

Improving the simulation of permanent grasslands across Germany by using multi-objective uncertainty-based calibration of plant-water dynamics

Bahareh Kamali^{1,2}, Tommaso Stella¹, Michael Berg-Mohnicke¹, Jürgen Pickert¹, Jannis Groh^{1,3}, Claas Nendel^{1,4,5}

¹ Leibniz Centre for Agricultural Landscape Research (ZALF), Eberswalder Str. 84, 15374 Müncheberg, Germany

² Institute for Crop Science and Resource Conservation (INRES), University of Bonn, Katzenburgweg 5, 53115 Bonn, Germany

³ Forschungszentrum Jülich, Agrosphere, Institute of Bio- and Geoscience IBG-3, Wilhelm-Johnen-Straße 52428 Jülich, Germany

⁴ Institute of Biochemistry and Biology, University of Potsdam, Am Mühlenberg 3, 14476 Potsdam, Germany

⁵ Global Change Research Institute, Academy of Sciences of the Czech Republic, Bělidla 986/4a, 603 00 Brno, Czech Republic

Corresponding author

Bahareh Kamali

University of Bonn (UNIBONN)

Agricultural faculty, the Institute of Crop Science and Resource Conservation (INRES)

Add: Katzenburgweg 5, 53115 Bonn

Tel: +49 (0) 228 73 3258

E-mail: bahareh.kamali@zalf.de and bkamali@uni-bonn.de

Abstract

The dynamics of grassland ecosystems are highly complex due to multifaceted interactions among their soil, water, and vegetation components. Precise simulations of grassland productivity therefore rely on accurately estimating a variety of parameters that characterize different processes of these systems. This study applied three calibration schemes – a Single-Objective (SO-SUFI2), a Multi-Objective Pareto (MO-Pareto), and, a novel Uncertainty-Based Multi-Objective (MO-SUFI2)– to estimate the parameters of MONICA (Model for Nitrogen and Carbon Simulation) agro-ecosystem model in grassland ecosystems across Germany. The MO-Pareto model is based on a traditional Pareto optimality concept, while the MO-SUFI2 optimizes multiple target variables considering their level of prediction uncertainty. We used measurements of leaf area index, aboveground biomass, and soil moisture from experimental data at five sites with different intensities of cutting regimes (from two to five cutting events per season) to evaluate model performance. Both MO-Pareto and MO-SUFI2 outperformed SO-SUFI2 during calibration and validation. The comparison of the two MO approaches shows that they do not necessarily conflict with each other, but MO-SUFI2 provides complementary information for better estimations of model parameter uncertainty. We used the obtained parameter ranges to simulate grassland productivity across Germany under different cutting regimes and quantified the uncertainty associated with estimated productivity across regions. The results showed higher uncertainty in intensively managed grasslands compared to extensively managed grasslands, partially due to a lack of high-resolution input information concerning cutting dates. Furthermore, the additional information on the quantified uncertainty provided by our proposed MO-SUFI2 method adds deeper insights on confidence levels of estimated productivity. Benefiting from additional management data collected at high resolution and ground measurements on the composition of grassland species mixtures appear to be promising solutions to reduce uncertainty and increase model reliability.

Keywords: Intensively managed grasslands; extensively managed grasslands; Grassland productivity; Pareto optimality.

1. Introduction

Grasslands occupy 40.5% of the world's total terrestrial area (White et al., 2000) and contribute greatly to terrestrial biodiversity and carbon storage, which accounts for approximately 28–37% of the global soil organic carbon pool (Lal, 2004). A recent study in tree-sparse grasslands such as California has demonstrated that grasslands are more reliable carbon sinks than forest and wetland ecosystems with respect to climate change (Dass et al., 2018). Economically less profitable than arable crops (Wimberly et al., 2017), permanent grasslands are usually cultivated where arable cropping is limited by, e.g. steep slopes, the risk of frequent flooding, adverse climatic or hydraulic conditions (Lei et al., 2016), and poor soils. In Germany, managed grassland areas cover more than 28% of agricultural land (Griffiths et al., 2020; Statistisches Bundesamt, 2019) and represent remarkable values for fodder production, biodiversity conservation, recreation, and ecosystem services. In fact, about 50% of plant species in Germany depend on grassland habitats (BFN, 2014). Grasslands also contribute to the replenishment of groundwater resources, as they allow more water to percolate into deeper soil zones compared to forests or cropland (von Wilpert et al., 2016). Permanent grasslands (in the form of meadows or pastures), which comprise more than 90% of Germany's grassland areas, have been under pressure from land use changes since at least 1990. Since then, the total area has been reduced by 12% (Statistisches Bundesamt, 2019) as a result of land use changes in favor of intensive agriculture. At the same time, low-productivity grasslands on marginal sites have been abandoned for economic reasons, threatening some rare plant communities (BFN, 2014). The various types of grassland in Germany nevertheless still provide a large range of ecosystem services and biodiversity, if managed appropriately. Therefore, preservation of grassland as an important part of agricultural land use in Germany remains of crucial importance, and a better understanding of their environmental functioning is required to manage these areas more sustainably, especially in the context of global warming.

The dynamics of grassland ecosystems are highly complex, due to the multifaceted interactions among their soil, water, and vegetation components. Equipped with understanding of these interrelations, it is easier to design grassland management strategies that do not interfere with the provision of ecosystem services. Over the last few decades, various process-based agro-ecosystem models have been developed to simulate plant–soil–water relations and to explore the productivity and functioning of grassland systems. Such model application studies have focused on different aspects, such as the spatial distribution of grassland productivity (Chang et al., 2015b; Zheng et al., 2020), the impacts of environmental factors and climate change on grassland productivity (Gomara et al., 2020; Graux et al., 2011), and their potential as sinks for carbon storage and the mitigation of greenhouse gas emissions (Chang et al., 2015a; Jones and Donnelly, 2004; Sandor et al., 2018). In these studies, the most commonly simulated variables are aboveground biomass (*AGB*) or net primary productivity, providing a synthetic indicator to evaluate the productivity of grassland and the applied models were calibrated and validated based on vegetative variables, such as *AGB*.

however, vegetative variables are tightly linked to water dynamics in soils constraining ecosystem productivity (Archontoulis et al., 2020; Tang et al., 2018), and thus a model calibrated using solely vegetation variables may not be sufficient to predict soil-related services or disservices and may result in unrealistic representations of soil processes. Including soil moisture as a constraint to plant growth in the calibration process is critical to add explanatory power to the plant-related target variables. The main difficulty in this context is the interdependency of soil water and vegetation dynamics and the simultaneous consideration of parameters and variables related to both vegetation and soil processes in the calibration. Multi-objective approaches help the model to simulate various observation data concurrently and close to reality (Houska et al., 2017), since, in contrast to single criterion calibration procedures, they also take into account the existence of compensating effects when calibrating the model. Considering multiple variables combined with global optimization routines/algorithms in calibration decreases the risk for the model getting trapped in local minima, which may lead to a better fit for one variable, but to an unsuitable parameter set representing other processes (i.e. right fit for wrong reason). To account for this feedback, it is crucial to implement methods for model calibration capable of taking into account information that describes the interconnected processes. In most studies with integrative aspects, different observation variables are combined as a single objective (SO) which subsequently forms a basis to estimate the model's parameters. The multi-objective (MO) optimization perspective (Groh et al., 2018; Wöhling et al., 2015), which examines the trade-offs between different conflicting objectives, has received less attention in the application of grassland models, and to the best of our knowledge, we offer the first study of an integrative perspective in grassland simulations.

