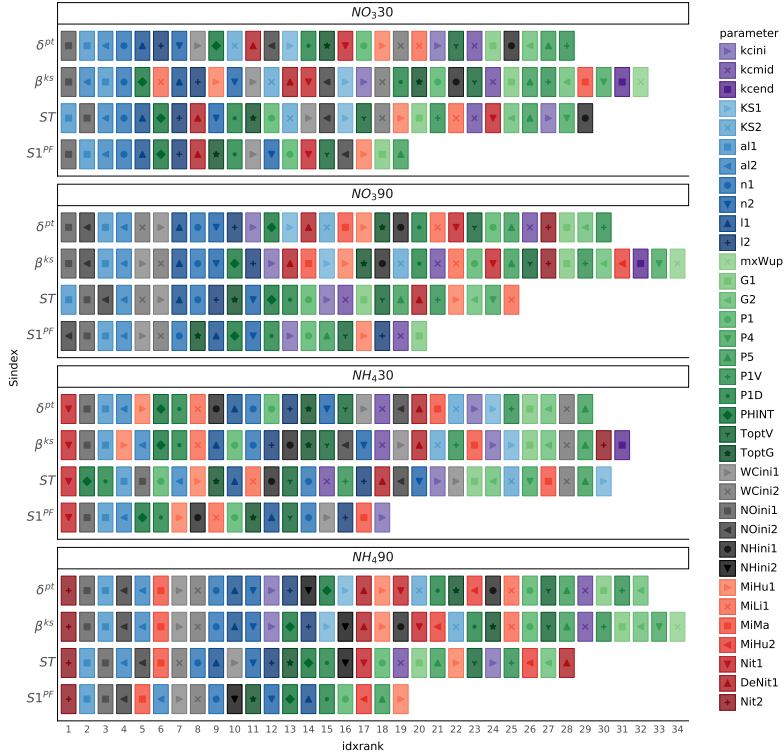


Figure 3.22: Parameter ranks of the weighted and averaged sensitivity index values for the target variables of the nitrogen group for the four different sensitivity indices. Markers and colors represent the parameters.

TGV_{90}^N the initial water content is also listed within the first five to eight ranks. Both TGV^{NH} have in common that the most important parameter is the depth-specific nitrification rate, whereas the $TGV^{NO_3^-}$ are more impacted by $l1$, $l2$, $n1$ and $n2$. Nevertheless, the color pattern for the TGV_{90}^N is better ordered than that of TGV_{30}^N . For the NO_{30} , β^{ks} reaches a different parameter ranking; it classifies mineralization rates as more important than nitrification rates and crop group parameters, which are holistically ranked higher by the SI^{VB} s. At least the first

four ranks include parameters from a common set across SIs. Beyond that, the parameter ranking of the different SIs differs, and the color pattern crop, nitrogen and flux parameters are classified as sensitive. For the NO_3 90, now both SI^{MI} identify nitrification and mineralization rates and $kcini$ as more important than SI^{VB} , which rank crop parameters higher.

The parameter ranking for NH_4 30 shows a diverse pattern. Parameters of all groups except flux are represented in the top ranks. Furthermore, the individual SIs neither necessarily agree on an order nor a common set. This observation is in accordance with the findings in section 3.2.3. The only pattern is that SI^{MI} classify mineralization rates as more important, whereas SI^{VB} s rank crop parameters higher.

The parameter ranking for NH_4 90 is a bit more ordered, where at least the SI^{MI} agree on the parameter ranking. Additionally, the first nine ranks form a common set (except $WCini1$ for ST). Again SI^{MI} assign mineralization rates a higher impact on NH_4 90, whereas SI^{VB} s assign a higher rank to crop parameters. The high rank of $MiMa$ of the lower NH_4^+ soil content is interesting.

Time-dependent Parameter Sensitivities

Figure 3.23 presents the time-dependent parameter sensitivities for NO_3 30. As seen in figure 3.9, (1) in the beginning, the UCSR is mainly controlled by $NOini1$, whereby its decrease in importance coincides with the increase in importance of $al1$ and $al2$, which take over the role as key drivers of the UCSR. (2) Similarly, crop parameters influence the UCSR from February to August, and $PHINT$ is sensitive all the time.

However, what could not be seen before is that (1) all parameters that have a first-order impact also have a poor $S1^{PF}$ to ST ratio (except $NOini1$ in November). (2) Crop parameters ($PHINT$, $P1$, $P1D$, $ToptV$) and mineralization rates show a time dynamic pattern in their sensitivity (3) according to $S1^{PF}$ and the two SI^{MI} . This pattern resembles the time course of the UCSR (cf. fig. 3.1) and increasing

idxvalues coincide with the fertilization dates (cf. tab. 2.2). (3) The main difference in the ranking between the β^{ks} and δ^{pt} originates from the time between March and May, where δ^{pt} rates most parameters as non-influential. (4) Initial water content again becomes influential during May and June, although its impact fades four weeks after the simulation start. The initial NO_3^- content impacts NO_330 in the whole simulation period although its importance decreases.

The results from figures 3.22 and 3.9 are also reflected in the time series of the parameter sensitivities for NO_390 , shown in figure 3.24.

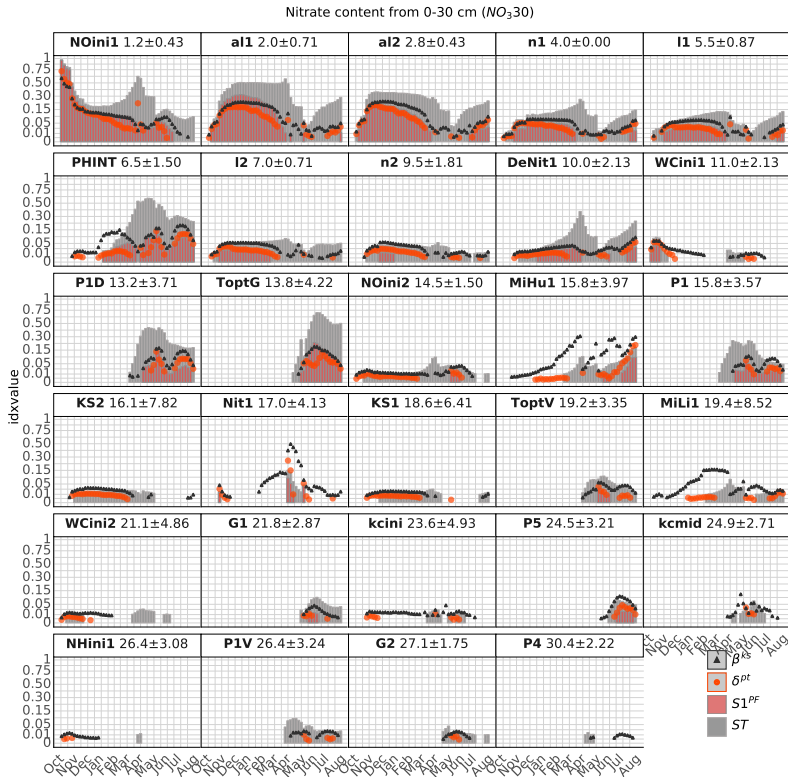


Figure 3.23: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable nitrate content in the top soil (NO_330).

(1) The initial NO_3^- content is, again, the most important parameter according to all SIs and accounts for 38% and 98% of the VSR. In addition, $NOini1$ has a direct impact at all times. (2) As with NO_{30} , according to δ^{pt} , the sensitivity of many parameters is interrupted from April to June. This often coincides with the parameters' loss of the first-order effect. (3) Furthermore, the importance of the ini-

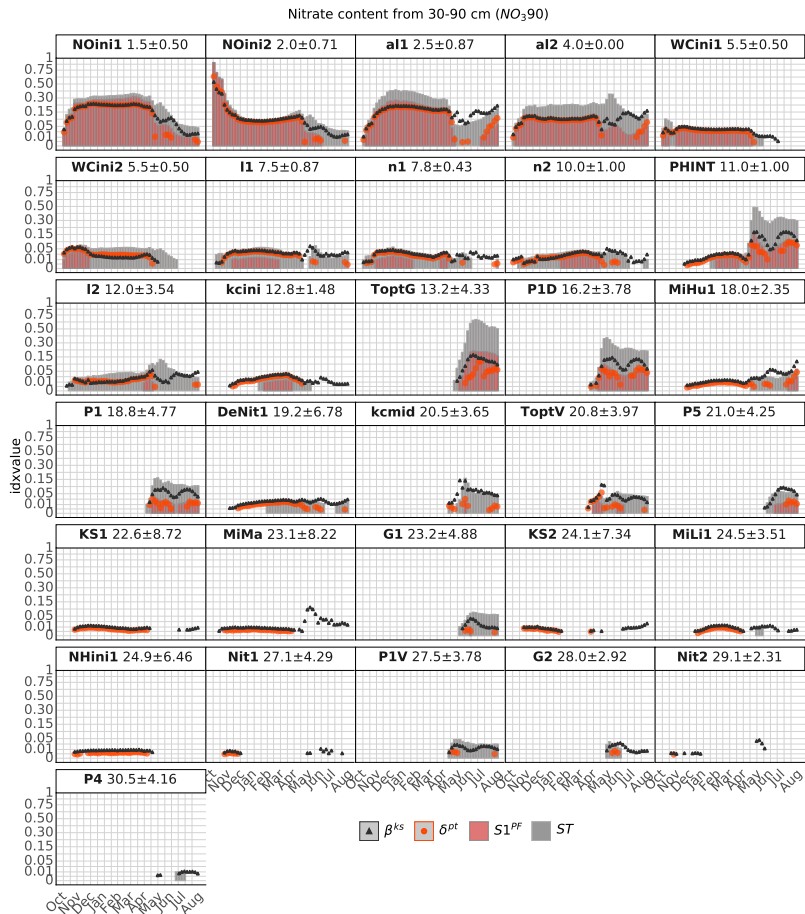


Figure 3.24: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable nitrate content in the subsoil (NO_{390}).

tial water content does not fade after a few weeks but remains constant until the middle of the vegetation period. (4) The sensitivity of the MvG parameters changes from a direct to a total impact. This either means that these parameters totally lose their direct impact or that

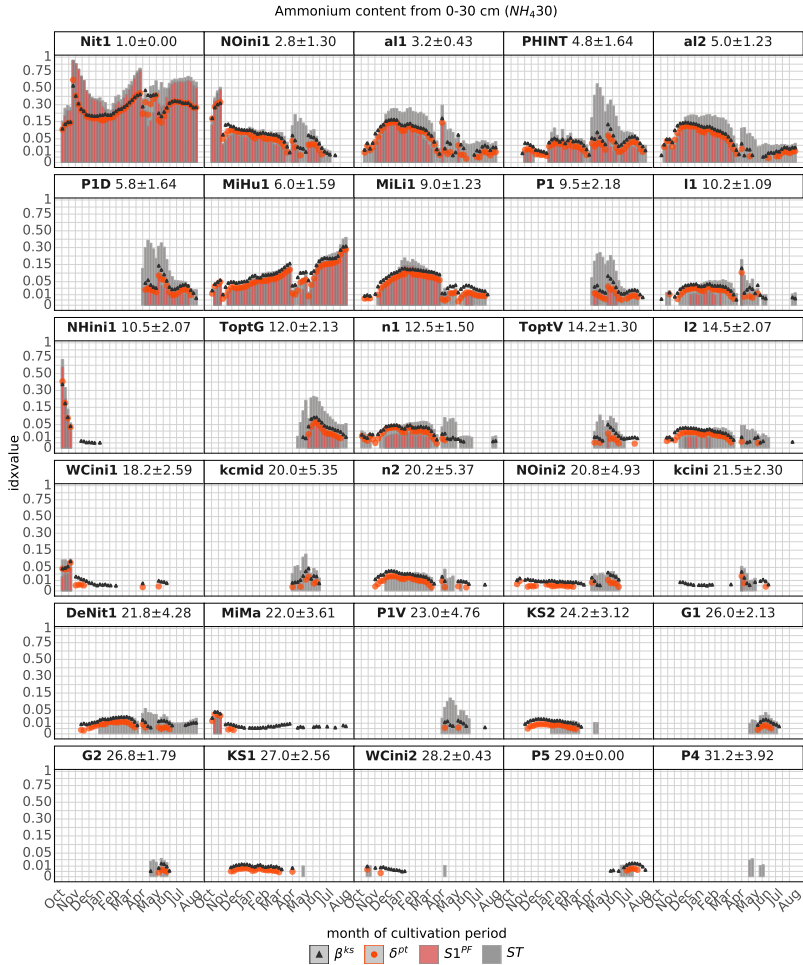


Figure 3.25: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable ammonium content in the top soil (NH_430).

the $S1^{PF}$ to ST ratio decreases to a values below 0.5. This happens in April and May. At the same time, crop, flux, and nitrogen group parameters affect the UCSR. (5) Again, β^{ks} assigns mineralization rates with higher idxvalues.

In figure 3.25, we observe that the time series of the parameter sensitivities for NH_430 correspond to the dynamic evolution of the UCSR.

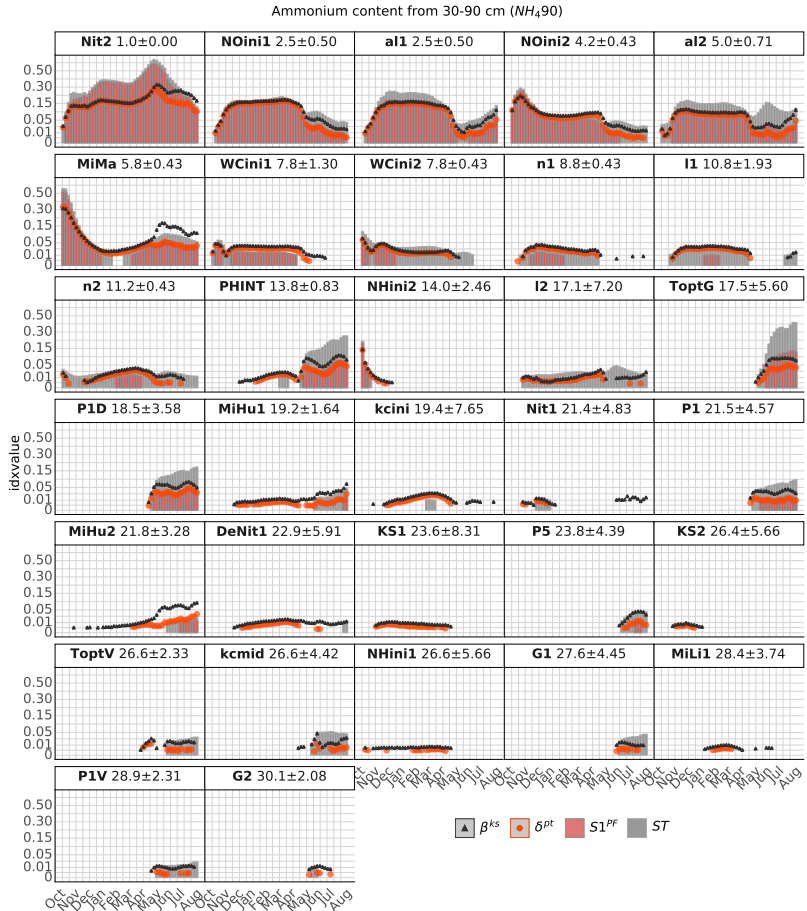


Figure 3.26: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable ammonium content in the sub soil (NH_490).

shown in figure 3.1. (1) First, it appears that the MvG parameters are less important for NH_330 than for the NO_3^- content, and that in terms of importance, they are replaced by nitrification and mineralization rates. (2) Regardless the temporal pattern and the S1 to ST ratio of the crop parameters' idxvalues are comparable, which also holds for the MvG parameters. (3) β^{ks} and δ^{pt} correspond quite well. (4) Initial water content as well as $NHini1$ impact the UCSR for only a few weeks, whereby $NHini1$ is briefly of great importance, and the initial water content only plays a minor role.

(1) The time series of the parameter sensitivities, presented in figure 3.26, clearly shows that the parameters ranked high have a good $S1^{PF}$ to ST ratio most of the time. If not, this situation is temporary and restricted to the time between April and August. (2) In addition, the crop group parameters have a remarkably direct impact on NO_490 . (3) As with NO_390 , the initial water and NO_3^- contents remain influential until the end of the vegetation period. Furthermore, they are among the ten highest-ranked parameters. (4) In contrast, the initial NH_4^+ content fades at the latest by February. (5) Furthermore, β^{ks} and δ^{pt} tend to identify parameters as earlier sensitive than SI^{VB_s} . (6) $MiMa$ is, for the first time, an important parameter.

3.3.2.4 Flux Group

Time-independent Parameter Ranking

Figure 3.27 shows that (1) the crop parameters are the most important parameters for the TGVs of the flux group. (2) *PHINT*, *P1D*, *P1* and *ToptG* are, as for the crop group, the most important crop parameters, and (3) *P4* at no point in time has a first-order effect. (4) For the TGV *aET*, *kcmid* is the top-ranked parameter. (5) MvG parameters, the initial water content and *NOini1* follow in the ranking, whereas nitrogen parameters are only sensitive according to β^{ks} . (6) In general, each SI

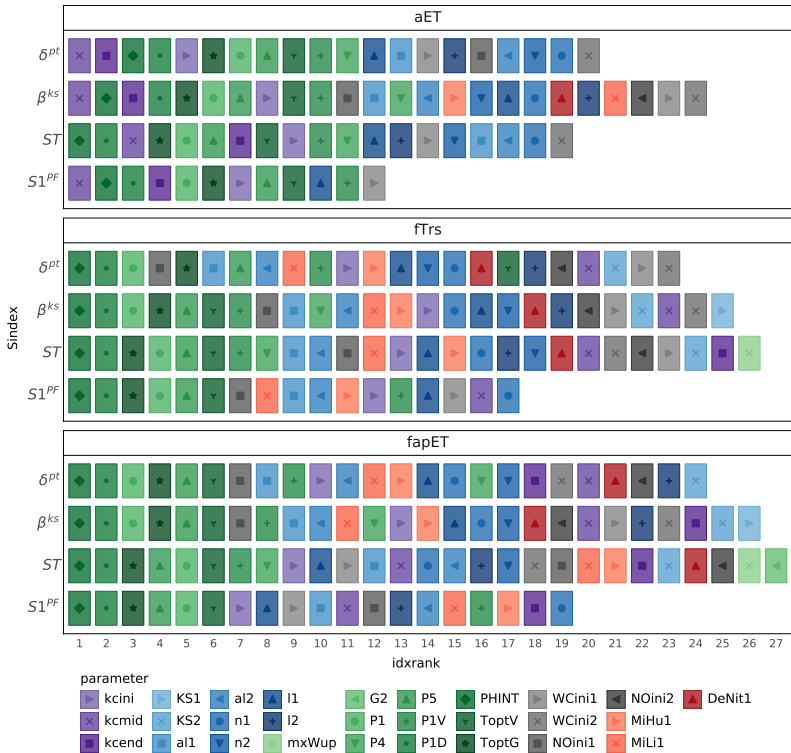


Figure 3.27: Ranking of the target variables of the flux group for the four different sensitivity indices.

yields a different parameter ranking order. (7) For the TGV_{frc}^{fx} , i.e., $fTrs$ and $fapET$, flux parameters are "randomly" scattered between MvG parameters and mineralization rates in the parameter ranking of the four SIs. (8) The most important MvG parameter is again al . (9) Likewise the matric potential ST produces more sensitive parameters than SI^{MI} . Since the picture of the TGV_{frc}^{fx} is similar, we only present time-resolved parameter sensitivities for $fTrs$. The parameter sensitivities for $fapET$ can be found in the appendix.

Time-dependent Parameter Sensitivities

In figure 3.28, the time-dependent parameter sensitivities for aET are presented. (1) $kcini$ is the most influential parameter, and it explains between 75% and 100% of the model output variance during half of the simulation period. However, it does not appear within the first

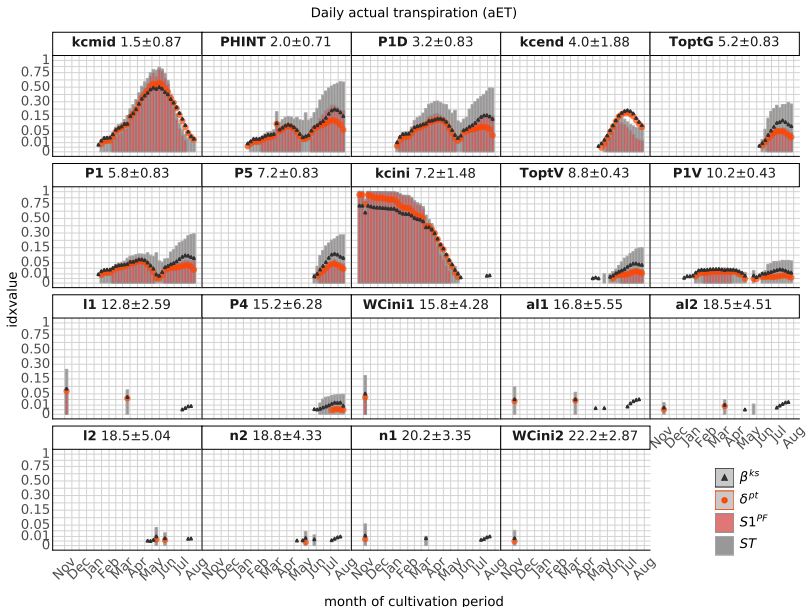


Figure 3.28: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the target variable daily actual Evapotranspiration (aET).

three ranks in figure 3.27. This is an example of how important the weighting of the aggregated idxvalues is. Indeed, $kcini$ is the most important parameter in the winter but the UCSR is nearly 0 (cf. 3.1).
 (2) Furthermore, most of the time crop parameters have a poor $S1^{PF}$

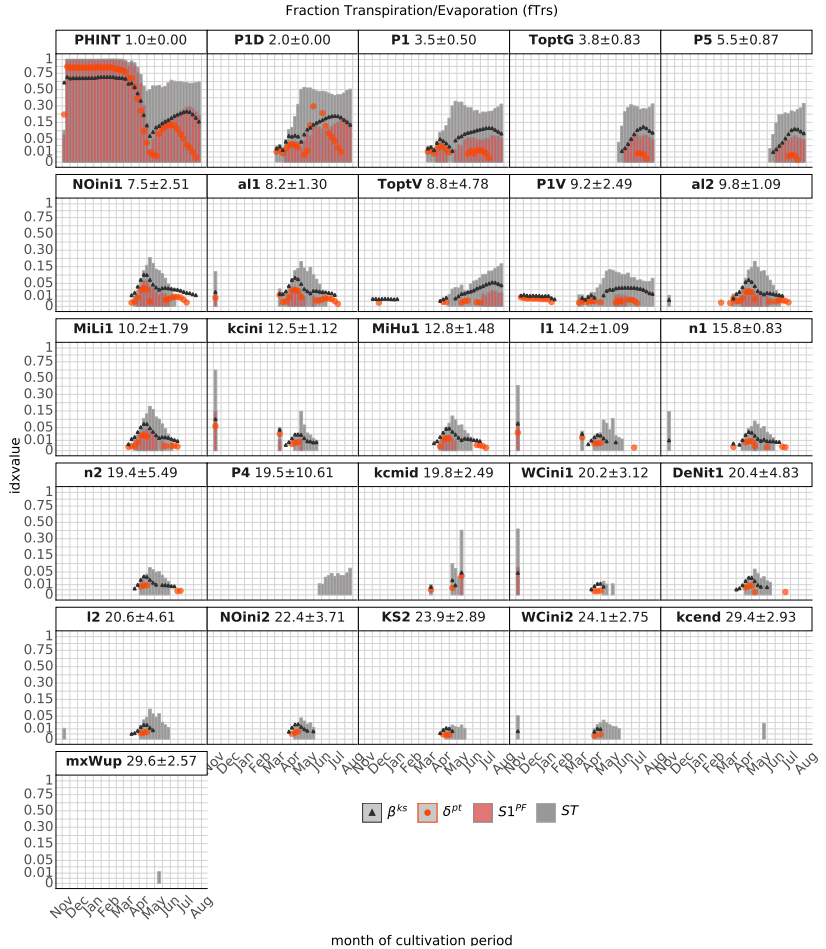


Figure 3.29: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the ratio of daily actual transpiration to daily actual evaporation ($fTrs$).

to ST ratio, with (3) the SI^{MI} again following the time course of $S1^{PF}$. (4) The result that aET depends on 12 to 24 parameters can be deduced from the parameters from row three on. These parameters only appear on the fringes once or twice during the simulation period and only have a small idxvalue.

In comparison with the time series of the parameter sensitivities for aET with $fTrs$ (figure 3.29), it becomes clear that, again, most of the sensitive, low-ranked parameters are influential in the short term. Crop parameters have a poor $S1^{PF}$ to ST ratio. Nevertheless, the accordance between SIs is volatile; in particular δ^{pt} is different from the other SIs. Its parameter sensitivities are less constant over time and are not comparable to any other SI's time course. However, the sensitivities and the high idxvalue for ST coincide with the time periods in which the variation in the simulated $fTrs$ is almost 0.

3.3.3 Trend Identification

This section concerns trend identification, i.e., the linearity and the direction of the parameter's impact. Instead of showing all the results again, we select only the remarkable ones. The results for the individual TGVs can be found in the Appendix (cf. chapter VII). Table 3.2 shows the linearity factor 1_n for parameters where this factor is uniformly different from 0.5 within each group. The parameter name, the average 1_n and its standard deviation are given. Furthermore, we provide the parameter value (x_{1n}) at position 1_n of its range. The standard deviation for x_{1n} is given, as well. The values shown are only for parameters with an $S1^{GD}$ idxvalue higher 0.05.

For *VegBm* and *GenBm*, the time series of the 1_n , the ω and $S1^{GD}$ are shown in figure 3.30 for the parameters *P1*, *P1D* and *PHINT*. The red line indicates the point of 0.5, which is 1_n 's value at perfect linearity. For *GenBm*, these parameters all show the same pattern. First, they all have a decreasing impact. Hence, with an increasing parameter value, the simulated biomass becomes smaller. Second, the

Table 3.2: Linearity parameter l_n , its standard deviation, the corresponding parameter value (x_{1n}) and its standard deviation for selected model parameters within the target groups.

parameter	l_n		x_{1n}		parameter	l_n		x_{1n}	
	mean	std	mean	std		mean	std	mean	std
flux									
TopV	0.38	0.030	21.60	0.358	G2	0.38	0.029	2.13	0.086
l1	0.60	0.013	4.41	0.121	al2	0.45	0.002	0.01	0.000
NOini1	0.43	0.014	8.54	0.273	TopV	0.39	0.021	21.69	0.251
WCini1	0.30	0.022	19.04	0.647					
nitrogen									
all	0.41	0.032	0.01	0.001	all	0.39	0.040	0.01	0.001
l1	0.43	0.029	2.85	0.262	al2	0.35	0.038	0.01	0.001
n1	0.43	0.016	1.46	0.010	KS2	0.44	0.018	99.59	1.66
Nit1	0.37	0.048	0.44	0.044	l2	0.43	0.020	2.91	0.184
Nit2	0.31	0.039	0.22	0.022	n1	0.43	0.037	1.46	0.022
DeNit1	0.45	0.029	0.50	0.026	n2	0.44	0.038	1.46	0.023

linearity parameter moves from left-skewed to linear to right-skewed. Thus, in the beginning of yield formation, smaller values have a higher impact on *GenBm*, whereas in the end, greater parameter values have a higher impact on *GenBm*. Third, a linear impact, the crossing of the red line with the orange dots, coincides with the moment of highest

parameter impact.

For *VegBm*, the pattern is different compared to the pattern of *GenBm*, but it is the same across parameters. Regardless, the time from March to June resembles the pattern described before. With an increasing $idxvalue$, 1_n approaches 0.5, and with a decreasing impact, it diverges from 0.5. The impact is also negative, i.e., *VegBm* decreases with increasing parameter value. In June, however, the direction of the impact switches from negative to positive, and the approach towards linearity from below restarts.

To show that this behavior is reproducible for other TGVs, we present the CUSUNORO values for *NH₃30*, *aET*, *mp15* and *wc30* in figure 3.31. The described pattern for *VegBm* is valid for *mp15*, *wc30* and

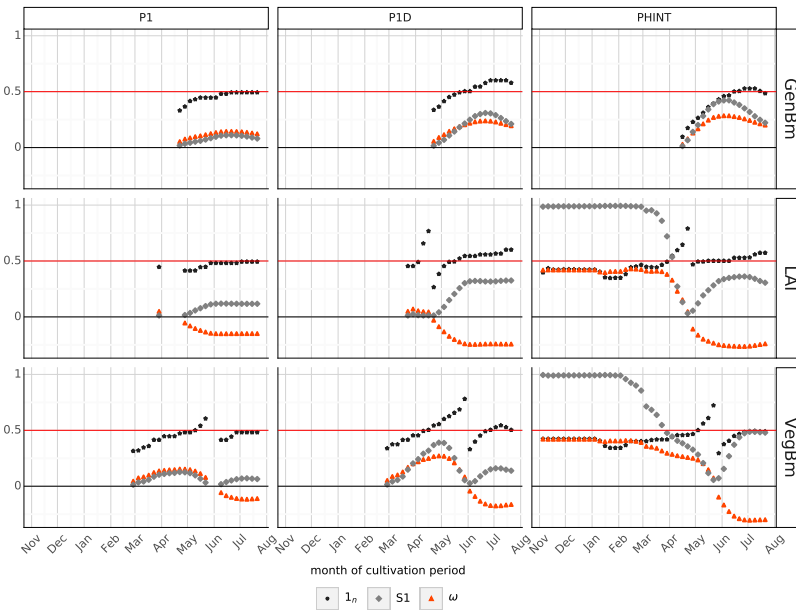


Figure 3.30: Values extracted from the CUSUNORO curve for the vegetative biomass (*VegBm*) and the generative biomass (*genBm*) for characteristic parameters.

aET , with the exceptions described below. The pattern for the $P1$ and $P1D$ is also comparable to the pattern shown in figure 3.30 (not shown). For $mp15$ and $wc30$, the direct influence of $PHINT$ is interrupted during June, before it restarts with a positive effect. This is the same time window in which these parameters always have only an interaction effect (cf., among others, 3.18). For aET , the pattern is identical. However, it is different for NO_330 , where at first the impact of $PHINT$ is positive, with a first turning point in April. Afterward, its impact increases and decreases two times. Each decrease and increase coincides with 1_n being less than 0.5 and being greater than 0.5, respectively.

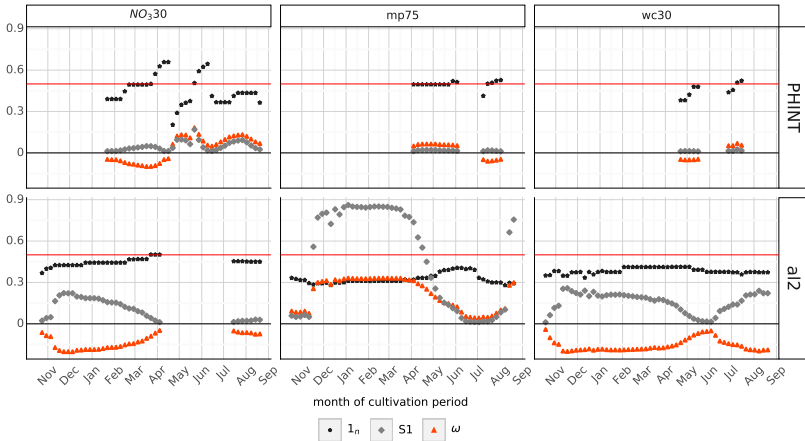


Figure 3.31: Values extracted from the CUSUNORO curve for the ammonium and water content in the top soil (NH_430 , $wc30$), the matric potential in 15 cm ($mp15$) and the actual evapotranspiration (aET) for the parameter $PHINT$.

3.3.4 Summary

In this section, we summarize the results of section 3.3.2 to provide a conclusive overview of the main findings observed in time series of parameter sensitivities and derived prescriptions for the application of different GSA methods. In table ??, we list the five characteristics for each TGV. First, the table shows which GSA method identifies more parameters as sensitive. Second, it shows if, when and for how long the initial conditions impact the model output. Third, table ?? indicates the time of highest interaction, i.e., the worst determinability of the estimation of the parameters. Fifth, it states whether the S1 to ST ratio is poor. We summarized *VegBm* and *LAI* as vegetative and *GenBm* and $N - gBm$ as generative in the crop group and did not further differentiate TGVs of the flux group. Time periods give the duration of an impact, beginning at the simulation start. We only briefly describe the findings because they are discussed in detail in Chapter IV.