Process-based agro-ecosystem modeling is based on the simultaneous simulation of different interconnected biophysical processes, which depends on various parameters describing behaviors of various conceptual processes and the way they relate to each other (Fenicia et al., 2007). Some model parameters are extremely difficult to measure or even cannot be directly measured and often have to be estimated inversely, through an optimization procedure. Such procedures are intended to minimize deviations between simulated and observed target variables, such as *AGB*, soil moisture (*SM*), or leaf nitrogen concentration. Since the processes in agro-ecosystem models are closely interlinked, these minima are often only local, and optimization procedures may reveal only one out of many existing minima, but not necessarily the global minimum (or maximum). This means that the parameter values identified are associated with high levels of uncertainty, which in turn may lead to poor model predictions. Reducing the uncertainty of these parameters requires methods which consider A) multiple components of the model in an MO calibration framework and B) quantification of the confidence level of output variables.

MO optimization approaches use tradeoffs to determine a set of non-dominated parameters that cannot be improved for one objective without compromising the other objective. During optimization, these approaches incorporate multiple objectives on the basis of different information such as: 1) multi-variable

data representing different interrelated processes, 2) multi-site data, or 3) multi-response models considering independent criteria of one model aspect (Efstratiadis and Koutsoyiannis, 2010; Kamali et al., 2013). Therefore, instead of converging around a single optimum, MO approaches spread search in parameter space in a way to detect a number of feasible parameter sets (solutions) with acceptable trade-offs along the Pareto front (Proximedia, October 8, 2018). Given the fact that different parameters activate different processes in the model (Efstratiadis and Koutsoyiannis, 2010), exploring the trade-off between different objectives (considering various variables) assists models in finding a more realistic and robust parameter estimates. Furthermore, evaluating the trade-off among solutions located on the Pareto fronts provides additional indications on probable limitations of a model (Efstratiadis and Koutsoyiannis, 2010). For example, an irregular shape of a Pareto-front can be a sign for an ill-posed model, or a significant trade-off points to a probable ill model parameterization. A proper evaluation of these additional information obtained from MO approaches help modelers to better assessment of model performance and consistency and find robust solutions to reduce the uncertainty in the model.

The most common algorithms for MO calibration include MO particle swarm optimization (Kennedy and Eberhart, 1995), MO genetic algorithm (Fonseca and Fleming, 1993), and MO complex evolution (Yapo et al., 1998). While most MO methods offer superior performance compared to SO calibration (Kamali et al., 2013), they also suffer from their inability to provide information on the uncertainty of model predictions. Bayesian approaches can account for parameter uncertainty in optimization, and their superior performance vis-à-vis SO approaches has recently been demonstrated for eco-hydrological models (Tang et al., 2018; Wöhling et al., 2013) where mostly two variables (*LAI* and *SM*) have been considered. However, their application for calibration of grassland models have been limited mostly to SO approaches where the errors from different variables were aggregated to overall one single function (Höglind et al., 2016; Korhonen et al., 2018). The complex dynamics of grasslands, strongly vary depending on the types of their management (intensively and extensively managed grasslands), species composition and prevailing soil. Therefore, we assume that model calibration will benefit a lot from considering multiple key variables e.g. *LAI* as a model state variable, *AGB* and *SM*. Despite the importance, the current studies lack reflection on uncertainty-based calibration and parameter estimation approaches which capitalize on: 1) information that describes the multiple variables and find the trade-off among different variables; 2) the uncertainty ranges of each variable; and 3) the comparison of these approaches on intensively and extensively managed grasslands.

In this paper, we aim to 1) implement an uncertainty-based MO calibration procedure to a process-based models simulation model for grassland (Nendel et al., 2011) to understand how it helps to improve the simulation of different soil and vegetation processes; 2) analyze the implications (scaling/uncertainty) derived from the simulation of managed grasslands in Germany, and use that information to quantify the level of uncertainty associated with biomass estimates across different regions of Germany. The method proposed and applied accounts for all sources of uncertainty, such as model input, model structure, model

parameters, and measured data, and does not disentangle different types of uncertainty. The suitability of the method is tested against data from grassland sites with two/three cutting events per growing season (extensively managed) and sites with four/five cutting events per growing season (intensively managed). Overall, five experimental sites located in different parts of Germany were selected. In this paper, we maintain our assumption that grassland management (i.e. intensively and extensively managed grasslands) relates directly to only the number of cutting events.

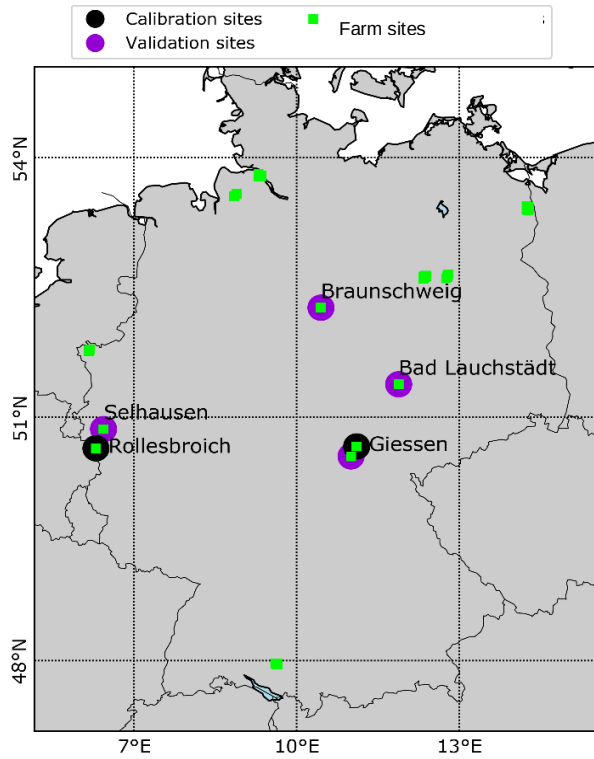
2. Methodology

2.1. Data sources

Two data sets were used in this study. For the sensitivity analysis, model calibration and model validation, we used data from five experimental grassland sites that have been originally designed for different projects (see Table 1). The second data set includes national-scale soil and climate data for German-wide simulations of grasslands (see section 2.4 for details).

The five experimental sites are: 1) Braunschweig (to understand the increasing importance of grassland areas for sustainable agricultural practice), 2) Rollesbroich and 3) Selhausen (long-term observation from lysimeters to investigate hydrosphere, biosphere and atmosphere interactions), 4) Bad Lauchstädt (to investigate the consequences of future climate scenarios for ecosystem functioning in different land-use types on large field plots), and 5) Gießen (a free-air carbon dioxide enrichment experiment to investigate the effects of multiple environmental factors on C3 plants, including the CO₂ fertilization effect). These data had all been collected within different research projects. The temporal resolution of the data varied depending on the project, variable, and year. For *SM*, the temporal resolution of data were daily or weekly most of the time, whereas the temporal resolution of *LAI* and *AGB* varied from weekly to bi-monthly within growing season (Table 1). The sites are located in different parts of Germany (Fig. 1a), featuring different climatic conditions, cutting regimes, and soil types (Table 1). The climate data used for these experimental sites were obtained from the closest station from the German Weather Service (Deutscher Wetterdienst, DWD) to each site. The average yearly precipitation at these sites over 30 years during 1990–2020 varied from 494 mm yr⁻¹ in Bad Lauchstädt to 910 mm yr⁻¹ in Rollesbroich. The average annual air temperature ranged from 8.75 °C in Rollesbroich to 10.6 °C in Selhausen.

a) Experimental sites used for model calibration/validation



b) Natural unites used for German-wide simulations

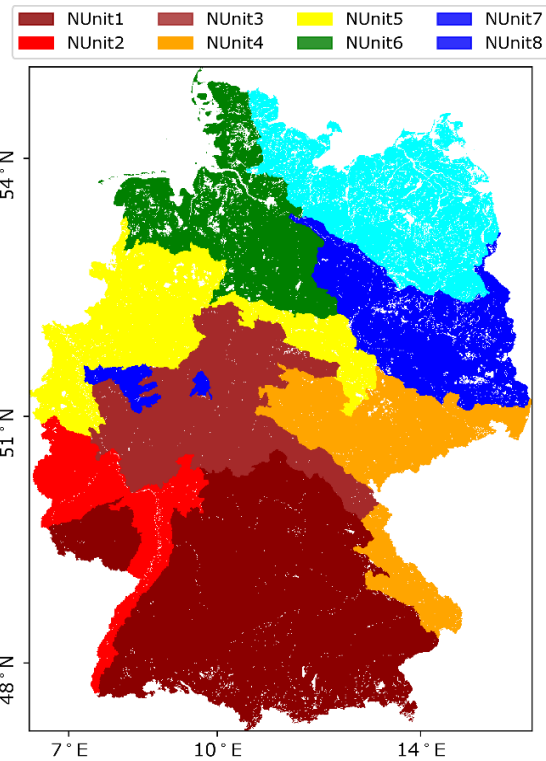


Fig. 1: a) Geographic location of the five experimental grassland sites selected for model calibration and validation and farm sites used for defining management for national-scale simulations; b) the eight natural units for Germany-wide simulation (NUnits 1-8) specifying the different landscape types across Germany that were used to compare grassland productivity at the regional scale.

163 **Table 1:** Location and characteristics of grassland study sites. The numbers in brackets show the number of plots for which measurement data were
164 available. *LAI*: leaf area index; *AGB*: above ground biomass; *SM*: soil moisture. The average yearly precipitation and temperature were obtained from
165 DWD climate stations over 30 years during 1990–2020

	Site	[Latitude, longitude]	Number of cuts per season	Years of available Data	Measured variables (Temporal resolution)	Soil type	Precipitation mm yr ⁻¹	Temperature mm yr ⁻¹	Sources
Calibration	Rollesbroich	[50.6, 6.3]	Four cuts	2015-2018	<i>LAI, AGB</i> (Bi-monthly) <i>SM</i> (Daily)	Stagnic Cambisol	910	8.75	(Pütz et al., 2016)
	Gießen1	[50.5, 8.7]	Two cuts	1998-2018	<i>LAI, AGB</i> (Bi-monthly) <i>SM</i> (Daily)	Fluvic Gleysol	624	9.93	(Obermeier et al., 2017)
Validation	Gießen 2-3	[50.5, 8.7]	Two cuts	1998-2018	<i>LAI, AGB</i> (Bi-monthly) <i>SM</i> (Daily)	Fluvic Gleysol	624	9.93	(Obermeier et al., 2017)
	Braunschweig	[52.3,10.4]	Three cuts	2013-2018	<i>LAI, SM</i> (weekly)	Albeluvisol	613	9.83	(Agrometeorological Research Center (ZAMF))
	Bad Lauchstädt 1-5	[51.4,11.9]	Four cuts	2015-2017	<i>LAI, AGB</i> (Bi-monthly)	Haplic Chernozem	494	9.93	(Schadler et al., 2019)
	Bad Lauchstädt 6-10	[51.4,11.9]	Two cuts						
	Selhausen	[50.8, 6.5]	Four cuts	2015-2018	<i>LAI, AGB</i> (Bi-monthly) <i>SM</i> (Daily)	Cutanic Luvisol	674	10.60	(Pütz et al., 2016)

166

The soil data of each experiment have been measured within the mentioned projects (See Table 1) and became available for this study. More detailed information on the soil profiles is presented in Table S.1. Similarly, management data (cutting dates and fertilization) were recorded for these projects (Table S.2). The measured values for *LAI* (leaf area index) and *AGB* were available to evaluate simulations of vegetation dynamics. The *SM* values, measured at different depths across the soil profile (depending on the experiment), were used to evaluate simulations of soil water dynamics.. The grassland sites also differ with respect to the number of cutting events per growing season (Table 1).

To simulate grassland production across Germany, we collected information on cutting dates from different grassland farms across Germany (Fig. 1). Daily records of minimum and maximum temperature, radiation, precipitation, wind speed, and relative humidity were obtained from the German Weather Service (Deutscher Wetterdienst, DWD) and interpolated in a 1 km² grid as described in Zhao et al. (2015). Data were available for the period 1990-2018. The original soil data source is the soil map of Germany, at the scale of 1:200,000, differentiated by land use (BÜK200N, “Nutzungsdifferenzierte Bodenübersichtskarte”, German Federal Institute for Geosciences and Natural Resources (BGR, 1995). For input into the process-based model, soil information was delineated according to DWD grids. The soil parameters required for model simulations included soil texture in mass fraction, as well as volumetric (%) soil water content at saturation, field capacity, and permanent wilting point. We simulated grassland types with different numbers (two, three, four, and five) of cutting events per growing seasons. For regional comparison of grassland production across country, the landscape of the country was divided into eight regions or natural units (NUnits 1–8) based on geomorphological, hydrological, and geological criteria (Fig. 1b and Table S.3) (DWD, 1994). In this paper, the grassland productivity of different natural units have been estimated from Germany-wide simulations.

2.2. The grassland model

The MONICA model is a process-based agro-ecosystem simulation model for simulating crop growth, water dynamics, and nitrogen dynamics for practical applications. It is an extended version of HERMES (Kersebaum, 2007) for the carbon cycle in soil and plants. The model’s code is generic, meaning that soil–plant–water processes can be described using crop-, species-, and soil-specific parameters. MONICA has been tested for a variety of crops in Central Europe (Nendel et al., 2014; Stella et al., 2019), and its generic structure can accommodate grassland characteristics to some degree. Grass growth processes are simulated by dividing growth into six phenological stages, from emergence to senescence (Table S.4). The transition from one stage to another is driven by the accumulation of temperature sum above a base temperature. Owing to differences in the vegetation development of perennial and annual plants, some adaptations were required to achieve the desired level of model performance for perennial grasses, drawing on earlier model implementations that simulated grapevine growth (Nendel and Kersebaum, 2004).

First, unlike annual crops, which are harvested once during the growing season, perennial grasses are often cut several times in a single growing season. Therefore, after each cutting, the growth stage of the plant is reinitialized to the leaf emergence stage. We added four options to determine the amount of stubble mass remaining after a cutting event: 1) percentage of stubble remaining on the field; 2) the amount of biomass (kg ha^{-1}) remaining on the field; 3) grass height after cutting; and 4) the *LAI* value after cutting.

Second, grasses survive from one year to another, going dormant each fall and regrowing each spring. The stolon of grasses provides new growing points during tilling each spring as well as after cutting. To incorporate this trait of grasses, the model defines a perennial reserve organ for grass plants. Each day, the potentially available carbohydrate from the reserve organ is calculated and added to the daily assimilate production. The day's maintenance respiration costs and any assimilated loss due to growth-limiting factors are subtracted from the total; the remainder is the amount of carbohydrate available for regrowth. Two factors determine the contribution rate of the reserve organ during regrowth: 1) a coefficient called *StageMobilFromStorageCoeff*, which varies according to the developmental stage of the plant, and which is estimated during calibration; 2) a coefficient that modifies the contribution of the storage organ based on the average temperature of the day (Mitchell et al., 1995). Finally, the root systems of annual crops typically die after the growing season. For grasslands, however, to avoid unnatural accumulations of root mass in the soil the model uses the pre-defined pools in MONICA to simulate a small proportion (2-6%) of the root dying back every day, thus contributing to soil organic matter decomposition and mineralization.