ST identifies more parameters as sensitive than the SI^{MI} only for TGV_{frc}^{fx} s and the matric potentials. Otherwise, MI methods are more sensitive towards parameters that have a low impact. Furthermore, the initial water content impacts almost every TGV, but its impact often fades in time ranges of one week to three months. However, in some cases, it is also sensitive during the vegetation period. The initial water content affects the UCSR during the whole simulation period only for TGV_{90}^N . The initial NO_3^- content is a crucial parameter for the UCSR of the whole model, whereas the initial NH_4^+ content is of minor importance only for our TGVs. From the last two columns of table ??, it can be seen that the interesting time in our model is the vegetation period; there each TGV except the water content is impacted by interaction effects.

For the NO_3^- content of the soil and the matric potential, the $S1^{PF}$ to ST ratio is all-time low. In table 3.3, we provide the parameter ranks of each SI for the total model. The ranks refer to the weighted

Table 3.3: Parameter ranks for the weighted and averaged sensitivity indices $S1^{PF}$, ST, $\hat{\delta}$ and β^{ks} . The average is taken over time and target variables. Parameter count gives the number of target variables that the parameter influences.

#	parameter	parameter rank				parameter count			
		$S1^{PF}$	ST	δ^{pt}	β^{ks}	$S1^{PF}$	ST	δ^{pt}	β^{ks}
1	PHINT	1	1	1	1	16	16	16	16
2	P1D	2	2	2	2	16	16	16	16
3	kcmid	3	4	3	4	8	12	11	12
4	al1	4	5	4	3	11	15	14	14
5	NOini1	5	8	5	7	10	12	13	15
6	P1	6	3	7	6	16	16	16	16
7	al2	7	6	6	5	11	15	15	15
8	Nit1	8	14	10	13	2	5	4	4
9	NOini2	9	18	9	12	5	11	11	15
10	ToptG	10	7	12	8	12	15	15	15
11	n2	11	9	11	10	8	13	15	15
12	n1	12	10	8	9	10	13	14	14
13	Nit2	13	21	20	24	1	3	2	3
14	l2	14	13	13	11	9	12	13	13
15	WCini1	15	19	17	18	11	11	13	14
16	WCini2	16	20	18	22	6	12	13	12
17	P5	17	11	22	17	10	15	15	16
18	ToptV	18	12	19	14	12	16	16	16
19	kcini	19	17	16	16	9	13	14	14
20	G2	20	23	15	20	3	10	7	7
21	l1	21	15	14	15	10	11	14	14
22	MiMa	22	29	24	27	2	5	3	4
23	kcend	23	22	21	21	6	7	6	9
24	P1V	24	16	23	19	8	14	15	16
25	MiHu1	25	28	25	23	7	12	10	11
26	MiLi1	26	27	26	25	4	10	10	13
27	NHini1	27	33	32	32	1	4	4	5
28	KS2	28	25	28	26	2	9	12	11
29	DeNit1	29	30	29	28	2	11	10	11
30	NHini2	30	34	34	34	1	2	1	1
31	mxWup	31	26	33	33	1	4	4	7
32	G1	32	32	27	30	2	9	7	7
33	P4	33	24	30	29	1	12	7	12
34	MiHu2	34	35	35	35	1	2	1	2
35	KS1	37	31	31	31	-	4	7	9
36	mxNup	37	36	37.5	37.5	-	1	-	-
37	DeNit2	37	37	37.5	37.5	-	1	-	-
38	x	37	38	37.5	37.5	-	-	-	-

idxvalue averaged over the whole simulation period and all TGVs. The parameter count indicates the number of TGVs that the parameter affects. The table is ordered by the ranks of $S1^{PF}$. If a parameter is ranked higher by ST, one can conclude that it has a higher total effect on the model than a first-order effect. If SI^{MI} rank a parameter higher

than SI^{VB} , this indicates that they do affect the UCSR but not the VSR. In this case, variance might not be an appropriate representation of the UCSR. Parameter counts provide information regarding whether a parameter frequently has a first-order effect or if it predominantly affects the model by interaction. For example, *PHINT*, *P1* and *P1D* impact all TGVs according to all SIs. However, *P1* is ranked higher by ST, which is an indicator that it often has a poor $S1^{PF}$ to ST ratio. In contrast, *Nit1* is ranked higher by $S1^{PF}$ and directly impacts only two TGVs. This indicates that its first-order impact must be quite high because it is listed within the top ten ranked parameters. Another example is the parameter *n1*, which has a larger impact on the UCSR of more TGVs than it has an impact on the VSR.

Some general statements about parameter sensitivities can be made. In general, the different SIs rank the same seven parameters at the top, and these are also the parameters that affect at least 50% of our TGVs. Furthermore, it can be concluded that 34 parameters at some point have a first-order effect, and that only three parameters are not important to our model at all. We have 19 parameters with a direct cross-model first-order impact. The different rank order below rank 10 indicates that there are parameters whose determinability within the model is restricted.

Table 3.4: Comprehensive overview of the results of the global sensitivity analysis for the different target variables. A "+" simply means "has an impact," and "-" indicates "has no impact". An x in parenthesis means that the specific characteristic is true for only one SI, a "*" means that this is only true during the vegetation period. If there is no "~" in the last column, the ratio of SI to ST is only poor during the time specified in the time of interaction.

group	#VB>	initial conditions	time of	poor
target variable	#MI	NH_4^+	interaction	SI/ST
crop				
vegetative	-	-	Apr/May	x*
generative	-	-	May-Jun	x
water content				
30	-	-	May	-
90	-	-	Jun	-
matrix potential				
15	x	(x)*	> mid Feb	~ x
75	x	(x)*	> mid Mar	~ x
NO_3^- content				
30	-	2 month	> Jan	~ x
90	-	(x)*	mid Apr	~ x
NH_4^+ content				
30	-	2 month	> Jan	x
90	-	(4 month)	> mid Apr	-
flux	x	1 week/Apr	> Mar	x*

() not true for all sensitivity indices. * only true for the vegetation period. ~ true for the whole simulation period.

IV. Discussion

We discuss the highlights, commonalities and peculiarities with regard to the two objectives (cf. section 1.2). Furthermore, our goal is to relate the results to each other to inform the reader of what must be considered when applying a time-resolved GSA to a soil-crop model. In section 4.1, we discuss (1) why parameter sensitivities for the TGVs of the crop group strongly depend on *dev* and why temporally resolved parameter sensitivities are crucial to better understand the relations in XN-CERES; and (2) why the application of a GSA to soil-crop models is special, and the results must be evaluated carefully. In sections 4.2, 4.3 and 4.4, we discuss the results of the parameter sensitivities for the TGVs of the water, nitrogen and flux groups. We focus on cross-module parameter dependencies and evaluate the results with regard to the two goals. In section 4.5, we briefly discuss the different SIs with regard to their estimation and their reliability for the simulation with XN-CERES.

4.1 Crop Group

4.1.1 Vegetative Growth

The temporally resolved parameter sensitivities and the use of different GSA methods show a strong dependency of the UCSR of crop growth, i.e., leaf and biomass formation, on crop development, i.e., *dev*, in XN-CERES. We discuss this in detail in the context of the temporal course

of the model’s additivity (c.f. fig. 3.7), the model output distribution (c.f. fig. 3.1) and the parameter sensitivities (c.f. fig. 3.12 and 3.13) of LAI and $VegBm$. We distinguish between three different periods. Period I refers to the period from field emergence in November to the start of the vegetation period in mid-February. Period II includes the period from mid-February to the onset of yield formation in mid-April, and period III refers to the time from mid-April to harvest. The date of the onset of yield formation is based on the maximum simulated dev in the simulation set.

Period I As shown in section 3.3.1 and 3.3.2.1, during period I, the TGV_{veg}^{pl} are totally additive and dominated by $PHINT$. The simulated dev does not exceed BBCH 39, which means that the crop is either in the XN-CERES stage 1 or 2. These are the only two stages where leaf formation is simulated in XN-CERES and the corresponding equations depend on $PHINT$. First, we can conclude that the set parameter ranges do not allow for faster crop development. Put the other way around, parameter combinations exist that allow the crop to finish leaf formation within period I. Second, we can conclude that the parameter sensitivities and the UCSR of $VegBm$ and LAI are indirectly controlled by dev . This relation is due to the development stage-dependent selection of the currently valid mathematical description in the model, by design. Since in period I, the present dev is restricted to stage 1 and 2, all crops in the simulation set are affected by the parameter $PHINT$. While SI^{VB} s confirm this by apportioning the entire UCSR to $PHINT$, SI^{MI} additionally classify $P1V$ as sensitive, although the idxvalue is small ($P1V$ is at that time the key driver of the UCSR of dev). Hence, SI^{MI} support the conclusion that the UCSR of the TGV_{veg}^{pl} s is impacted by dev . Furthermore, on the one hand, it indicates that it makes a difference for the UCSR if the crop is in stage 1 or 2. On the other hand, this may also be an indicator that variance is not the perfect representation of the UCSR. It may even mean both. Nevertheless, this suggests that even small deviations from the normal distribution

(cf. fig. 3.2) affect the reliability of SI^{VB} s. For now, there are two conclusions we can retain. First, the simulation of *dev* is not only crucial for *dev* itself, but also influences the simulated *LAI* and *VegBm*. This impact is due to the BBCH-dependent selection of the currently valid mathematical description in the model. In general, it is a common feature of crop models that the mathematical description for TGVs of the crop group is not constant over time and therefore changes with ongoing crop development. Second, *PHINT* is the most important parameter; SI comprehensively, but MI methods give a more detailed insight into the relations in the model.

Period II In Period II, the parameter sensitivities show a more interesting pattern, and the above-discussed relation between *dev* and the simulated TGV_{veg}^{pl} becomes more apparent. Furthermore, *LAI* and *VegBm* have distinct patterns. The slow decrease of the model additivity and the loss of *PHINT*'s predominant first-order effect, for both TGV_{veg}^{pl} , is a result of the fact that, from February onwards, with each evaluated date, more and more crops reach a development stage beyond BBCH 39. The consequence is a broader range of simulated development stages, which increases the number of currently valid mathematical descriptions. In turn, this means that fewer and fewer crops are affected by *PHINT* and its direct impact decreases. This is special for soil-crop models and can be generalized for all TGVs of the crop group. A parameter's sensitivity at a specific point in time may be low simply because it only affects the crops in a subset of the total simulation set. Nonetheless, this parameter can explain 100% of the UCSR of that subset. Filtering the simulations set by *dev* would in fact only mark parameters as sensitive that affect a specific *dev*, but important correlations such as the dependence between *dev* and *VegBm* would remain undiscovered.

In February and March, respectively, the model additivity and the predominant first-order effect of *PHINT* decrease. At the same time,

P1D and *P1* gain influence. These two parameters are the key drivers of uncertainty for *dev*. Therefore, now, the dependence between the simulated *dev* and the TGV_{veg}^{pl} is directly perceptible from the parameter sensitivities of all SIs. The crucial role of *dev* for the SR of *LAI* becomes even clearer if we consider the fact that period II is also the prime time for the evolution of the simulated *LAI* in value and range (cf. fig: 3.1). Therefore, we can conclude that *LAI* is mostly affected by the total duration of stages 1 and 2. If this is too short, the crop stops leaf formation too early. If it is too long, the *LAI* might be too high. As a brief aside, the *LAI* formation is much faster here, because it primarily depends on the model parameter *total biomass growth rate*. It is calculated from the constant *light use efficiency*, the extracted light (50% of the given global radiation) and the *LAI* itself. Furthermore, it is temperature-dependent. With higher temperatures and longer days in spring, the biomass growth rate increases, which leads to an increase in *LAI*. This again increases the biomass growth rate. Thus, it is a reinforcing loop and explains the exponential increase in *LAI*.

However, according to SI^{VB} s, *LAI* is increasingly affected by parameter interactions in period II. With an increase in value, the parameter *PHINT* has a negative effect on *LAI*, whereas the parameters *P1V* and *P1D* have a positive effect. Therefore, multiple parameter combinations of the three parameters exist, that result in the same *LAI*. The UCSR of *LAI* also depends on the SNWC, which is indicated by the parameter sensitivities of *NOini1*, MvG parameters and mineralization rates. The dependence of *LAI* on the state of the soil water has also been found by DeJonge et al. (2012). Because these parameters show a first-order effect, the latter cannot be compensated by any other parameter. In a calibration process, this means that these parameters have to be considered and should be in accordance with *LAI* measurements. If they are fixed in advance, it is possible that the measured *LAI* cannot be simulated. It has to be tested if calibrating these parameters on *LAI* yields a correct representation of SNWC. If this is not the case, we could conclude that the relation in the model

is not adequately described. The results also show that measurements of *LAI* and SNWC should be assessed in April and May.

From the fact that these parameters do not show a first-order effect on *VegBm* and that, in XN-CERES, first leaf formation is reduced and then *VegBm*, two inferences can be made. On the one hand, N- and water stress occur within our simulation set. On the other hand, it is not severe enough to also impact directly *VegBm*.

The relation described above between *dev* and *LAI* is directly transferable to *VegBm*. The only difference is in time because biomass formation does not end with stage 2. In XN-CERES, it is considered until harvest, but it is secondary during yield formation. Then, assimilates are primarily transported to the grains. Between BBCH 40 and BBCH 79, the assimilates are partitioned between roots and *VegBm*, and the ratio is *dev* specific. To calibrate *VegBm*, additional measurements of *dev* must be available.

VegBm is also affected by the SNWC, that is, by group foreign parameters. Whereas SI^{VB} s only identify water stress as important, SI^{MI} additionally identify nitrogen stress. SI^{VB} s could not identify the relation between *VegBm* and the soil N condition. Since the output distribution is leptokurtic and bimodal, the use of MI-methods is useful in period II. However, fixing MvG and nitrogen parameters is not possible without affecting the values of the calibrated crop parameters. Hence, if the model is calibrated on measurements of *VegBm* and MvG and nitrogen parameters are fixed in advance, the optimized crop parameters are conditional on the fixed values.

Assuming that the relation between SNWC and *VegBm* is adequately described in XN-CERES, this would have the following consequences in a prediction context. Assume that the crop parameters have been calibrated on *VegBm* measurements with MvG and nitrogen parameters fixed in advance. If these fixed parameters now lead to, e.g., water stress in the simulations that was not there in reality, the crop parameters will compensate for the water stress-induced reduction of *VegGm* in the model. If these calibrated parameters are transferred to

a simulation study where no water stress in the field occurs, these parameters will overestimate *VegBm*. For calibration, however, *VegBm* measurements do not provide enough information to fix MvG and nitrogen parameters. Either the modeler needs an additional TGV to calibrate these parameters simultaneously, or the crop sub-model, in this case, as a standalone, would be over-parametrized.

Furthermore, the GSA should be repeated in a model setup where N and water stress are more severe. MvG and nitrogen parameters could then also directly impact *VegBm* because then the biomass growth rate would be reduced until it also affects *VegBm*. The results of XING et al. (2017) support the conclusion that more parameters are sensitive in water-limiting conditions; that study investigated the sensitivities of crop parameters for different water managements.

Period III The beginning of period III ends the model additivity decline. On the one hand, it corresponds to the point of inflection from which the simulated *LAI* decreases on average. On the other hand, it corresponds to the moment when 50% of the crops in the simulation set complete stage 2. Recall that stage 2 is the end of leaf growth in XN-CERES (cf. section 2.3). All higher stages simulate leaf senescence, and each development stage has a different mathematical description. With each stage, the senescence is accelerated. The mathematical description is fixed in the different stages and cannot be directly adjusted by any parameter. In addition to the low influence of the SNWC, *LAI*, from stage 3 onwards, directly depends on *dev*. This is induced by the dependence of *dev* and *LAI* on the same parameters. In period III, it is only important which equation is currently active. The decrease of the foreign group parameter sensitivities is again due to the fact that leaf formation is still simulated only in an ever-shrinking subset,. However, when this is the case, the crop is increasingly affected by SNWC.

Furthermore, the former negative effect of *PHINT* becomes a positive effect. As a consequence, this could lead to a competitive situation in a calibration scenario. The parameters *P1*, *P1V* and *PHINT* must be

estimated simultaneously to find the parameter combination that, on the one hand, does not underestimate LAI formation in stage 2 and, on the the other hand does not underestimate leaf senescence by making the crop stay too long in stage 3.