2.3. Model parameterization, sensitivity analysis, and calibration

We followed three steps to simulate grassland productivity across Germany (Fig. 2). As is typical for environmental modeling, the first step was to conduct a sensitivity analysis to define the most sensitive parameters. MONICA has over 200 parameters describing different processes such as plant growth, soil water, and soil carbon turnover processes. Some of them relate to growth stages (Table S.4) and plant organs (e.g. root, leaf, stem, and stolon for grasslands). In this study we selected 111 parameters which were relevant for simulating vegetation and soil moisture components (Table S.5). The sensitivity analysis was conducted for the five experimental sites presented in section 2.1. Soil properties were considered only in the first step of sensitivity analysis and of course not during the calibration. The reason to include these parameters in sensitivity analysis was to reflect some knowledge on uncertainty due to soil heterogeneity which is important for simulation at large scales simulation with 1 km^2 resolution. In fact, at such scale, the soil properties vary significantly at such spatial scale and its uncertainty should be considered. However, we did not use soil characteristics as parameters for calibration in any experimental results.

Sensitivity analyses have already been implemented for MONICA in previous studies (Specka et al., 2015, 2019), yet they have lacked analysis related to the new version of the model for perennial crops with newly

added parameters (Section 2.3.1). Having identified the most sensitive parameters, we implemented SO and MO calibration procedures (Section 2.3.2). The final parameter sets and their ranges were then used to simulate grasslands with different numbers of cuts per growing seasons (i.e. two, three, four, and five cuts) across Germany.

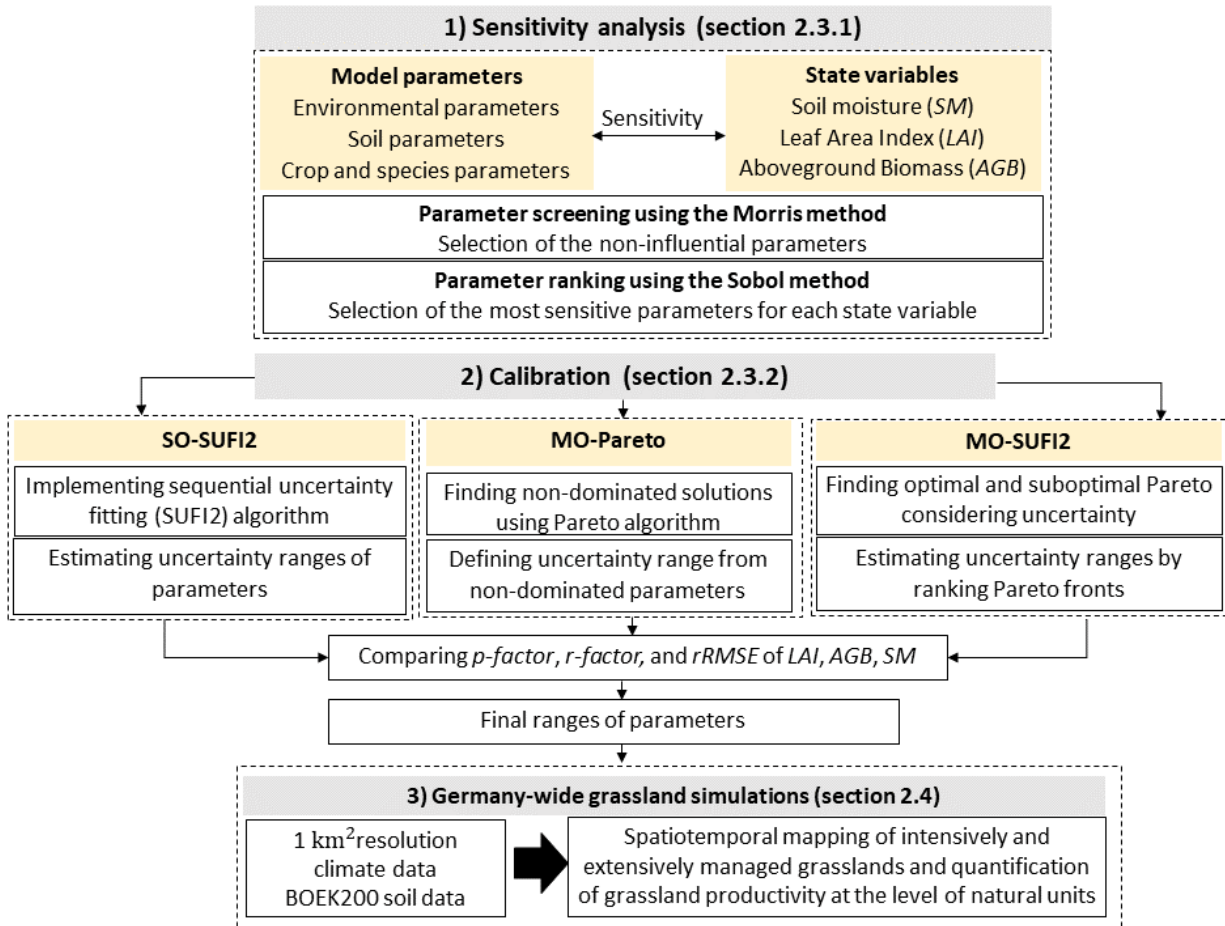


Fig. 2. Schematic representation of the proposed three-step framework (sensitivity analysis, calibration, Germany-wide grassland simulation) for mapping grassland productivity under various cutting regimes.

2.3.1. Sensitivity analysis

We conducted sensitivity analysis for all five experimental grassland sites mentioned in Table 1. Accordingly, the soil and weather data used for each experiment have been explained in section 2.1. Similar to methods widely-used in literature (Confalonieri et al., 2010), sensitivity analysis proceeded in two steps. The first step was to use a screening method to exclude non-influential parameters. The Morris method was selected for this purpose (Morris, 1991). In the process, parameter sensitivity is defined by making local changes at different points across possible parameter ranges. The main advantage of this method over variance-based methods (which are more reliable for ranking sensitivity) is that they require less computational effort in cases with huge numbers of parameters. Therefore, it provides a possibility to screen parameters and exclude those with non-influential effects. This subsequently reduces the number of

parameters when applying variance-based methods, such as the Sobol method (Sobol, 1967). The Morris method uses two measures – μ (the mean of the distribution of the elementary effects) and σ (the standard deviation of the distribution of the elementary effects) to identify the sensitivity of parameters. To calculate μ and σ , the method first calculates the Elementary Effect (EE), which is obtained from dividing the parameter space into p discrete levels. The model outputs are then calculated for r trajectories within the parameter space. EE for parameter i is defined as:

$$EE_i = \frac{f(X_1, \dots, X_{i-1}, X_i + \Delta, X_{i+1}, \dots, X_n) - f(X)}{\Delta}, \quad (1)$$

where $\Delta = p/(2(p - 1))$. After repeating the procedure r times, the mean and standard deviation of EE is used as a measurement to evaluate the sensitivity of parameters. That is, we obtain the Morris measures of μ and σ for each parameter i as:

$$\mu_i = \frac{\sum_{j=1}^r EE_{ij}}{r} \quad (2)$$

and

$$\sigma_i = \sqrt{\frac{\sum_{j=1}^r (EE_{ij} - \mu_i)^2}{r}} \quad (3)$$

Based on the combination of σ and μ , the parameters are grouped into three classes: 1) parameters with negligible effects (here, $\sigma < 0.05$ and $\mu < 0.05$); 2) parameters with large linear effects without interactions (high σ with low μ); and 3) parameters with large non-linear and/or interaction effects. We excluded from subsequent analysis the parameters that fell into groups of parameters with negligible effects i.e. $\sigma < 0.05$ and $\mu < 0.05$.