Starting at BBCH 70, i.e., XN-CERES stage 5, assimilates are primarily transported to the grains. They are only used for vegetative growth in the case of surplus. Furthermore, senescence decreases the overall biomass growth rate. Therefore, BBCH 70 is the divide where the mathematical model description differentiates itself from the previous one. The decrease in model additivity and the inflection point of $VegBm$ are exactly at the point where 50% of the crops in the simulation set have reached BBCH 70 (cf. fig. 3.1). Furthermore, the direction of the impact of the crop parameters changes (cf. fig. 3.30). We can conclude that the time of interaction is highest when the mathematical descriptions that are currently valid in the whole simulation set have a contrasting impact on the TGV_{veg}^{pl} and equally affect the two halves of the simulation set. If one mechanism predominates in the simulation set, the parameter sensitivity again increases. The change of the impact direction can be explained: The three parameters slow down the crop development and therefore biomass growth. However, if switching from vegetative to generative growth is decelerated, the crop can invest the assimilates in leaf and biomass formation for a longer time. This once again underlines the importance of accurately determining the moments when the major development stages change. In sum, dev cannot be considered independently from the other TGVs of the crop group.

4.1.2 Generative Growth

Unless otherwise specified, all statements in this section refer to the two TGV_{gen}^{pl} s, $GenBm$ and $N - gBm$. First, we note that, at harvest, 97.5% of the simulations reached the generative phase. Hence, the parameter ranges allow for a trustworthy reaching of the genera-

tive phase. Only a few parameter combinations fail. For the TGV_{gen}^{pl} , the use of MI methods is recommended because, at the onset of yield formation, the distributions are highly skewed and leptokrusic. The variance is a poor representation of the UCSR. The transition from vegetative to generative growth is a continuous process. In the beginning, the overall uncertainty is low because in most of the simulations, $GenBm$ and $N - gBm$ are zero. The two TGV_{gen}^{pl} s are non-zero only for a very small subset t . In that case, the variance can be quite large, although the uncertainty is actually small.

As for TGV_{veg}^{pl} s, parameter sensitivities are small in the beginning because they affect only a subset of the simulations. It increases as this subset increases with each date evaluated. Therefore, the low additivity in the beginning is of minor importance and should not be considered as an indicator of the low determinability of the two TGV_{gen}^{pl} s. This is supported by β^{ks} , which assigns only little sensitivity to $PHINT$, $P1D$, $P1$, $P1V$ and $P4$ in the beginning. Nevertheless, TGV_{gen}^{pl} s, like TGV_{veg}^{pl} , strongly depend on dev which is indicated by the increasing idxvalues of β^{ks} and $S1^{PF}$. The dependence of above-ground biomass at harvest on dev is in accordance with the results of Tan et al. (2016) and Specka et al. (2015).

The reliability of δ^{pt} at the onset of yield formation, is reduced because the used kde is not suited for highly leptokrusic functions, and the integral becomes infinite. However, one can also argue that $PHINT$, $P1$ and $P1D$ are most important for yield formation in the beginning because they are responsible for the generation of non-zero model output for TGV_{gen}^{pl} . Nevertheless, again, determining the date when the mathematical descriptions switch from vegetative to generative growth is crucial for the simulation of the yield formation. This is supported by the fact that $ToptV$ is only sensitive during the generative phase, and that it has a negative impact on $GenBm$ and a positive effect on $VegBm$. Thus, if the optimal temperature is too low, the crop reaches the generative phase too late because the crop development is too slow. In a calibration context, the optimal switching date for yield

formation need not coincide with the measured date. Therefore, dev and TGV_{gen}^{pl} should be calibrated simultaneously. If the modeler only has information about the yield, the parameters are not unequivocally determinable. The process of yield formation in the model is then over-parameterized. This has also been found by Confalonieri (2010).

If the majority of the crops in the simulation set are in the generative phase, the yield is mostly controlled by dev and the duration of the XN-CERES stages 1 to 5. Interestingly, the parameter $P4$ is not determinable with regard to any TGV of the crop group; it only interacts with other parameters. This is also true for the parameter $G1$. Therefore, the number of grains depend on a number of interacting parameters. A low $G1$ can be compensated by a longer grain-filling phase and vice versa. However, the maximum grain growth rate, i.e., $G2$, directly controls yield amount and yield quality.

In addition to the strong dependence on dev , yield formation and especially $N - gBm$ are affected by interaction effects, in particular by the SNWC. Furthermore, the initial nitrogen content is important over the total possible range, although the crop is fertilized. Therefore, assessing the initial soil N conditions is important, more so than mineralization rates. For $N - gBm$, $ToptG$ is important, whereas it has no direct impact on $GenBm$. From ω , we know that lower optimal temperatures promote higher N concentrations in the grain. Therefore, considering yield amount and quality simultaneously may help to increase the determinability of the crop parameters.

4.1.3 Development Stage

According to the parameter sensitivities and the model additivity, dev is perfectly determinable and impacted by eight parameters. However, we would like to briefly address why these parameters show high interactions with regard to TGV_{veg}^{pl} and TGV_{gen}^{pl} . First, dev can only increase, and parameters can only accelerate or decelerate this increase. Furthermore, at each stage, it is clearly defined which parameter con-

trols the pace of the crop development. Except for stage 1, all stages are controlled by only one parameter. In these stages, there can be no interaction. In stage 1, at least in the beginning, 10% of the VSR are caused by the interaction of $P1$, $P1V$ and $P1D$. The interaction between $P1D$ and $P1V$ decreases as soon as the temperature and day length increase because then $P1V$ loses its impact. Since the winter months are cold, the evolution of dev is slow, and the parameters have little affect on SR of dev . A small value can only be increased or decreased minimally by constants. This is supported by the fact that, dev and its simulation range hardly increase, while in the majority of the simulation set the crop is in stage 1.

Stages 2 and 3 are only controlled by $PHINT$, and the duration of the individual stages do not interact; hence, these two stages are auto-correlated. If the crop develops slowly in stage 2, it will also develop slowly in stage 3 because the impact of $PHINT$ is negative the entire time. The joint effect of different parameter values becomes first apparent when the majority of the crops reach stage 4 in May. This is also indicated by the decrease in the model's additivity. However, to adequately estimate the individual parameters, the modeler needs information on the moment each individual stage is reached. Otherwise, with higher stages, different combinations of parameters can lead to the same moment when a crop reaches stage 4 or 5. However, for the other TGV_{veg}^{pl} s, the time of attainment of each individual development stage is important. During calibration, dev and TGVs for the crop group should not be considered independent, and crop parameters should be fixed on measurements from more than one TGV of the crop group.

Since the distribution is skewed and leptokurtic in the beginning and bi- and trimodal during the vegetation period, the ratio of $S1^{PF}$ to ST is not fully reliable. The sensitivity of $P4$ is a second indicator that VB methods fail and that dev is not fully additive. It is only rated as sensitive by SI^{MI} . Nonetheless, both methods result in the same ranking and, therefore, the SR of dev is unambiguously controlled by the sensitive parameters.

4.2 Water Group

4.2.1 Soil Water Content

The determinability of the parameters that impact the soil water content is high for the whole simulation period (cf. fig. 3.8). The water content in top soil is governed by the MvG parameters $al1$ and $n1$. Higher values lead to a lower simulated $wc30$ because the air entry point is then at a lower matric potential, and the slope of the retention curve is steeper. In addition, the decrease of the unsaturated hydraulic conductivity is less steep. The positive effect of $al2$ on the $wc30$ is due to the lowered hydraulic conductivity leading to a capillary lock.

During the vegetation period, the soil water regime is strongly determined by the upper boundary condition, i.e., ET, which is scaled by the crop factors ($kcmid$ and $kcini$). Hence, a correct assessment of the actual ET could reduce the uncertainty during the vegetation period for the two soil horizons by up to 50% and 65%, respectively. The importance of the crop factors for the simulation of the soil water content and the high additivity of the water sub-model has also been found by Stahn et al. (2017). The huge throughput of the crop factors shows the importance of the soil water reservoir in the subsoil for the unrestricted ET. Since most water is transpired during that time (cf. fig. 3.1), it is important for the unlimited water supply of the crop. Nevertheless, the depletion of deeper soil layers affects the SR of crop growth because the crop has to invest assimilates into root growth. These assimilates are then not available for leaf and biomass formation. The stress-adapted partitioning scheme between roots and above-ground biomass is development stage-specific. Hence, the simulated water content impacts the crop growth by changing the partitioning, and changing the partitioning in turn impacts the water content. Indeed, in March, the simulated rooting depth diverges within the simulation set and ranges from 90 to 160 cm soil depth at harvest (results not shown). Since the stress-adapted partitioning is different in each stage, the water content

depend on *dev*. The water stress and adapting partitioning scheme is a reinforcing loop, and hence it expresses the interaction of crop and water parameters. In our model setup, the potential ET has not been reduced because the crops had the possibility of rooting deeper soil layers. This explains why the crop factors do not have an impact on any TGV of the crop group. The higher impact of crop parameters on the sub soil suggests that water parameters could also have a direct impact on *VegBm* if the crop cannot compensate the atmospheric water demand through enhanced root growth. This view is supported by the dependency of *VegBm* on the maximum allowed rooting depth, which was tested in a preliminary GSA with a different setup.

The importance of root growth simulation for the top soil water content has already been shown by Gayler et al. (2014) for the land surface model Noah-MP. To evaluate parameter sensitivities and their cross-model impact in drought situations, the GSA should be either repeated in a dry climate or with shallow soil. The combination of sufficient rainfall and shallow soil already leads to a small increase in the sensitivity of crop and water parameters on crop growth and soil water content, respectively. This has been tested in preliminary GSA (results not shown) and found in accordance with Richter et al. (2010), Confalonieri et al. (2010a) and Vanuytrecht et al. (2014). They found that the parameter sensitivities depend on the site and weather conditions. Nevertheless, the water content is mainly controlled by the hydraulic properties, which contrasts with the results of Stahn et al. (2017). However, since the reduction of the water uptake in the model depends on the matric potential in the soil, our result, that crop parameters are not of great importance, is not surprising. Hence, the calibration of the MvG parameters on measured water content works well, but it does not ensure that the cultivation conditions for the crop match reality.

MI methods do not provide any further insights, and methods do well compare (cf. fig. 3.17 and 3.6). However, it is difficult to state whether the low-ranked parameters are truly important because of the convergence behavior of the SIs. This has to be tested in a calibration study.

However, the impact of KS on the UCSR of the soil water content during the whole simulation period suggests that it cannot be fixed without influencing the calibration of the other parameters. Since this is not captured by SI^{VB}_s , KS has no effect on the VSR.

4.2.2 Matric Potential

During the winter, the matric potential depends on the MvG parameters of the respective depth. However, according to SI^{VB}_s these parameters highly interact with each other, and in the first weeks, they also interact with the initial water content. Hence, the measurement of the initial water content would significantly reduce the UCSR (cf. fig. 3.8). Otherwise, a spin-up of at least three months should be considered (cf. table ??). MvG parameters for $mp75$ may then be clearly identifiable. However, the determinability of $mp15$ is always low. This could be due to the dependence on the water state of the sub-soil (cf. fig. 3.20). However, the interaction effect might also be overestimated because the reliability of SI^{VB} is curtailed. The output distribution is highly leptokurtic and skewed, and variance is not an appropriate measure of uncertainty (cf. fig. 3.2). Nevertheless, it is conceivable that various combinations of the MvG parameters exist that lead to the same simulated matric potential and to the same shape of the retention curve.

During the vegetation period, the UCSR is caused by parameters from all groups but nitrogen. Therefore, the matric potential is the crucial TGV that impacts and is impacted by crop growth, the SNWC and the actual ET. According to SI^{VB}_s , the impact occurs mostly by interaction. Both the crop parameters' impact and the low determinability compare with the results of Stahn et al. (2017). As a consequence, the calibration of the sensitive parameters on the matric potential is difficult. Many parameter combinations exist that lead to the same simulated matric potential. For example, a high potential ET increases root water uptake, which decreases the matric potential. However, if

the decrease in matric potential leads to water-limiting conditions in the soil, is controlled by the MvG parameters. Therefore, it might or might not lead to a shift in the partitioning between roots and above-ground biomass. The partitioning again depends on *dev*. That crop parameters have a larger impact on the matric potential than the MvG parameters on the TGV of the crop group; however, they could be subjected to root water uptake from deeper soil layers. That was the case in at least 75% of the simulations, where the *mp15* and *mp75* were not below pF 2 and pF 2.3, respectively (cf. fig. 3.1). As a result, crop growth is only impacted by the matric potential in water-limiting conditions, whereas the matric potential is always impacted by constant occurring root water uptake. How the root water uptake changes the simulated matric potential depends on the MvG parameters. Hence, different combinations of crop and MvG parameters lead to the same SR of the matric potential, and VSR is only caused by the combination of these parameters. In a calibration context, the matric potential is not suitable to determine parameter values. Thus, the modeler needs measurements of *VegBm* and *GenBm* and therefore also of *dev* to have additional information to estimate crop parameters. Since the MvG parameters are clearly identifiable for the soil water content, one could include them in a calibration. Hence, the matric potential is an ideal TGV to test if the mathematical descriptions in XN-CERES are an accurate representation of reality.

The overall importance of interactions and the dependance on 35 parameters, however, should be viewed critically. Since the overall UCSR is much smaller than the variance, the determinability of the individual parameters may be better than SI^{VB} s suggest. This could explain the good determinability of the MvG parameters, despite the high ST and low $S1^{PF}$ in Stahn et al. (2017). The parameters further classified as sensitive by ST are also those that did not converge with regard to *cv_{gidx}*. Therefore, graphically inspecting the confidence intervals of each index before parameter selection is recommended. If idxvalues are

low, and the lower confidence bound is below 0, the parameter could be omitted from calibration, as suggested in Nossent et al. (2011).

4.3 Nitrogen Group

4.3.1 NO_3^- Content

The accurate measurement of the initial soil NO_3^- content would lead to a great reduction in the UCSR of $TGV^{NO_3^-}$, especially over the winter months. At the same time, the impact of crop parameters tends towards zero, and MvG parameters are the second major source of uncertainty. Thus, the initial NO_3^- content is most important for simulations of leached NO_3^- to groundwater. Although the impact is much lower, the initial NO_3^- content is also a source of uncertainty during the vegetation period for the NO_3^- content and the crops' vegetative and generative growth. For the simulation of a crop rotation, i.e., multi-year yield predictions, this means that the end soil NO_3^- content of the preceding crop (C1), which is the initial content for the following crop (C2), causes uncertainty in the predictions of the latter's (C2) yield. Since the NO_3^- content at the simulation end strongly depends on MvG and crop parameters, the initial NO_3^- content for the following crop (C2) depends also on the crop and MvG parameters used with the preceding crop (C1). As a consequence, optimized crop parameters must be in accordance with measurements of the NO_3^- content in the soil, at least at the end of the simulation. Otherwise, the optimized crop parameters may match crop measurements, but they may deplete the soil too much, and thus the following crop does not grow as well. Of course, the reverse may also be true. Therefore, the initial NO_3^- content introduces a large, propagating uncertainty in the NO_3^- content, the leached NO_3^- and the crop growth.

However, the MvG parameters are most important for the NO_3^- content, but according to SI^{VB} , it is difficult to calibrate on NO_3^- measurements. For the top soil, this is always true, whereas for the sub

soil, this is true for the vegetation period. Hence, in the vegetation period, the NO_3^- content depends on group-foreign parameters. For the estimation of these parameters, measurements of the NO_3^- content do not provide enough information. Hence, to estimate them, measurements of other TGVs such as *VegBm*, *GenBm* and therefore also *dev*, water content and matric potential are necessary. As a consequence, the NO_3^- content is the TGV that is impacted most by the state of the other TGVs. From a different perspective, the NO_3^- content is an ideal TGV to test how well the model represents the real system. Thus, calibrating the parameters on other measurements should ideally lead to an accurate approximation of the NO_3^- measurements.

Given that SI^{VB} agree with SI^{MI} , it is a reliable finding that crop and MvG parameters impact the UCSR of the NO_3^- content the most, although the output distributions are skewed and leptokurtic. However, the ratio of $S1^{PF}$ and ST is uncertain. Nevertheless, MvG parameters and crop parameters have a significant first-order effect on the NO_3^- content. Therefore, a multi-objective calibration should include the NO_3^- content, matrix potential and water content, and crop growth to enable a joint, satisfactory solution for the MvG parameters and *PHINT*, *P1* and *P1D*. The inclusion of the N content of the crop's biomass would additionally reduce the uncertainty in the simulated NO_3^- content. Interestingly, the parameter *mxNup* is unimportant for the simulation of crop growth and the N content, indicating that the simulated root system is large enough to supply the crop with enough N. Limiting N conditions can therefore only be attributed to low N contents in the soil, and not limitations in the crop. This suggests that *mxNup*'s value range is set too high, and that it has no effect on the SR, i.e., it is a superfluous parameter, at least in our model setup. This is also true for *mxWup*.

Reliable predictions of nitrate leaching require that the soil water condition and the crop's water and N uptake are known. Liang et al. (2017) have also found that the hydraulic properties and crop parameters are more important than the N-transformation parameters for the

VSR of N-leaching to the groundwater. Mineralization rates are more important for the top soil and do not cause much UCSR of NO_3 90. Furthermore, the mitigation of fertilization strategies is only possible if the soil-crop model's water dynamic adequately approximates the amount of leached NO_3^- and if one knows about the initial NO_3^- content. Otherwise, too much or too little would be leached, and fertilization rates would be adapted to an incorrect basis. MI methods are again useful; they show the increasing importance of mineralization rates with increasing distance from the fertilizer application date. Meanwhile, the overall UCSR of NO_3 30 decreases, but the shape of the distribution becomes more leptokurtic and skewed. Hence, VB methods are not reliable and do not show the importance of mineralization rates on the NO_3 30 content. Unfortunately, measurements of the N content do not provide enough information to calibrate mineralization rates. Again, δ^{pt} is not as reliable as β^{ks} because the construction and integration in highly tailed distributions leads to high inaccuracy of its estimates.

4.3.2 NH_4^+ Content

The NH_4^+ content largely depends on the nitrification rates and, to some extent on mineralization rates. Hence, if the N-transformation parameters are to be determined, one at least needs information about the NH_4^+ content. Again, the initial NO_3^- content is an important source of uncertainty. The interaction between fertilization, soil water regime and initial condition is most apparent for NH_4 30 but is transferable to NO_3 30. With each fertilization, the range of the SR increases. The state of the applied NH_4^+ depends on the crop and MvG parameters. Therefore, the amount of fertilizer needed strongly depends on the combination of crop and MvG parameters. The determinability of crop and MvG parameters on the NH_4^+ content could be better than SI^{VB} suggest because, due to the high leptokurtic distributions, the overall UCSR is low. In general, we can conclude that the set parameter ranges always lead to the same N contents and a low UCSR of the

TGV of the nitrogen group. Hence, all simulations of the simulation set approach the same final state. How the N is partitioned between crop and leaching depends on the parameter combination of MvG and crop parameters.

4.4 Flux Group

The SR of the TGVs of the flux group are also closely related to *dev* because the development stage controls the selection of the crop factor. With the increase of the variability of *dev* within the simulation set, the model's additivity decreases. Crop factors and *dev* can compensate for each other to some degree. Therefore, they cannot be set independently, and crop parameters should be calibrated in accordance with measured fluxes. Including crop factors could therefore compensate for the poor capability of the model observed in Wöhling et al. (2013) to simulate the reduction in transpiration due to leaf senescence. Since *kcmid* is the most important parameter that affect the matric potential, calibrating it on matric potential and flux measurements is another possibility to test the quality of the model's structure. However, since both TGVs depend on the simulated crop, these parameters cannot be disregarded.

Since the ratio of evaporation to transpiration depends on the ground cover, i.e., *LAI*, *fTrs* is strongly coupled with the parameters that cause UCSR of *LAI*. The higher variability in the SR at the end of the vegetation period is a consequence of the different dates of maturity. Hence, in a subset of simulations, the crop does not transpire anymore, whereas the rest is still in stages in which the crop assimilates and transpires. However, as is apparent from the model output's variability and distribution shape (cf. fig. 3.1 and 3.2), the UCSR of *fTrs* is small during the vegetation period. The use of MI methods is recommendable, although the difference in the identification of influential and non-influential parameters between methods is small.

The low uncertainty in the ratio of actual to potential ET (cf. fig.

3.1 and 3.2) is a clear indicator that only a small subset of crops in our simulation set suffer from water stress. However, XN-CERES becomes numerically unstable when the soil matrix potential becomes too low. The generation of an LHS with higher values of $kcmid$, al and n led to an increase of aborted simulations of about 20%. Furthermore, aborted simulation cluster where high values of the three parameters are combined. This violates the assumption that the model inputs are independent. Therefore, the sample should be generated considering correlated inputs.

Specifying the upper boundary flux, i.e., actual ET, would strongly reduce the overall uncertainty of XN-CERES. It would reduce the uncertainty in the SR of the state of the soil water and thus the interaction between the crop, water and flux modules could be reduced. Furthermore, the simulation of the ET would no longer depend on dev . That would additionally reduce the dependence between the sub-models considered. However, flux measurements are likewise uncertain and hence introduce a new source of uncertainty.

4.5 Comparison of GSA methods

The different SIs, i.e., SI^{PF} , SI^{GD} , $\hat{\delta}$, δ^{pt} , β^{ks} and β^{kui} are all suitable for the Factor Prioritization setting. They all identify the key drivers of uncertainty and in general agree on a common set. Since the convergence of the SI^{MI} and $S1^{GD}$ is reached at much lower sample sizes and hence with less computational effort, SI^{GD} are superior to SI^{PF} . The importance of parameter interactions can still be assessed by checking the model additivity for the different TGVs. Parameters without a first-order effect cause UCSR by interacting. Information concerning whether parameters that have a first-order effect also have an interaction effect cannot be provided. In our case, we only have parameters with a poor $S1^{PF}$ to ST ratio in situations where the model output distribution is skewed or leptokurtic. These are cases where variance is not an accurate representation of the UCSR. Furthermore,

in some cases, SI^{MI} follow the time track of $S1^{PF}$, and in other cases, they follow the time track of ST. Hence, a parameter ranked higher by SI^{MI} than by $S1^{PF}$ could indicate, that this parameter not only has a direct impact.

However, in the Factor Fixing setting, all SIs designed for this setting have drawbacks. The idxvalue's convergence of SI^{PF} is slow and not fully reached at sample sizes of 1,000,000. δ and β^{ks} (also β^{kui}) do converge according to the convergence criteria used, but with an increasing sample size, more and more parameters change from non-influential to influential. The reasons have been discussed in section 2.4.4.1 and are in accordance with the results in Plischke et al. (2013) and Mara et al. (2017). Therefore, with both methods, large sample sizes are needed in case of XN-CERES to achieve reliable results. The unstable ranking of low sensitive parameters has also been found by Khorashadi Zadeh et al. (2017) and Nossent et al. (2011). Despite the extensive post-processing and the use of critical values, there are still a significant number of parameters at the lower sensitivity threshold ($SI^{MI} < 0.001$). Here, the modeler still has to decide if these parameters are truly important or if the impact is negligible. The latter is supported because idxvalues are often rounded to the second decimal place. For the Factor Fixing setting, both methods create uncertainty due to convergence behavior and sample size. The modeler is then left with the decision of which parameter to select at the lower sensitivity threshold. To achieve a higher certainty for the SIs, one could repeat the assessment of the critical value for each bootstrapping sample and only choose the parameters that are above a given threshold, as was done in Khorashadi Zadeh et al. (2017). However, in our case, that would have further increased the time of analysis and the amount of data to be evaluated in the post-processing. Since the confidence intervals of the SI^{MI} are small in any case, we do not expect a different result. Graphical inspection of the parameter sensitivities and their time course provides a remedy, and graphical inspection is an intuitive decision support tool.

The use of a meta-model in our case would not reduce the computational costs because the GSA itself already took more than a month on the given HPC resources. A meta-model is an empirical model that is trained on the model output of the process model, e.g., XN-CERES. It usually takes only seconds to run, and hence the model output of larger samples can be generated much faster, and the computational costs decrease. Nevertheless, we conclude that both methods are similarly suitable for the Factor Fixing setting, which corresponds with the results of Mara et al. (2017) and Khorashadi Zadeh et al. (2017). However, in both studies, the setup was considerably simpler and implied only 10 and 26 parameters, respectively, two TGVs and no time resolution.

With XN-CERES, the use of cdf-based methods is favored over pdf-based methods, whereby the choice of the metric measuring the distance between the distribution functions is of no importance for identifying and ranking sensitive parameters. In our case, β^{kui} and β^{ks} are interchangeable. However, the used kde is unreliable with highly skewed distributions, as is the case for, e.g., *GenBm*. Furthermore, the integration of the distribution causes additional numerical noise. In comparison, $\hat{\delta}$ works better at smaller sample sizes, whereas it was not restrictive enough at higher sample sizes. At a sample size of 200.000 all 39 parameters were influential for each TGV. In contrast, δ^{pt} is highly restrictive at small sample sizes and classified most parameters as non-influential but works better for larger sample sizes. Its estimation accuracy could be improved by using an integration method other than trapezoidal rule and another kde. Last, the estimation of δ compared to the estimation of β^{ks} (β^{kui}) is computationally costly. In all, SI^{MI} based on distances between cdfs are reliable and computationally cheap.

V. Conclusion

A number of GSA methods were successfully applied to the soil-crop model XN-CERES. We have identified the key drivers of uncertainty and non-sensitive parameters for all TGVs considered and the overall model. The results suggest that the crop parameters *PHINT*, *P1* and *P1D*; the flux parameter *kcmid*; the MvG parameter *al*; and the initial NO_3^- content are the cross-module drivers of the overall uncertainty of the XN-CERES. Only three parameters could be identified and can be excluded from a calibration. Very few parameters cannot be estimated at all by measuring of the TGVs considered. Hence, parameters of XN-CERES can reliably be estimated if measurements from all groups are available. Time-resolved parameter sensitivities and the consideration of different TGVs showed that, not surprisingly, measurements should primarily be taken during the vegetation period. However, the interpretation for TGVs of the crop group is complicated by the peculiarity of crop models, which use different mathematical descriptions for one and the same TGV over time.

Including the initial conditions in the GSA revealed the great importance of the initial NO_3^- content for overall uncertainty. Furthermore, crop parameters cannot be calibrated on different TGVs of the crop group independently. Time-resolved information about the development stage is crucial. Overall, mineralization and nitrification rates play a minor role in the uncertainty, but they can be estimated from measurements of the NH_4^+ content. Furthermore, the calibration on either the matric potential or the NO_3^- content is ideal to test the

model's structure and identify its potential deficiencies if TGVs not considered in the calibration are compared to measurements.

MI methods proved to be useful because the model output distributions of XN-CERES are often skewed and leptokurtic. Furthermore, they converge faster. Using cdfs instead of pdfs is more stable in the case of highly skewed distributions. SI^{GD} converge much faster, but both methods require large sample sizes to achieve a stable ranking of low-ranked parameters. The separation of influential and non-influential parameters worked best for β^{ks} and β^{kui} . Hence, we conclude that cdf-based GSA methods and the chosen SIs β^{ks} and β^{kui} are best suited for assessing the parameter sensitivities of XN-CERES.

The GSA should be repeated in a drier situation or a shallower soil to evaluate if the derived relations between sub-models and parameter sensitivities hold for further model setups. Furthermore, in a multi-objective calibration study, the parameter determinability should be verified. In particular, it should be evaluated whether parameters can be estimated in accordance to all TGVs or if the model cannot comply with manifold TGVs measurements simultaneously. The GSA showed that fixing parameters from other groups impacts the calibration of parameters that belong to the group of interest. Hence, as an absolute requirement, values of fixed parameters should always be specified together with calibrated parameter values to make the values usable for other modelers.

Overall, for the calibration of soil-crop models, modelers should not calibrate parameters from sub-models independently, and they should focus on the multi-objective calibration of the models with at least one measurement from each group. Furthermore, to cope with the interaction effects of the MvG and crop parameters, Bayesian updating, as in Schöniger et al. (2014), should be considered. Multi-objective calibration and Bayesian updating for multiple TGVs can provide information about the quality of the models' mathematical descriptions with regard to the real-world situation. Regarding the crop modeling package XN, the GSA should also be repeated with other choices of sub-models to

enable more general conclusions about the predictive power of soil-crop simulations.

VI. Summary

Soil-crop models enjoy ever-greater popularity as tools to assess the impact of environmental changes or management strategies on agricultural production. Soil-crop models are designed to coherently simulate the crop, nitrogen (N) and water dynamics of agricultural fields. However, soil-crop models, with their sub-models for the crop, the N dynamic and the water regime, depend on a vast number of uncertain model inputs, i.e., initial conditions and parameters. To assess the uncertainty in the simulation results (UCSR) and how they can be apportioned among the model inputs of the XN-CERES soil-crop model, an uncertainty and global sensitivity analysis (GSA) was conducted. We applied two different GSA methods, moment-independent and variance-based methods in the sense of the Factor Prioritization and the Factor Fixing setting. The former identifies the key drivers of uncertainty, i.e., which model input, if fixed to its true value, would lead to the greatest reduction of the UCSR. The latter identifies the model inputs that cannot be fixed at any value within their value range without affecting the UCSR. In total we calculated six sensitivity indices (SIs).

The overall objective was to assess the cross-sub-model impact of parameters and the overall determinability of the XN-CERES applied on a deep loess soil profile in Southwest Germany. Therefore, we selected 39 parameters and 16 target variables (TGVs) from the four sub-models to be included in the GSA. Furthermore, we assessed a weekly time series of the parameter sensitivities. The sub-models were crop, water, nitrogen and flux. In addition, we also compared moment-independent

(MI) and variance-based (VB) GSA methods for their suitability for the two settings.

The results show that the parameters of the TGVs of the four groups cannot be considered independently. Each group is impacted by the parameters of the other groups. Crop parameters are most important, and they have an impact on each of the 16 TGVs. They are followed by the Mualem van Genuchten (MvG) parameters, of which *alpha* is by far the most important parameter. The nitrate (NO_3^-) content and the matric potential are the two TGVs that are most affected by the interaction of parameters, especially crop and MvG parameters. However, the model output of these two TGVs is highly skewed and leptokurtic. Therefore, the variance is an unsuitable representation of the UCSR, and the reliability of the variance-based sensitivity indices SI^{VB} is curtailed. Nitrogen group parameters play an overall minor role for the uncertainty of the whole XN-CERES, but nitrification rates can be calibrated on ammonium (NH_4^+) measurements. Considering the initial conditions shows the high importance of the initial NO_3^- content. If it could be fixed, the uncertainty of crop groups' TGVs, the matric potential and the N content in the soil could be reduced. Hence, multi-year predictions of yield suffer from uncertainty due to the simulated NO_3^- content.

Temporally resolved parameter sensitivities provide important insights into the model behavior. The big dependence between the crop's development stage and the other 15 TGVs becomes visible. High temporally resolved measurements of the development stage are important to univocally estimate the crop parameters and reduce the uncertainty in the vegetative and generative biomass. Furthermore, potential periods of water and N-limiting situations are assessed, which is helpful for deriving management strategies. In addition, it becomes clear that measurement campaigns should be conducted at the simulation start and during the vegetation period to have enough information to calibrate the XN-CERES.

Regarding the performance of the different GSA methods and the dif-

ferent SIs, we conclude that the sensitivity measure relying on the Kolmogorov-Smirnov metric (β^{ks}) is most stable. It converges quickly and has no issues with highly skewed and leptokurtic model output distributions. The assessments of the first-effect index and the β^{ks} provide information on the additivity of the model and parameters that cannot be fixed without impacting the simulation results.

In summary, we could only identify three parameters that have no direct impact on any TGV at any time and are hence not determinable from any measurements of the TGVs considered. Furthermore, we can conclude that the groups' parameters should not be calibrated independently because they always affect the uncertainty of the selected TGV directly or via interacting. However, no TGV is suitable to calibrate all parameters. Hence, the calibration of the XN-CERES requires measurements of TGVs from each group, even if the modeler is only interested in one specific TGV, e.g., yield.

The GSA should be repeated in a drier climate or with restricted rooting depth. The convergence of the values for the Sobol indices remains an issue. Even larger sample sizes, another convergence criteria or graphical inspection cannot alleviate the issue. However, we can conclude that the sub-models of the XN-CERES cannot be considered independently and that the model does what it is designed for: coherently simulating the crop, N and water dynamics with their interactions.