In the second step, the remaining parameters from the Morris method (group 2 and 3) were ranked to determine the contribution of each factor to the total variability. To do so, we used the Sobol variance-based approach. The probabilistic framework of this method decomposes the variance of the model output into fractions, which can then be attributed to inputs or sets of inputs. The method makes it possible to calculate two indices: 1) the main effect (ME) sensitivity index, which measures the direct effect of individual factors on the variance of the outputs:

$$\text{Main effect}_i = \frac{\text{Variance}(E[\text{Yield} | X_i])}{\text{variance}(Y)} \quad (4)$$

where $E[\text{Yield} | X_i]$ denotes the expected value of yield across all factors X_i , and 2) the total effect (TE) sensitivity index which corresponds to a single factor and the interaction of multiple factors:

$$\text{Total effect}_i = 1 - \frac{\text{Variance}(E[\text{Yield}|X_{-i}])}{\text{Variance}(Y)} \quad (5)$$

where $E[\text{Yield}|X_{-i}]$ is the expected value of yield across all factors except X_i .

2.3.2. Model calibration

We calibrated the model by using one SO and two MO calibration algorithms and then compared the results. The three optimization algorithms differ in their strategy for searching parameter ranges. To compare the performance of all three algorithms, we used three criteria: one to compare the simulated and observed variables and to identify the best-performing parameter set using relative root mean square error (*rRMSE*), and two additional criteria (*p-factor* and the *r-factor*) to show the goodness-of-fit and the level of uncertainty of the model. In all three algorithms and to evaluate the accuracy of the different calibration procedures and find the best performing models, we compared the observed and simulated values of three reference variables v (*LAI*, *AGB*, and *SM*) using *rRMSE* which is calculated as:

$$rRMSE_v = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (\text{Observed}_{i,v} - \text{Simulated}_{i,v})^2}}{\text{Std}(\text{Observed}_v)} \quad (6)$$

The respective error functions for the three reference variables in this study are $rRMSE_{LAI}$, $rRMSE_{AGB}$, and $rRMSE_{SM}$ for *LAI*, *AGB*, and *SM*.

The performance of models regarding the level of uncertainty in the model were measured using the *p-factor* and the *r-factor* Criteria. The *p-factor* criterion represents the fraction of measured data bracketed by the 95PPU uncertainty band and varies from 0 to 1, where 1 means that 100% of the measured data is bracketed by the model simulation (expressed as the 95PPU). Values around 0.5 are usually considered acceptable for crop simulation (Kamali et al., 2018). The *r-factor* criterion is the average width of the 95PPU band divided by the standard deviation of the measured variable, which is a measure of prediction uncertainty. The ideal value for *r-factor* is 0, with an acceptable practical value of around 2 for crop yields (Kamali et al., 2018). Detailed information describing this method can be found in Abbaspour (2015).

The two experimental sites of Gießen1 (a site with two/three cuts per growing season) and Rollesbroich (a site with four/five cuts per growing seasons) were selected as calibration sites, because they both had measured variables for *LAI*, *AGB*, and *SM*. The parameter ranges obtained from the calibration sites were then transferred to other sites to test model suitability at the validation stage. The parameter sets obtained from the calibration procedure were then used to test the performance of the model in terms of simulating other variables in the validation sites with an independent dataset.

Single-objective uncertainty-based calibration (SO-SUFI2)

This study used the sequential uncertainty fitting algorithm (SUFI-2) technique (Abbaspour et al., 2015) for SO calibration which considers on one variable at a time. The algorithm maps all uncertainties in the output on the parameter ranges. Uncertainty was quantified by the 95% prediction uncertainty (95PPU) calculated at the 2.5% and 97.5% levels of cumulative distribution of an output variable obtained using Latin Hypercube sampling in the parameters' space. We started the SO-SUFI2 procedure with 1,000 parameter sets sampled from wide ranges in the first iteration. The *p-factor*, *r-factor*, and *rRMSE* values were accordingly calculated. If they were not within the acceptable ranges (*p-factor* larger than 0.5 and *r-factor* between 0.9 and 2). The algorithm calculated a new range for each parameter, based on the comparison of simulated and observed variable e.g. the *rRMSE_{LAI}* of 1,000 samples. The new ranges of parameters were then used to create another 1,000 samples of parameter sets in the second iteration. We continued the iterations until we reached a compromised balance between the *p-factor* and the *r-factor*. In this paper, the best compromise between two criteria was reached after three iterations. The objective functions used in the second and third iteration were the same as the first iteration (*rRMSE*, *p-factor*, and *r-factor*).

Multi-objective calibration based on Pareto optimality (MO-Pareto)

The SUFI-2 algorithm was developed for SO calibration. The structure of the algorithm combines multiple objectives into one, and was not designed in a way to optimize multiple objectives at the same time (Abbaspour, 2015). In the case of multiple objectives, the algorithm calculates the average of different objectives and considers it as a single final value for optimization. We are interested in the simultaneous optimization of multiple objectives (*rRMSE_{LAI}*, *rRMSE_{AGB}*, and *rRMSE_{SM}*). Therefore, we tested the suitability of the MO calibration approach, which is based on minimizing more than one objective function according to Pareto optimality concepts (hereafter MO-Pareto). The Pareto optimality concept provides a set of possible solutions based on the tradeoff among two or more error functions – that is, there is no unique set of optimal parameters, but a number of solutions exist in which it is not possible to improve one objective function without compromising at least one other. Such solutions are called Pareto, non-dominated, non-inferior, or efficient solutions.

The non-dominated solutions comprise a front called an optimal Pareto front (Fig. 3a), which divides the search space of solutions into feasible and infeasible regions. As shown in Fig. 3a (as an example), Solution 1 has the lowest *rRMSE_{LAI}*, but the highest *rRMSE_{SM}* among all solutions located on the optimal Pareto front. In contrast to Solution 1, Solution 2 has the lowest *rRMSE_{SM}* and the highest *rRMSE_{AGB}*. Therefore, the two solutions are non-dominated vis-à-vis each other, meaning that one is ideal for one objective function and the other is ideal for another objective function. The MO-Pareto algorithm explores the search space of the parameters and aims to define the optimal Pareto front. The Pareto optimality conditions are defined as:

$$\forall v = LAI, AGB, SM: rRMSE_v(\theta_i) \leq rRMSE_v(\theta_j) \quad (7)$$

$$\exists v = LAI, AGB, SM: rRMSE_v(\theta_i) < rRMSE_v(\theta_j) \quad (8)$$

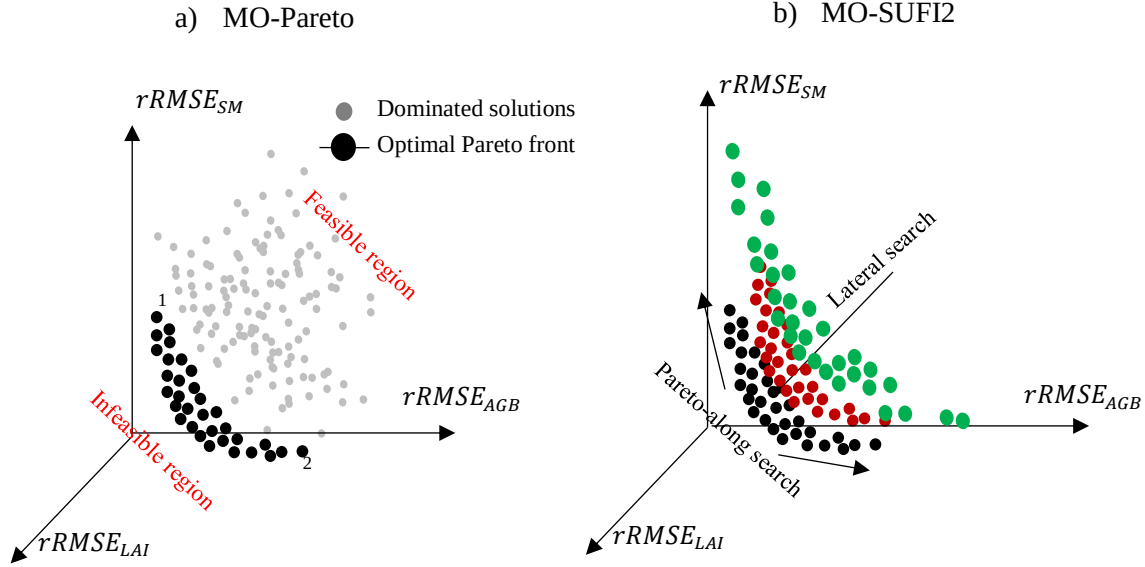


Fig. 3: Illustrations of the MO optimization method based on Pareto-optimal concepts (MO-Pareto) for three objectives. In a), the optimal Pareto front divides the search space into feasible and infeasible regions; b) shows the schematic representation of ranking Pareto fronts used to measure the model uncertainty (MO-SUF12) by searching in two orthogonal dimensions i.e. Pareto-along and lateral search.

The first expression divides the feasible parameter space into two sets of feasible (Pareto) and infeasible solutions. The second expression states that, in the absence of additional information, it is not possible to distinguish among any of the efficient solutions as being objectively better than any other. The MO-Pareto algorithm considers only $rRMSE$ to search among feasible solutions and find the optimal Pareto front. During the search, the algorithm does not take into account p -factor and r -factor (as the representative criteria in our study) to evaluate the model performance in terms of capturing uncertainty band. We here calculated p -factor and r -factor according to those non-dominated solutions located on the optimal Pareto front as a surrogate to parameter ranges. By calculating p -factor and r -factor, we tested if the parameter sets located on the optimal Pareto front yield satisfactory p -factors and r -factors or not. This provides a solid base to compare the performance of this technique with others in terms of capturing the uncertainty ranges of parameters.

Novel uncertainty-based multi-objective SUFI-2 (MO-SUFI2)

The MO-SUFI2 with ν state variables (i.e. ν objective functions) combines the Pareto optimality concept with the uncertainty-based concept of SO-SUFI2. During the search process, the algorithm not only considers all three variables, but also search parameters spaces strategically in a way that finds those ranges giving acceptable p -factor and r -factor values. The Pareto optimality concept focuses on considering multiple objectives during optimization, which when combined with the SO-SUFI2 concept in the novel MO-SUFI2 algorithm, makes it possible to find the uncertainty ranges of model parameters. This integrative perspective in MO-SUFI2 is possible by exploring the parameter space using two types of searching orthogonally to each other (Fig. 3b): 1) the Pareto alongside search, which identifies those non-dominated solutions that satisfy multiple objectives; and 2) the lateral search, which divides feasible solutions into sub-solutions by defining multiple Pareto fronts and ranking them based on the distance from optimal Pareto front (Fig. 3b). The ranking of candidate parameters allows to stratify the search space along lateral direction and adjust the ranges of the parameters until the satisfactory p -factor and r -factor values are achieved. We first start with defining rank-1 (which is the optimal Pareto front in MO-Pareto). We first start with separating parameter sets with rank-1 from all candidate sets using Equations 7 and 8. These sets are actually the optimal Pareto front already obtained in MO-Pareto. Having excluded them from the candidate parameter sets, we define rank-2 with the remaining sets using the same equations. The more specific stepwise procedure of the MO-SUFI2 algorithm is as follows:

- 1) The MO-SUFI2 algorithm started with N numbers (here $N=100$) of parameter sets, i.e. feasible solutions stored in one main repository.
- 2) The ν objective functions, i.e. $rRMSE_{LAI}$, $rRMSE_{AGB}$, and $rRMSE_{SM}$, were calculated for all N feasible solutions.
- 3) The Pareto rules (Equations 7 & 8) were applied and the non-dominated solutions located on optimal Pareto fronts were selected. The solutions located on the optimal Pareto front were assigned to the first rank.
- 4) The first-rank solutions were removed from the main repository.
- 5) The Pareto rules were reapplied to the remaining solutions in main repository to identify a new set of non-dominated solutions, labeled as the second rank.
- 6) Steps 4 and 5 were repeated until all N solutions were labeled with a rank, as illustrated in Fig. 3b.
- 7) In solutions with the same rank, we took out 5% of those solutions with highest residual between objective functions, i.e. solutions that performed worse in at least one objective function.
- 8) The p -factor and r -factor were calculated based on the first-rank solutions stored in the repository. If the p -factor and r -factor exceeded their pre-specified thresholds, the algorithm stopped and the p -factor and r -factor became the final values. In this paper, the threshold defined for the p -factor was a value above 0.5, with the r -factor varying between 0.9 and 2.0. If these conditions were not fulfilled, the

second-rank solutions were combined with solutions from the first rank and the *p-factor* and *r-factor* values were recalculated. The stepwise procedure of recalculating *p-factor* and *r-factor* was continued until the final *p-factor* and *r-factor* met the specified threshold for the *p-factor* and *r-factor*.

2.4. The simulation of grasslands across Germany

The parameters ranges obtained from MO-SUFI2 calibration were used to simulate the fresh *AGB* of grasslands across Germany. We simulated the MONICA model for all grids across Germany and calculated the annual amount of fresh *AGB* exported from the field/grid (hereafter *Yield*) as model output. The next step was to determine the spatiotemporal dynamics of grasslands and uncertainty levels for *Yield* estimation in Germany's eight nature units (NUnits 1-8) during 1991-2018 (Fig. 1b). We extracted the productivity of grasslands for all grids within each natural units and then calculated the average fresh *AGB*. Simulations were conducted for four cutting regimes: two and three cutting events per year as representative of extensively managed grasslands and four and five cutting events as representative of intensively managed grasslands. We assumed fixed cutting dates across all grids, based on information collected from farmers throughout Germany (farm sites in Fig. 1). The amount of fertilization applied in the different cutting regimes was 30 kg N ha⁻¹ at the beginning of the growing season and 15 kg N ha⁻¹ five days after each cutting event (Table S.6).

3. Results

3.1. Sensitivity analysis

The Morris sensitivity method for experimental sites did not show differences in the identified most sensitive parameters among case studies and the result for all were found to be similar to each other. The σ and μ indices were calculated according to daily information within this simulation time. Fig. 4a shows the maximum σ values during the growing season for different parameters in relation to the *LAI*, *AGB*, and *SM* state variables. We screened parameters based on the calculated σ and μ values and excluded those with values smaller than 0.05, owing to their negligible influence on simulating those state variables. As Fig. 4a shows, the most sensitive parameters varied depending on the state variable. For example, the soil texture parameters (*Field Capacity* and *Pore Volume*) and *StageKcFactor* were most relevant when simulating *SM*, whereas *AGB* and *LAI* were most affected by the parameter *SpecificLeafArea*. Additionally, some parameters, such as *OrganGrowthRespiration* and *BaseDayLength*, were sensitive for the simulation of all three state variables.

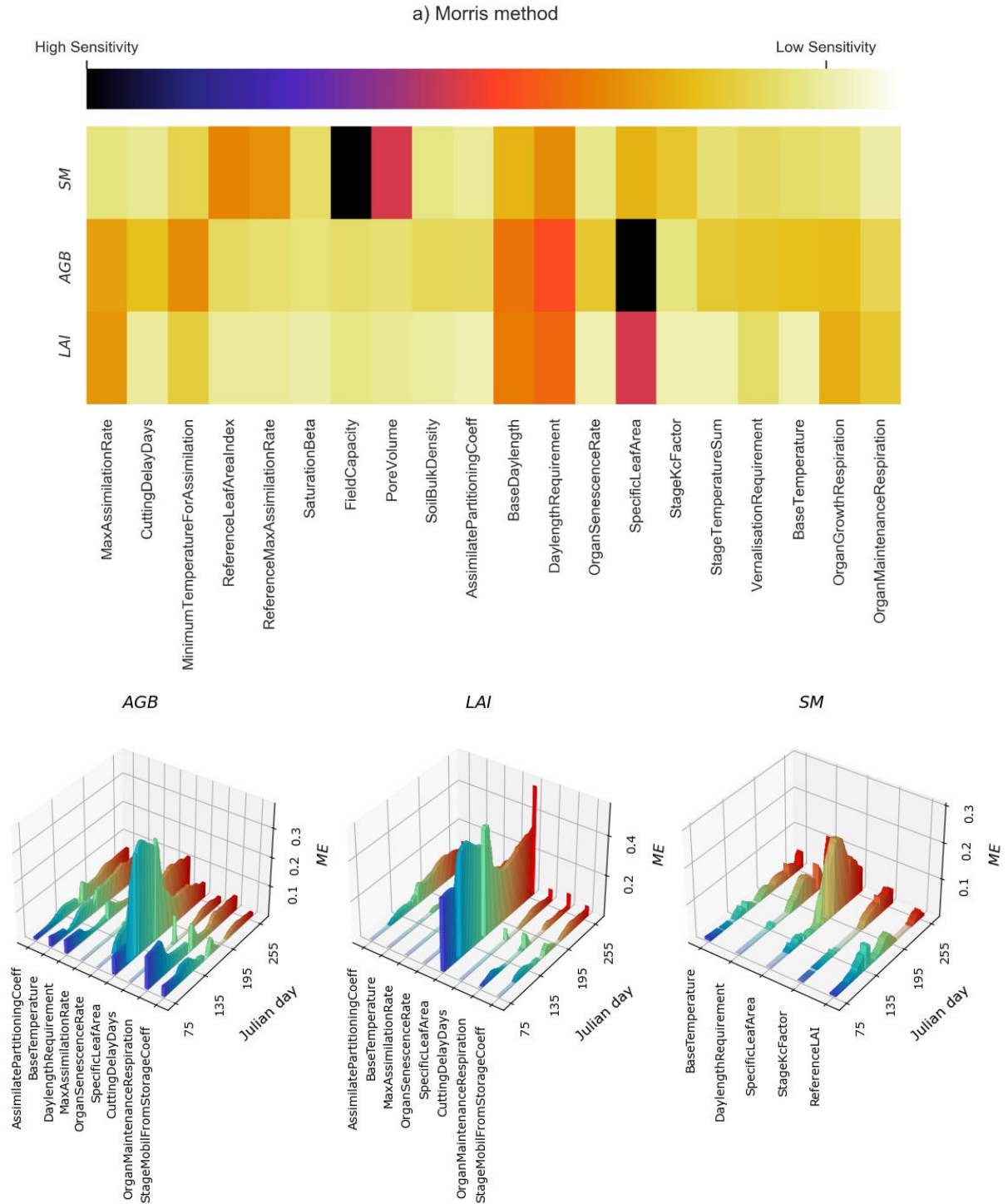


Fig. 4. a) Screening method: the sensitivity of different parameters to the *LAI*, *AGB*, and *SM* variables obtained from the Morris screening method. The results were obtained from those parameters with a maximum μ (i.e. maximum during the growing season) exceeding 0.05. b) **Ranking method:** the temporal variability of the *ME* (main effect) sensitivity index during the growing season (shown in terms of Julian days) obtained from the Sobol ranking method. Only parameters with maximum *ME* > 0.05 are depicted.

We also looked at the relationship between μ and σ to check if there was any interaction between sensitive parameters. As shown in Fig. S.1, a high value of μ occurs with an equally high (or even higher) σ , meaning that nonlinear relationships exist among variables. This means that the most sensitive model parameters (parameters with high μ) also have high levels of interaction with other parameters. We did not further analyze the interaction of parameters in the first step of the sensitivity analysis, because in this paper the Morris screening method was only used as a preliminary screening analysis to eliminate those parameters with negligible effects, which was critical for reducing computational costs. Overall, after excluding parameters with σ and μ smaller than 0.05, 20 parameters (out of 111 initial parameters) were found to be sensitive. It is worth remarking that the model dimensionality was even higher than 20, because some parameters, such as *OrganScenesenceRate* and *AssimilatePartitionCoeff*, are not individual values but are an array or a matrix, as their values vary depending on the growth stage of the plant (Table S.4) and on the plant organ (root, shoot, leaf, and stolon). Here, we only presented the average values obtained from all organs and growing stages.

The Sobol method was carried out next, based on the most relevant parameters obtained from the Morris method for each day during the growing season. Fig. 4b shows the daily temporal variation of *ME* during one growing season, revealing that the ranking of most parameters varied during growing seasons. Moreover, the number of sensitive parameters and their sensitivity differ depending on the state variables (*LAI*, *AGB*, or *SM*). Some parameters, such as *SpecificLeafAreaIndex*, were identified as the most sensitive parameter for simulating *AGB* and *LAI* during the growing season. However, other parameters, such as *OrganMaintenanceRespiration*, exhibited a higher sensitivity than *MaximumAssimilationRate* in early stages of the growing season, but at the end of the growing season, *MaximumAssimilationRate* was found to have been more sensitive than in early stages. This clearly demonstrates the temporal sensitivity of parameters during the observation/growing season.

3.2. Single-objective calibration

The results of calibration shows promising performance for both grassland sites with two/three cuts and four/five cuts (Figs. 5a and b). The best performing parameter set (the set with the lowest $rRMSE_{LAI}$) with an $rRMSE_{LAI}$ of 0.67 for Gießen1 and 0.61 for Rollesbroich shows that MONICA is able to replicate the dynamics of *LAI*. The *p-factors* of 0.65 for Gießen1 and 0.61 for Rollesbroich indicate that over 60% of the observed values are covered by the 95PPU band. The *r-factor* values of 1.64 and 1.42 for Gießen1 and Rollesbroich respectively are within the range of acceptable values in the literature (Kamali et al., 2018). As seen in Figs. 5c-5h, the results are promising in both validation grassland sites with two/three cuts and four/five cuts. The $rRMSE_{LAI}$ in validation sites exhibited an approximately similar performance to those of the calibration sites. The *p-factors* with values varying between 0.50 and 0.64 are close to the calibration

sites, with slightly outperforming values from Bad Lauchstädt1 and Bad Lauchstädt2. The r -factors reached their lowest value of 1.63 in Bad Lauchstädt1 and Braunschweig, with the highest value of 1.96 in Selhausen, both of which are quite close to the value of 1.42 at their corresponding calibration sites.