VII. Zusammenfassung

Boden-Pflanze Modelle erfreuen sich immer größerer Beliebtheit, um die Auswirkungen von Umweltveränderungen und Managementstrategien auf die landwirtschaftliche Produktion zu bestimmen. Boden-Pflanzen Modelle sind so konzipiert, dass sie kohärent die Pflanzen-, Stickstoff- (N) und Wasserdynamik in landwirtschaftlichen Feldern simulieren. Leider hängen Boden-Pflanze Modelle - mit ihren Teilmodellen für die Pflanze, die N Dynamik im Boden, der Evapotranspiration und dem Bodenwasserhaushalt - von einer Vielzahl unsicherer Modellinputs wie Anfangs- und Randbedingungen sowie Parametern ab. Zur Bestimmung der Unsicherheit in den Simulationsergebnissen (UCSR) und in welchem Ausmaß diese von den Modellinputs des Boden-Pflanze Modells XN-CERES abhängt, wird in dieser Arbeit eine Unsicherheits- und Global Sensitivitäts Analyse (GSA) durchgeführt. Wir verwendeten zwei verschiedene GSA-Methoden, momentunabhängige und varianzbasierte Methoden, im Sinne der Settings: Faktor Priorisierung und Faktor Fixing. Ersteres identifiziert die Parameter, die zur größten Reduktion der UCSR führen, wenn man deren richtigen Wert bestimmt. Letzteres identifiziert die Parameter, die nicht auf einen beliebigen Wert innerhalb ihres Wertebereichs fixiert werden können, ohne die UCSR zu beeinflussen. Insgesamt haben wir sechs verschiedene Sensitivitäts Indices (SIs) berechnet.

Das übergeordnete Ziel der Arbeit war es die Teilmodell-übergreifende Wirkung der Parameter und die allgemeinen Bestimmbarkeit des Boden-Pflanzen Modells XN-CERES auf einem Lössstandort in Südwest Deutsch-

land zu quantifizieren. Wir haben insgesamt 39 Parameter und 16 Zielvariablen (TGV) aus den vier Teilmodellen für die GSA ausgewählt. Darüber hinaus lösen wir die Parametersensitivitäten für die vier Teilm Modelle Pflanze, Wasser, Stickstoff und Flüsse wöchentlich auf. Darüber hinaus vergleichen wir Moment unabhängige (MI) und Varianz basierte (VB) GSA Methoden und ihre Eignung für die beiden Settings für ein Boden-Pflanze Model.

Die Ergebnisse zeigen, dass die Parameter der vier Gruppen im hohen Maße voneinander abhängen. Die TGV jeder Gruppe werden von gruppenfremden Parametern beeinflusst. An der Spitze stehen die Pflanzenparameter. Sie haben einen Einfluss auf jede der 16 TGVs. Es folgen die Mualem van Genuchten (MvG) Parameter, wobei α der mit Abstand wichtigste Parameter ist. Der Nitrat (NO_3^-) Gehalt und das Matrixpotential sind die beiden TGVs, die am stärksten von Parameterinteraktionen betroffen sind, insbesondere von Pflanzen- und MvG Parametern. Allerdings sind die Verteilungen dieser beiden TGVs schief und leptokurtisch. Daher ist die Varianz eine schlechte Representation für die UCSR und die Zuverlässigkeit der Varianz basierten Sensitivitätsindices (SI^{VB}) entsprechend eingeschränkt. Die Parameter der Stickstoffgruppe spielen insgesamt eine untergeordnete Rolle für die Unsicherheit des gesamten Modells. Die Nitrifikationsraten können allerdings anhand von Ammonium (NH_4^+) Messungen kalibriert werden. Die Betrachtung der Anfangsbedingungen zeigt, dass die Unsicherheit in der Simulation der TGVs der Pflanzengruppen, des Matrixpotentials und des N-Gehalts im Boden durch deren akurate Messung stark reduziert werden kann. Vorhersagen für Fruchtfolgen sind folglich unsicher, da der simulierte Ertrag der Hauptfrucht vom Zustand des Bodens nach der Vorfrucht abhängt.

Zeitaufgelöste Parametersensitivitäten liefern wichtige Erkenntnisse über das Modellverhalten. Die große Abhängigkeit zwischen dem Entwicklungsstadium der Pflanze und den andern 15 TGVs wird sichtbar. Hochauflösende Messungen des Entwicklungsstadiums der Pflanze sind wichtig, um die Pflanzenparameter eindeutig kalibrieren zu können und die

Unsicherheit in der Simulation von vegetativer und generativer Biomasse zu reduzieren. Ebenfalls können durch zeitaufgelöste Parametersensitivitäten, Zeiträume von möglicher Wasser- und N-Knappheit identifiziert werden. Dies ist besonders wichtig für die Erstellung von Managementstrategien. Darüber hinaus wird deutlich, dass Messungen vorrangig zu Simulationsbeginn und während der Vegetationsperiode durchgeführt werden sollten, um genügend Informationen für die Kalibrierung des Modells zu erhalten.

Bezüglich der Leistung der verschiedenen GSA Methoden und der unterschiedlichen SIs, kommen wir zu dem Ergebnis, dass das auf der Kolmogorov-Smirnov Metrik basierte Sensitivitätsmaß (β^{ks}) am stabilsten ist. Es konvergiert schnell und hat keine Probleme mit stark schiefen und leptokurtischen Verteilungen. Die Kombination aus First-Effect Index und β^{ks} gibt Aufschluss über die Additivität des Modells und identifiziert Parameter, die nicht fixiert werden können, ohne das Simulationsergebnis zu beeinflussen.

Zusammenfassend lässt sich sagen, dass wir nur drei Parameter identifizieren konnten, die keinen direkten Einfluss auf eine der untersuchten TGV haben und daher durch Messungen dieser TGVs nicht bestimmbar sind. Der direkte Einfluss weiterer acht Parameter ist so gering, dass deren Kalibrierung schwierig ist. Darüber hinaus kommen wir zu dem Schluss, dass die Parameter der verschiedenen Gruppen nicht unabhängig voneinander kalibriert werden können, da sie immer - direkt oder über Interaktion - die Unsicherheit der ausgewählten TGV beeinflussen. Allerdings ist nicht jede TGV zur Kalibrierung aller Parameter geeignet. Für die Kalibrierung der gewählten Modellkombination sind daher Messungen von TGVs jeder Gruppe erforderlich, auch wenn nur Interesse an einer bestimmten TGV wie zum Beispiel dem Ertrag besteht.

Aus der Arbeit ergeben sich einige generelle Empfehlungen. So sollte die GSA in einem trockeneren Klima oder mit eingeschränkter Durchwurzelungstiefe durchgeführt werden. Die Konvergenz der Werte für die Sobol-Indizes ist problematisch. Noch größere Stichprobengrößen, wei-

tere Konvergenzkriterien oder grafische Prüfungen könnten hier Abhilfe schaffen.

References

References

- AgMIP (2017). The agricultural model intercomparison and improvement project.
- Allen, R. G., Pereira, L. S., Dirks, R., and Smith, M. (1998). Crop evapotranspiration - guidelines for computing crop water requirements - fao irrigation and drainage paper no. 56. *FAO - Food and Agriculture Organization of the United Nations*.
- Anderson, T. W. and Darling, D. A. (1952). Asymptotic theory of certain goodness of fit criteria based on stochastic processes. *The Annals of Mathematical Statistics*, 23(2):193–212.
- Archer, G. E. B., Saltelli, A., and Sobol, I. M. (1997). Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58(2):99–120.
- Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B. A., Ottman, M. J., Wall, G. W., White, J. W., Reynolds, M. P., Alderman, P. D., Prasad, P. V. V., Aggarwal, P. K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., de Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., Hunt, L. A., Izaurrealde, R. C., Jabloun, M., Jones, C. D., Kersebaum, K. C., Koehler, A.-K., Müller, C., Naresh Kumar, S., Nendel, C., O’Leary, G., Olesen, J. E., Palosuo, T., Priesack, E.,

- Eyshi Rezaei, E., Ruane, A. C., Semenov, M. A., Shcherbak, I., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Thorburn, P. J., Waha, K., Wang, E., Wallach, D., Wolf, J., Zhao, Z., and Zhu, Y. (2014). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5(2):143–147.
- Asseng, S., Ewert, F., Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P. J., Rötter, R. P., Cammarano, D., Brisson, N., Basso, B., Martre, P., Aggarwal, P. K., Angulo, C., Bertuzzi, P., Biernath, C., Challinor, A. J., Doltra, J., Gayler, S., Goldberg, R., Grant, R., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurralde, R. C., Kersebaum, K. C., Müller, C., Naresh Kumar, S., Nendel, C., O’Leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stöckle, C., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., Wallach, D., White, J. W., Williams, J. R., and Wolf, J. (2013). Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, 3(9):827–832.
- Baucells, M. and Borgonovo, E. (2013). Invariant probabilistic sensitivity analysis. *Management Science*, 59(11):2536–2549.
- Baudin, M., Dufloy, A., Iooss, B., and Popelin, A.-L. (2015). Open turns: An industrial software for uncertainty quantification in simulation.
- Beven, K. and Freer, J. (2001). Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the glue methodology. *Journal of Hydrology*, 249(1-4):11–29.
- Borgonovo, E. (2006). Measuring uncertainty importance: Investigation and comparison of alternative approaches. *Risk analysis : an official publication of the Society for Risk Analysis*, 26(5):1349–1361.

- Borgonovo, E. (2007). A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784.
- Borgonovo, E. (2017). *Sensitivity Analysis: An Introduction for the Management Scientist*. International Series in Operations Research & Management Science. Springer.
- Borgonovo, E., Hazen, G. B., and Plischke, E. (2016). A common rationale for global sensitivity measures and their estimation. *Risk Analysis*, 36(10):1871–1895.
- Borgonovo, E., Lu, X., Plischke, E., Rakovec, O., and Hill, M. C. (2017). Making the most out of a hydrological model data set: Sensitivity analyses to open the model black-box. *Water Resources Research*, 53(9):7933–7950.
- Brilli, L., Bechini, L., Bindi, M., Carozzi, M., Cavalli, D., Conant, R., Dorich, C. D., Doro, L., Ehrhardt, F., Farina, R., Ferrise, R., Fitton, N., Francaviglia, R., Grace, P., Iocola, I., Klumpp, K., Léonard, J., Martin, R., Massad, R. S., Recous, S., Seddaiu, G., Sharp, J., Smith, P., Smith, W. N., Soussana, J.-F., and Bellocchi, G. (2017). Review and analysis of strengths and weaknesses of agro-ecosystem models for simulating c and n fluxes. *The Science of the total environment*, 598:445–470.
- Castaings, W., Borgonovo, E., Morris, M. D., and Tarantola, S. (2012). Sampling strategies in density-based sensitivity analysis. *Environmental Modelling & Software*, 38:13–26.
- Confalonieri, R. (2010). Monte carlo based sensitivity analysis of two crop simulators and considerations on model balance. *European Journal of Agronomy*, 33(2):89–93.
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., and Acutis, M. (2010a). Comparison of sensitivity analysis techniques: A case study with the rice model warm. *Ecological Modelling*, 221(16):1897–1906.

- Confalonieri, R., Bellocchi, G., Tarantola, S., Acutis, M., Donatelli, M., and Genovese, G. (2010b). Sensitivity analysis of the rice model warm in europe: Exploring the effects of different locations, climates and methods of analysis on model sensitivity to crop parameters. *Environmental Modelling & Software*, 25(4):479–488.
- Crnkovic, C. and Drachman, J. (1996). Quality control. *Risk*, 9(9):139–143.
- Cukier, R. I., Levine, H. B., and Shuler, K. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, 26:1–42.
- Damblin, G., Couplet, M., and Iooss, B. (2013). Numerical studies of space filling designs: optimization of latin hypercube samples and subprojection properties. *Journal of Simulation*.
- DeJonge, K. C., Ascough, J. C., Ahmadi, M., Andales, A. A., and Arabi, M. (2012). Global sensitivity and uncertainty analysis of a dynamic agroecosystem model under different irrigation treatments. *Ecological Modelling*, 231:113–125.
- Dumont, B., Basso, B., Bodson, B., Destain, J.-P., and Destain, M.-F. (2015). Climatic risk assessment to improve nitrogen fertilisation recommendations: A strategic crop model-based approach. *European Journal of Agronomy*, 65:10–17.
- Efron, B. and Stein, C. (1981). The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596.
- Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- European Commission (2009). Impact assessment guidelines.
- Gabriel, E., Fagg, G. E., Bosilca, G., Angskun, T., Dongarra, J. J., Squyres, J. M., Sahay, V., Kambadur, P., Barrett, B., Lumsdaine,

- A., Castain, R. H., Daniel, D. J., Graham, R. L., and Woodall, T. S. (2004). Open mpi: Goals, concept, and design of a next generation mpi implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, Budapest, Hungary.
- Gamboa, F., Janon, A., Klein, T., and Lagnoux, A. (2014). Sensitivity analysis for multidimensional and functional outputs. *Electronic Journal of Statistics*, 8(1):575–603.
- Gayler, S., Wöhling, T., Grzeschik, M., Ingwersen, J., Wizemann, H.-D., Warrach-Sagi, K., Högy, P., Attinger, S., Streck, T., and Wulfmeyer, V. (2014). Incorporating dynamic root growth enhances the performance of noah-mp at two contrasting winter wheat field sites. *Water Resources Research*, 50(2):1337–1356.
- Haefner and Hartmut (2019). Kit - scc - bwunicluster.
- Helton, J. C. and Davis, F. J. (2002). Illustration of sampling-based methods for uncertainty and sensitivity analysis. *Risk Analysis*, 22(3):591–622.
- Helton, J. C. and Davis, F. J. (2003). Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. *Reliability Engineering & System Safety*, 81(1):23–69.
- Helton, J. C., Johnson, J. D., Sallaberry, C. J., and Storlie, C. B. (2006). Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10-11):1175–1209.
- Herman, J. and Usher, W. (2017). Salib: An open-source python library for sensitivity analysis. *The Journal of Open Source Software*, 2(9):97.
- Homma, T. and Saltelli, A. (1996). Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17.

- Hora, S. C. and Iman, R. L. (1986). Comparison of maximum/bounding and bayes/monte carlo for fault tree uncertainty analysis.
- Hu, K., Li, Y., Chen, W., Chen, D., Wei, Y., Edis, R., Li, B., Huang, Y., and Zhang, Y. (2010). Modeling nitrate leaching and optimizing water and nitrogen management under irrigated maize in desert oases in northwestern china. *Journal of environmental quality*, 39(2):667–677.
- Iman, R. L. and Hora, S. C. (1990). A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Analysis*, 10(3):401–406.
- Ingwersen, J., Steffens, K., Högy, P., Warrach-Sagi, K., Zhunusbayeva, D., Poltoradnev, M., Gäbler, R., Wizemann, H.-D., Fangmeier, A., Wulfmeyer, V., and Streck, T. (2011). Comparison of noah simulations with eddy covariance and soil water measurements at a winter wheat stand. *Agricultural and Forest Meteorology*, 151(3):345–355.
- Jin, R., Chen, W., and Sudjianto, A. (2005). An efficient algorithm for constructing optimal design of computer experiments. *Journal of Statistical Planning and Inference*, 134:268–287.
- Johnsson, H., Bergstrom, L., Jansson, P.-E., and Paustian, K. (1987). Simulated nitrogen dynamics and losses in a layered agricultural soil. *Agriculture, Ecosystems & Environment*, 18(4):333–356.
- Khorashadi Zadeh, F., Nossent, J., Sarrazin, F., Pianosi, F., van Griensven, A., Wagener, T., and Bauwens, W. (2017). Comparison of variance-based and moment-independent global sensitivity analysis approaches by application to the swat model. *Environmental Modelling & Software*, 91:210–222.
- Kuiper, N. H. (1960). Tests concerning random points on a circle. *Indagationes Mathematicae (Proceedings)*, 63:38–47.

- Kutschera, L. and Lichtenegger, E. (1960). *Wurzelatlas mitteleuropaeischer Ackerun- kraeuter und Kulturpflanzen*, volume 1 of *Wurzelatlas*. DLG-Verl.-GMBH, Frankfurt am Main.
- Lamboni, M., Monod, H., and Makowski, D. (2011). Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, 96(4):450–459.
- Liang, H., Qi, Z., DeJonge, K. C., Hu, K., and Li, B. (2017). Global sensitivity and uncertainty analysis of nitrate leaching and crop yield simulation under different water and nitrogen management practices. *Computers and Electronics in Agriculture*, 142:201–210.
- Makowski, D., Naud, C., Jeuffroy, M.-H., Barbottin, A., and Monod, H. (2006). Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. *Reliability Engineering & System Safety*, 91(10-11):1142–1147.
- Mara, T. A., Belfort, B., Fontaine, V., and Younes, A. (2017). Addressing factors fixing setting from given data: A comparison of different methods. *Environmental Modelling & Software*, 87:29–38.
- Mara, T. A. and Tarantola, S. (2012). Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering & System Safety*, 107:115–121.
- Marrel, A., Saint-Geours, N., and de Lozzo, M. (2016). Sensitivity analysis of spatial and/or temporal phenomena. In Ghanem, R., Higdon, D., and Owhadi, H., editors, *Handbook of uncertainty quantification*, pages 1–31. Springer International Publishing, Switzerland.
- Martre, P., Wallach, D., Asseng, S., Ewert, F., Jones, J. W., Rotter, R. P., Boote, K. J., Ruane, A. C., Thorburn, P. J., Cammarano, D., Hatfield, J. L., Rosenzweig, C., Aggarwal, P. K., Angulo, C., Basso, B., Bertuzzi, P., Biernath, C., Brisson, N., Challinor, A. J.,

- Doltra, J., Gayler, S., Goldberg, R., Grant, R. F., Heng, L., Hooker, J., Hunt, L. A., Ingwersen, J., Izaurrealde, R. C., Kersebaum, K. C., Muller, C., Kumar, S. N., Nendel, C., O'leary, G., Olesen, J. E., Osborne, T. M., Palosuo, T., Priesack, E., Ripoche, D., Semenov, M. A., Shcherbak, I., Steduto, P., Stockle, C. O., Stratonovitch, P., Streck, T., Supit, I., Tao, F., Travasso, M., Waha, K., White, J. W., and Wolf, J. (2015). Multimodel ensembles of wheat growth: many models are better than one. *Global change biology*, 21(2):911–925.
- Mason, D. M. and Schuenemeyer, J. H. (1983). A modified kolmogorov-smirnov test sensitive to tail alternatives. *The Annals of Statistics*, 11:933–946.
- McKay, M. D., Beckman, R. J., and Conover, W. J. (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics*, 21(2):239–245.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092.
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33(2):161.
- NextFarming (2019). Next geodaten service — next farming.
- Nolan, B. T., Puckett, L. J., Ma, L., Green, C. T., Bayless, E. R., and Malone, R. W. (2010). Predicting unsaturated zone nitrogen mass balances in agricultural settings of the united states. *Journal of environmental quality*, 39(3):1051–1065.
- Nossent, J., Elsen, P., and Bauwens, W. (2011). Sobol' sensitivity analysis of a complex environmental model. *Environmental Modelling & Software*, 26(12):1515–1525.
- Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33:1065–1076.

- Pearson, K. (1895). Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58:240–242.
- Pianosi, F. and Wagener, T. (2015). A simple and efficient method for global sensitivity analysis based on cumulative distribution functions. *Environmental Modelling & Software*, 67:1–11.
- Plauborg, F., Manevski, K., Zhou, Z., and Andersen, M. N. (2015). The use of computer simulation models in precision nutrient management. In Stafford, J. V., editor, *Precision agriculture '15*. Wageningen Academic Publishers, The Netherlands.
- Plischke, E. (2012). An adaptive correlation ratio method using the cumulative sum of the reordered output. *Reliability Engineering & System Safety*, 107:149–156.
- Plischke, E., Borgonovo, E., and Smith, C. L. (2013). Global sensitivity measures from given data. *European Journal of Operational Research*, 226(3):536–550.
- Priesack, E. (2006). *Expert-N Dokumentation der Modellbibliothek: Zugl.: Göttingen, Univ., Habil.-Schr., 2006*, volume 60 of *FAM-Bericht*. Hieronymus, München.
- Ratto, M., Pagano, A., and Young, P. C. (2009). Non-parametric estimation of conditional moments for sensitivity analysis. *Reliability Engineering & System Safety*, 94(2):237–243.
- Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Physics*, 1(5):318–333.
- Richter, G. M., Acutis, M., Trevisiol, P., Latiri, K., and Confalonieri, R. (2010). Sensitivity analysis for a complex crop model applied to durum wheat in the mediterranean. *European Journal of Agronomy*, 32(2):127–136.
- Ritchie, J. and Godwin, D. (1987). Ceres wheat 2.0.

- Rosenzweig, C., Jones, J. W., Hatfield, J. L., Ruane, A. C., Boote, K. J., Thorburn, P., Antle, J. M., Nelson, G. C., Porter, C., Janssen, S., Asseng, S., Basso, B., Ewert, F., Wallach, D., Baigorría, G., and Winter, J. M. (2013). The agricultural model intercomparison and improvement project (agmip): Protocols and pilot studies. *Agricultural and Forest Meteorology*, 170:166–182.
- Ruget, F., Brisson, N., Delécolle, R., and Faivre, R. (2002). Sensitivity analysis of a crop simulation model, stics, in order to choose the main parameters to be estimated. *Agronomie*, 22(2):133–158.
- Saltelli, A. (2002a). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145(2):280–297.
- Saltelli, A. (2002b). Sensitivity analysis for importance assessment. *Risk Analysis*, 22(3):579–590.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., and Gatelli, D. (2008). *Global Sensitivity Analysis: The Primer*. Wiley-Interscience, s.l., 1. Aufl. edition.
- Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M. (2004). *Sensitivity Analysis in Practice : A Guide to Assessing Scientific Models*. John Wiley & Sons, Ltd.
- Sarrazin, F., Pianosi, F., and Wagener, T. (2016). Global sensitivity analysis of environmental models: Convergence and validation. *Environmental Modelling & Software*, 79:135–152.
- Schöniger, A., Wöhling, T., Samaniego, L., and Nowak, W. (2014). Model selection on solid ground: Rigorous comparison of nine ways to evaluate bayesian model evidence. *Water Resources Research*, 50(12):9484–9513.
- Sexton, J., Everingham, Y. L., and 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.
- Sheather, S. J. (2004). Density estimation. *Statistical Science*, 19(4):588–597.
- Shin, M.-J., Guillaume, J. H., Croke, B. F., and Jakeman, A. J. (2013). Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in r. *Journal of Hydrology*, 503:135–152.
- Silverman, B. W. (1998). *Density Estimation for Statistics and Data Analysis*. CHAPMAN & HALL/CRC, Boca Raton London New York Washington, D.C., 1 edition.
- Simunek, J., K. Huang, and M. van Genuchten (1998). The hydrus code for simulating the one-dimensional movement of water, heat, and multiple solutes in variables-saturated media. *Tech. Rep. 144, U.S. Salinity Lab., United States Dep. of Agriculture, Agricultural Research Service*, (version 6).
- Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling and Computational Experiment*, 1:407–414.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology*, 15(1):72–101.
- Specka, X., Nendel, C., and Wieland, R. (2015). Analysing the parameter sensitivity of the agro-ecosystem model monica for different crops. *European Journal of Agronomy*, 71:73–87.
- Stahn, P., Busch, S., Salzmann, T., Eichler-Löbermann, B., and Miegel, K. (2017). Combining global sensitivity analysis and multiobjective optimisation to estimate soil hydraulic properties and representations

- of various sole and mixed crops for the agro-hydrological swap model. *Environmental Earth Sciences*, 76(10):26.
- Strong, M., Oakley, J. E., and Chilcott, J. (2012). Managing structural uncertainty in health economic decision models: A discrepancy approach. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 61(1):25–45.
- Tan, J., Cui, Y., and Luo, Y. (2016). Global sensitivity analysis of outputs over rice-growth process in oryza model. *Environmental Modelling & Software*, 83:36–46.
- Tan, J., Cui, Y., and Luo, Y. (2017). Assessment of uncertainty and sensitivity analyses for oryza model under different ranges of parameter variation. *European Journal of Agronomy*, 91:54–62.
- Tarantola, S., Giglioli, N., Jesinghaus, J., and Saltelli, A. (2002). Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making? *Stochastic Environmental Research and Risk Assessment (SERRA)*, 16(1):63–76.
- US EPA (2009). Guidance on the development, evaluation, and application of environmental models.
- van Genuchten, M. T. (1980). A closed-form equation for predicting the hydraulic conductivity of unsaturated soils1. *Soil Science Society of America Journal*, 44(5):892.
- Vanuytrecht, E., Raes, D., and Willems, P. (2014). Global sensitivity analysis of yield output from the water productivity model. *Environmental Modelling & Software*, 51:323–332.
- Wallach, D. and Thorburn, P. J. (2017). Estimating uncertainty in crop model predictions: Current situation and future prospects. *European Journal of Agronomy*, 88:A1–A7.
- Wallor, E., Kersebaum, K.-C., Ventrella, D., Bindi, M., Cammarano, D., Coucheney, E., Gaiser, T., Garofalo, P., Giglio, L., Giola, P.,

- Hoffmann, M. P., Iocola, I., Lana, M., Lewan, E., Maharjan, G. R., Moriondo, M., Mula, L., Nendel, C., Pohankova, E., Roggero, P. P., Trnka, M., and Trombi, G. (2018). The response of process-based agro-ecosystem models to within-field variability in site conditions. *Field Crops Research*, 228:1–19.
- Wang, J., Li, X., Lu, L., and Fang, F. (2013). Parameter sensitivity analysis of crop growth models based on the extended fourier amplitude sensitivity test method. *Environmental Modelling & Software*, 48:171–182.
- Wang, Z., Qi, Z., Xue, L., and Bukovsky, M. (2016). Rzwqm2 simulated management practices to mitigate climate change impacts on nitrogen losses and corn production. *Environmental Modelling & Software*, 84:99–111.
- Wei, P., Lu, Z., and Song, J. (2014). Moment-independent sensitivity analysis using copula. *Risk analysis : an official publication of the Society for Risk Analysis*, 34(2):210–222.
- Wizemann, H.-D., Ingwersen, J., Högy, P., Warrach-Sagi, K., Streck, T., and Wulfmeyer, V. (2015). Three year observations of water vapor and energy fluxes over agricultural crops in two regional climates of southwest germany. *Meteorologische Zeitschrift*, 24(1):39–59.
- Wöhling, T., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H.-D., Högy, P., Cuntz, M., Attinger, S., Wulfmeyer, V., and Streck, T. (2013). Multiresponse, multiobjective calibration as a diagnostic tool to compare accuracy and structural limitations of five coupled soil-plant models and ckm3.5. *Water Resources Research*, 49(12):8200–8221.
- XING, H.-m., XU, X.-g., LI, Z.-h., CHEN, Y.-j., FENG, H.-k., YANG, G.-j., and CHEN, Z.-x. (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(11):2444–2458.

Xu, C. and Gertner, G. (2007). Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics & Data Analysis*, 51(12):5579–5590.

Zhao, G., Bryan, B. A., and Song, X. (2014). Sensitivity and uncertainty analysis of the apsim-wheat model: Interactions between cultivar, environmental, and management parameters. *Ecological Modelling*, 279:1–11.

Appendix

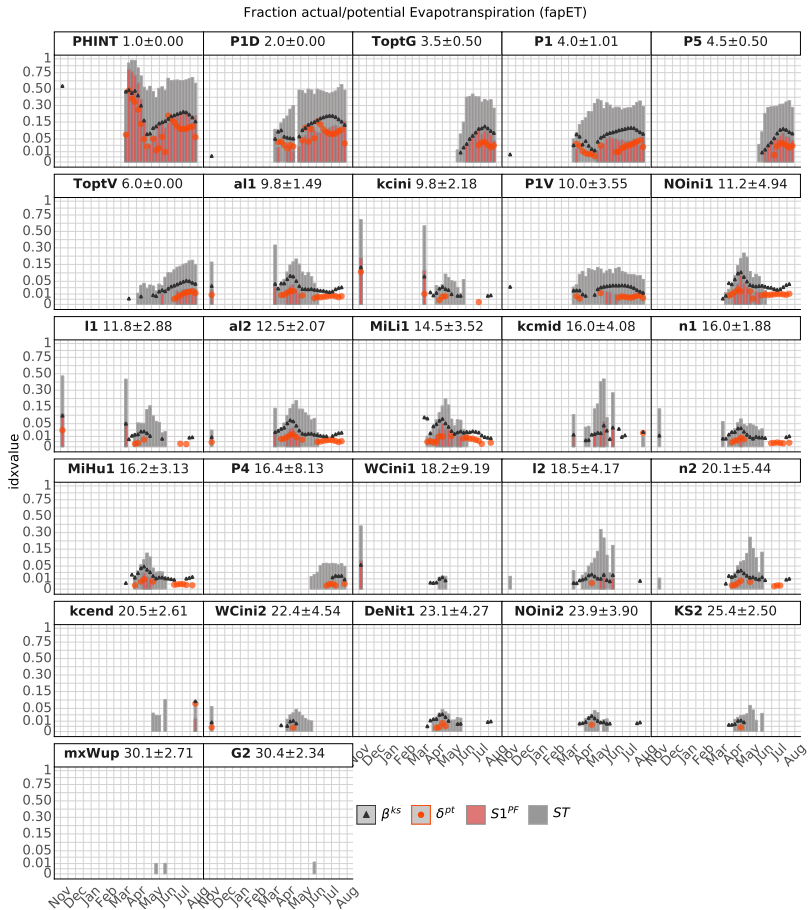


Figure 7.1: Time series of sensitivity indices $S1^{PF}$, ST , δ^{pt} and β^{ks} for the ratio of daily actual Evapotranspiration to daily potential Evapotranspiration ($fapET$).

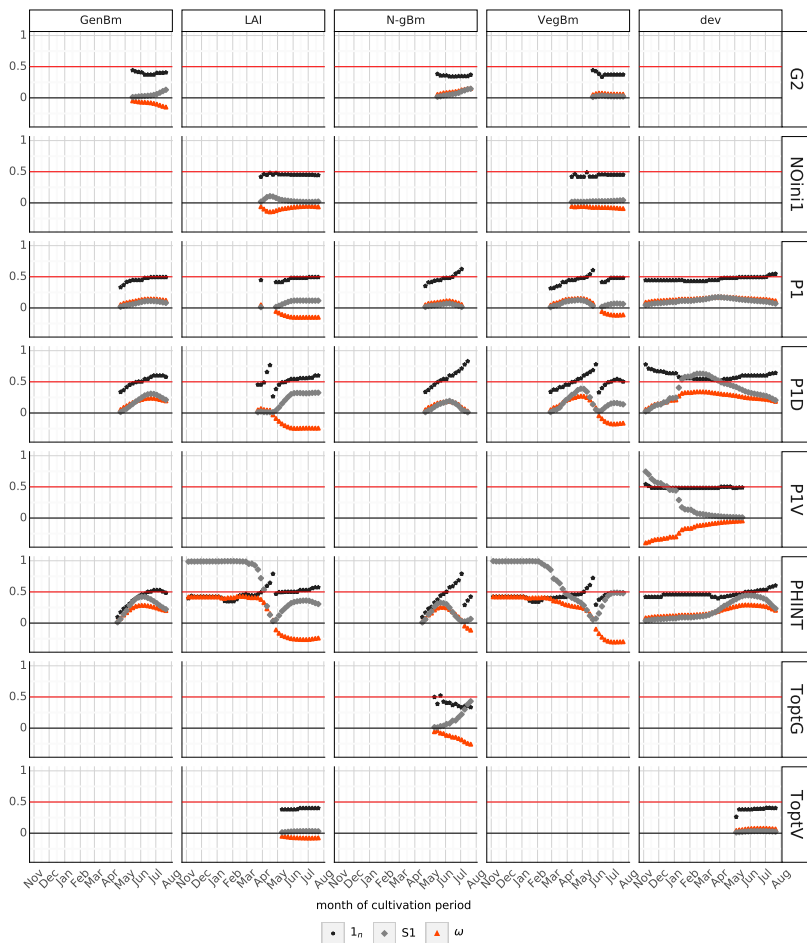


Figure 7.2: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the plant group for all parameters that have a first-order effect.

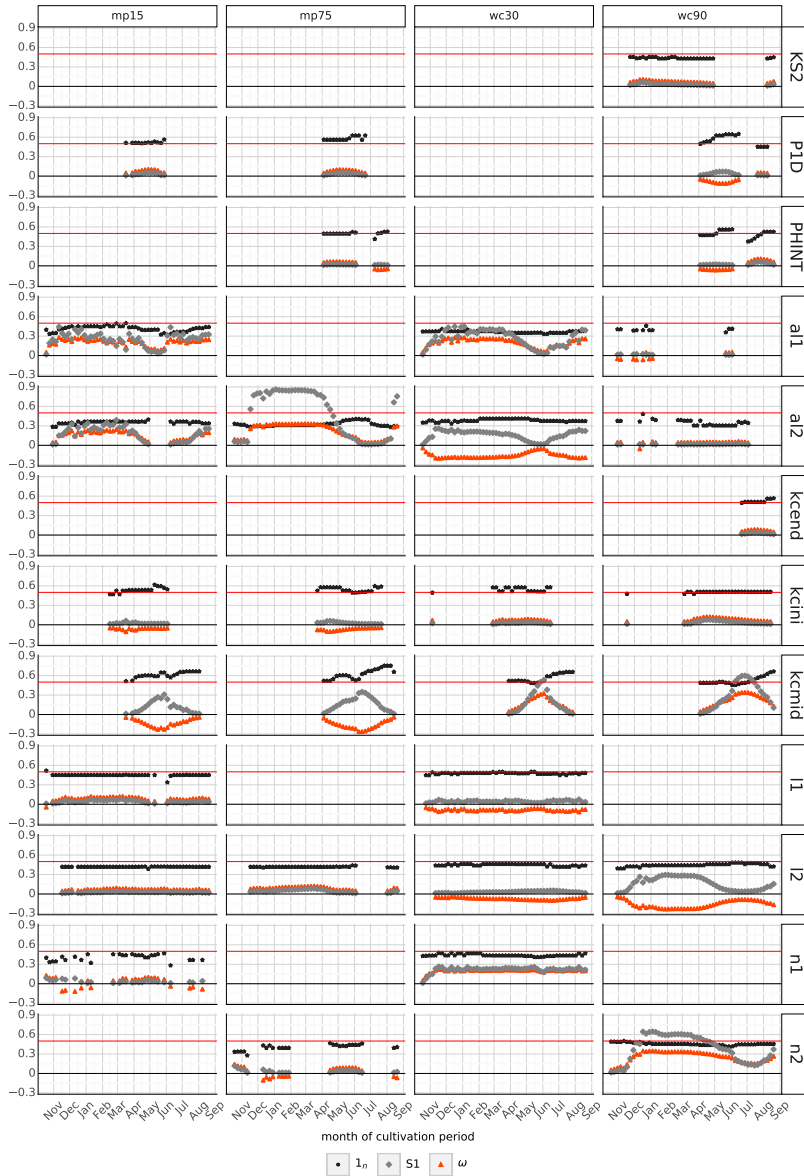


Figure 7.3: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the water group for all parameters that have a first-order effect.

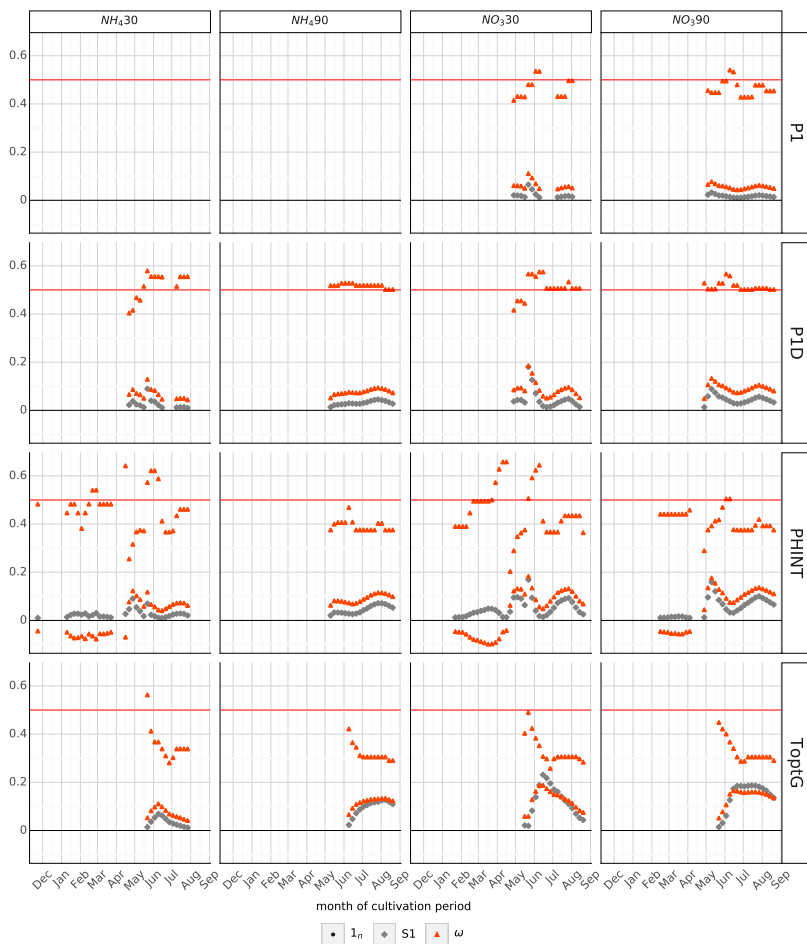


Figure 7.4: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for the plant group's parameter that have a first-order effect.

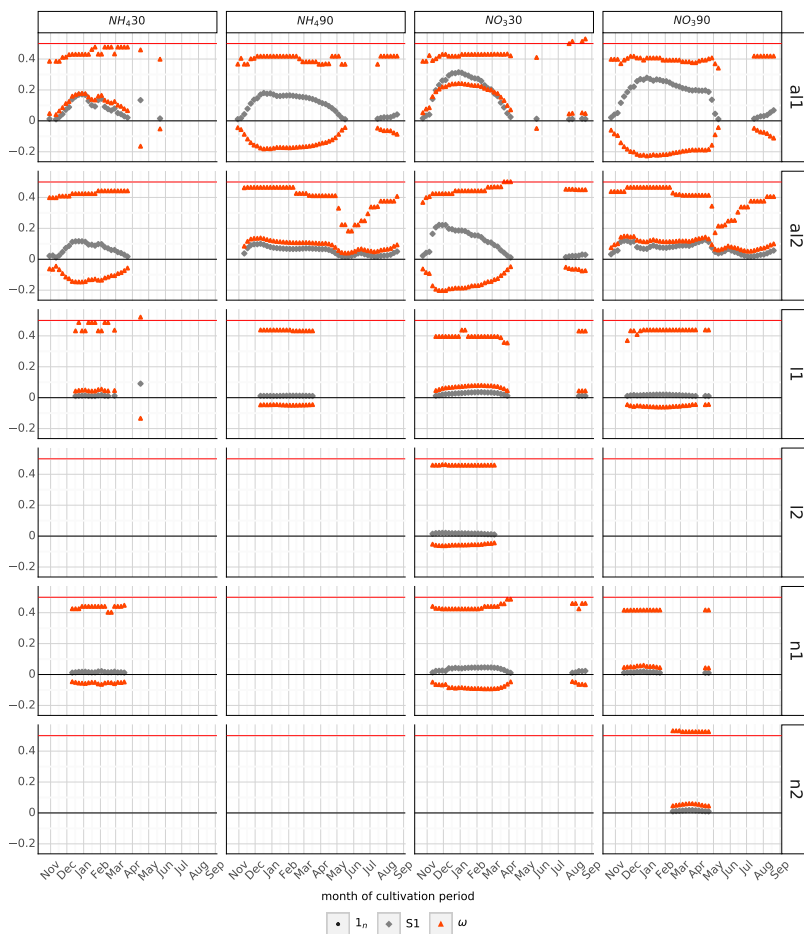


Figure 7.5: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for for the water group's parameter that have a first-order effect.

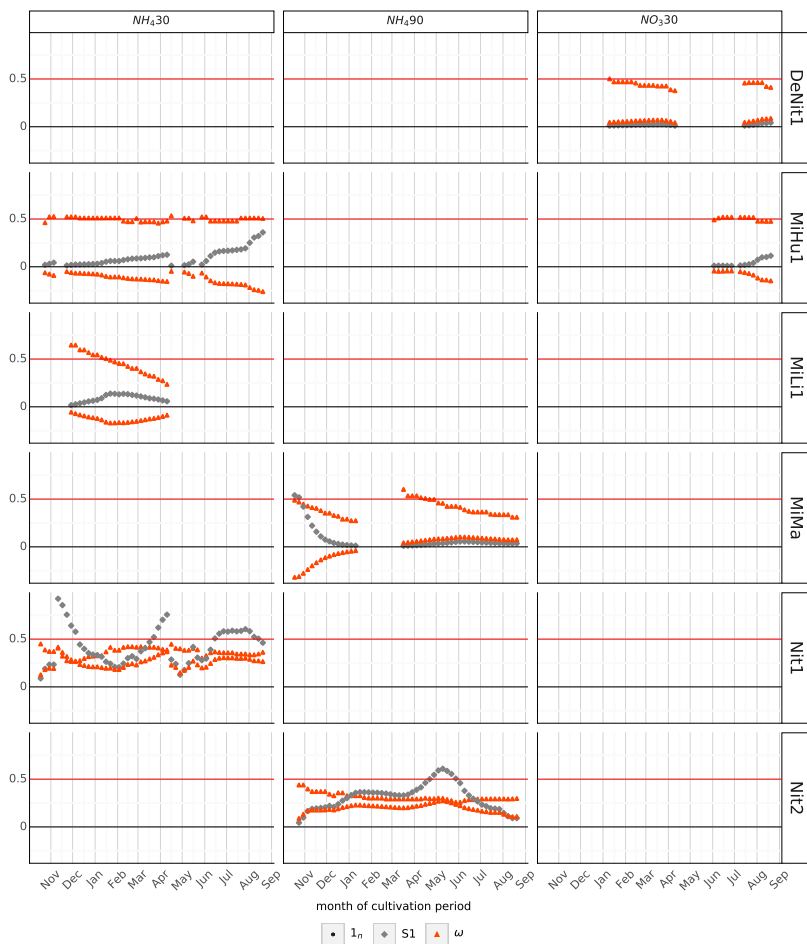


Figure 7.6: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for the nitrogen group's parameter that have a first-order effect.

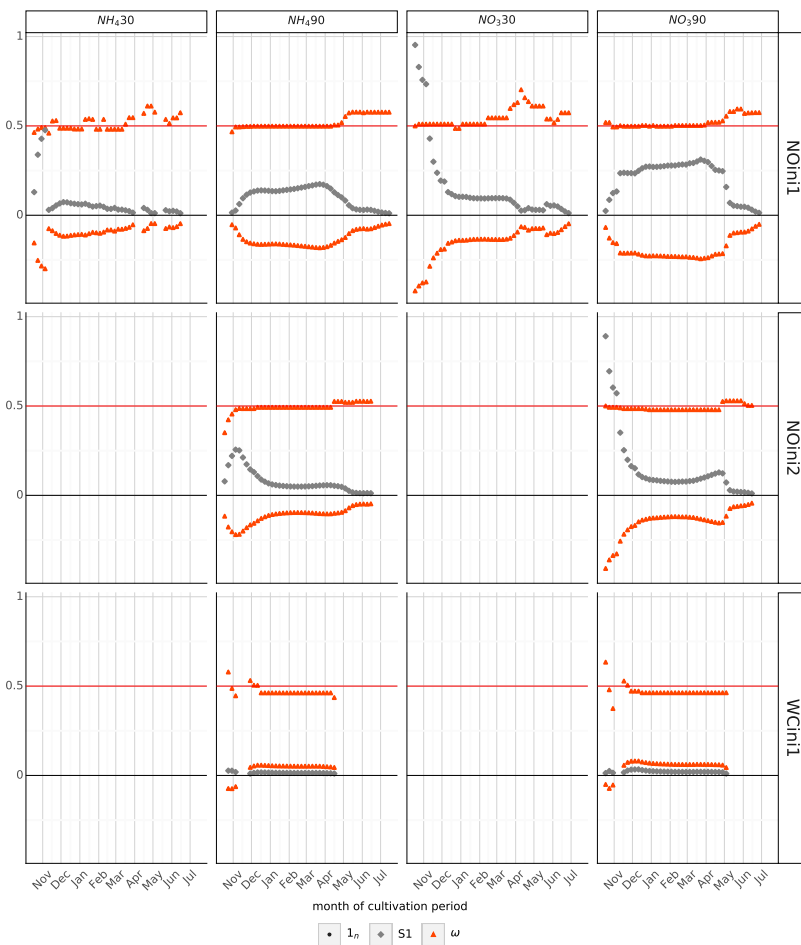


Figure 7.7: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the nitrogen group for the ini group's parameter that have a first-order effect.

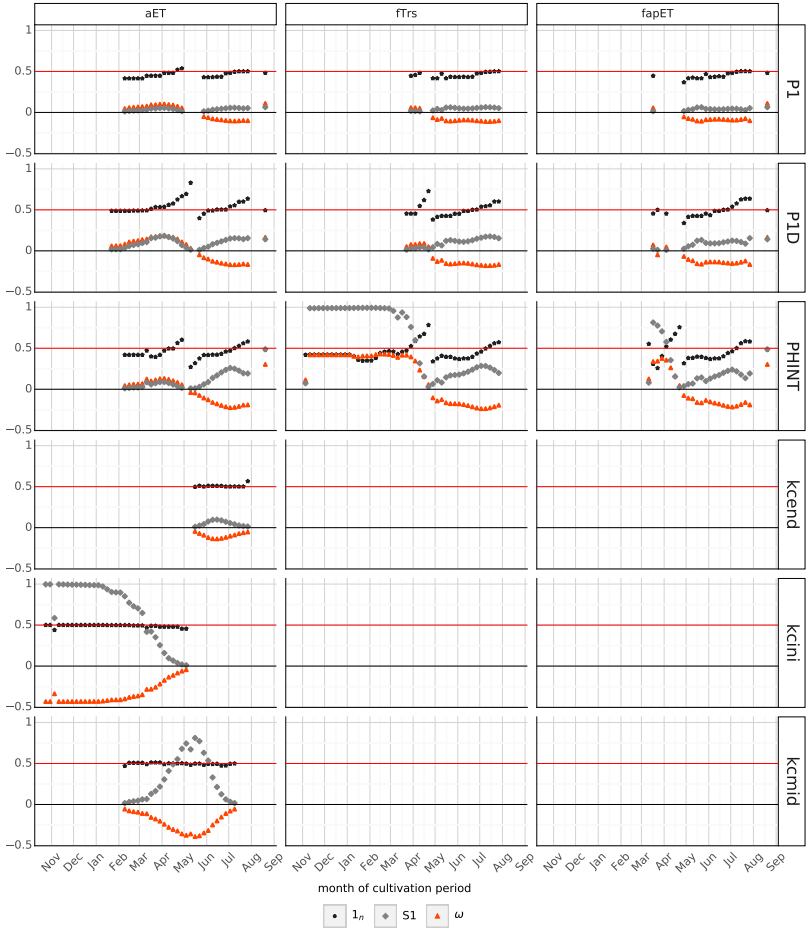


Figure 7.8: Time series of the values extracted from the CUSUNORO curve, i.e., 1_n and ω for the target variables of the flux group for all parameters that have a first-order effect.

Curriculum Vitae

Personal Data

Name: Irene Witte
City: Karlsruhe
Date and place
of birth: 31.01.1990
Aurich

Education

06.2014 Master of Science in Agriculture, Soil Science
University of Hohenheim, Germany
10.2011 Bachelor of Science in Agriculture
University of Hohenheim, Germany
Major: Soil Science
06.2008 Abitur, Gymnasium Ulricianum Aurich

Working Experience

since 15.08.19 Sr. Data Science Consultant, Blue Yonder
07.2018 – Research assistant, University of Hohenheim
12.2018 Biogeophysics
10.2014 – Research assistant, University of Hohenheim
05.2018 Institute of Soil Science and Land Evaluation
04.2012 – Student assistant, University of Hohenheim
01.2014 Institute of Soil Science and Land Evaluation
06.2012/13 Leader of single-day excursions into
the Black Forest
06.2010 – Student assistant in the project
12.2012 “Entwicklung erosionsmindernder
Anbauverfahren im Feldgemüsebau”,
University of Hohenheim
Institute of Crop Science and
of Soil Science and Land Evaluation
08 - 10.2009 Internship at “Spezialberatungsring für
Rinderhaltung und Futterbau e.V.”, Aurich