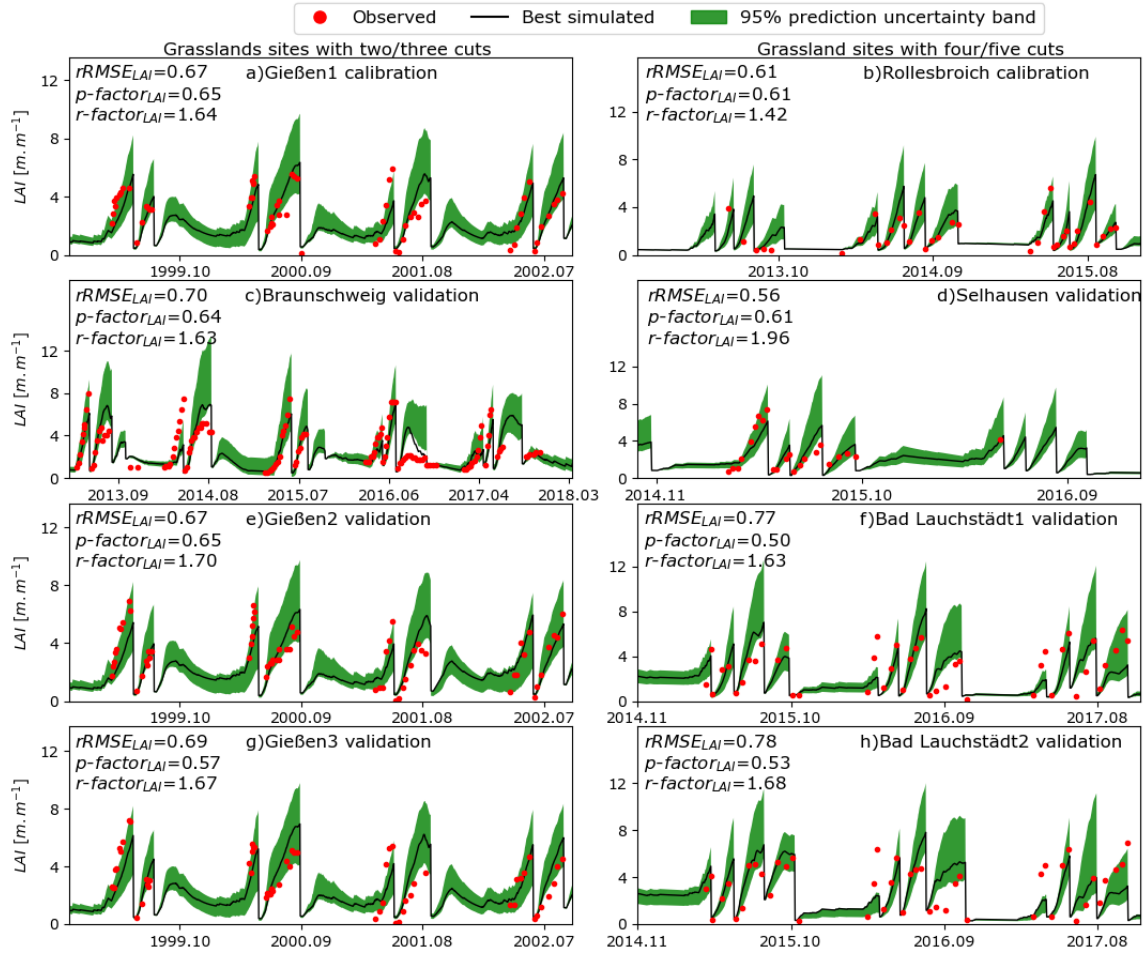


Fig. 5. MONICA simulations of Leaf Area Index (LAI) dynamics based on data from the calibration (a, b) and validation (c-h) sites, after the model was calibrated only for LAI using SO-SUFI2, i.e. minimizing $rRMSE_{LAI}$. The left column shows the results for grassland sites with two/three cuts, the right column for sites with four/five cuts. The $p\text{-factor}_{LAI}$, the $r\text{-factor}_{LAI}$, and the relative root mean square error ($rRMSE_{LAI}$) are given as the performance indicators.

We also checked the performance of the calibrated model (against LAI) in terms of simulating AGB and SM variables. The results for AGB and SM were not as robust as for LAI (Figs. S.2 and S.3). As shown in Fig. S.2, the $rRMSE_{AGB}$ values for all sites (with the exception of Rollesbroich and Selhausen) were above 2.0, which is considerably higher when compared with $rRMSE_{LAI}$. The p -factors with values smaller than 0.1 confirmed the inadequacy of the calibrated model (against LAI) in terms of capturing observed AGB and SM within 95PPU (Figs. S2 & S3). In the next step, we calibrated the model based on AGB , and found that better performance could be achieved in terms of simulating AGB , but the calibrated model using AGB as

the target variable (i.e. minimizing $rRMSE_{ABG}$) this time did not perform adequately well for LAI (Table 2). For example in Gießen1, the $rRMSE_{AGB}$ improved from 2.42 to 1.01, but the $rRMSE_{LAI}$ increased from 0.67 to 1.30. A similar pattern emerged when we calibrated the model based on only SM (minimizing $rRMSE_{LAI}$) (Table 2). For example, one can see improvement in $rRMSE_{SM}$ from 0.97 to 0.75 in Bad Läuchstadt1. Using the sum of the objective functions ($rRMSE_{LAI} + rRMSE_{AGB} + rRMSE_{SM}$) did not improve the results either, as the optimization algorithm may have tended to optimize based on one criterion. In Bad Läuchstadt1, while integrating all three objectives improved the overall performance, the model performed worse in terms of $rRMSE_{LAI}$ with a highest value of 0.89. This emphasizes the implementation of a multi-objective calibration procedure, which considers the impact of improving one objective at the expense of another objective function.

3.3. Multi-objective calibration

3.3.1. Traditional multi-objective calibration (MO-Pareto)

Fig. 6a shows the results of the calibration using Gießen1 data. The non-dominated solutions on the optimal Pareto front indicate solutions for which improving one objective occurs at the expense of compromising another objective. The solutions located on the optimal Pareto front (optimal non-dominated solutions) were used to calculate the p -factor and r -factor values, and the best performing solution (i.e. the lowest $rRMSE$) was extracted from that. Comparing the dynamics of best-performing LAI , AGB and SM in MO-Pareto (Fig. 6a) with SO-SUFI2 (Figs. 5a and S.3a) showed a lower $rRMSE_{AGB}$ (from 2.42 to 1.89) and $rRMSE_{SM}$ (from 0.77 to 0.72). However, no improvement was seen in $rRMSE_{SM}$ as well as in the p -factor and the r -factor values. The model behavior in Braunschweig as validation site improved (Fig. 6b). The $rRMSE_{LAI}$ with a value of 0.67 in the MO-Pareto approach decreased to 0.51 in SO-SUFI2. At the same time, $rRMSE_{SM}$ values improved from 0.73 in the SO-SUFI2 calibration scheme to 0.66 in the MO-Pareto calibration scheme. We did not see improved values for p -factor and r -factor values when we applied the MO-Pareto calibration scheme, i.e. p -factor values were smaller than 0.5 and r -factor values were very low. This indicates that the non-dominated solutions located on the optimal Pareto front may not be representative of parameter uncertainty.

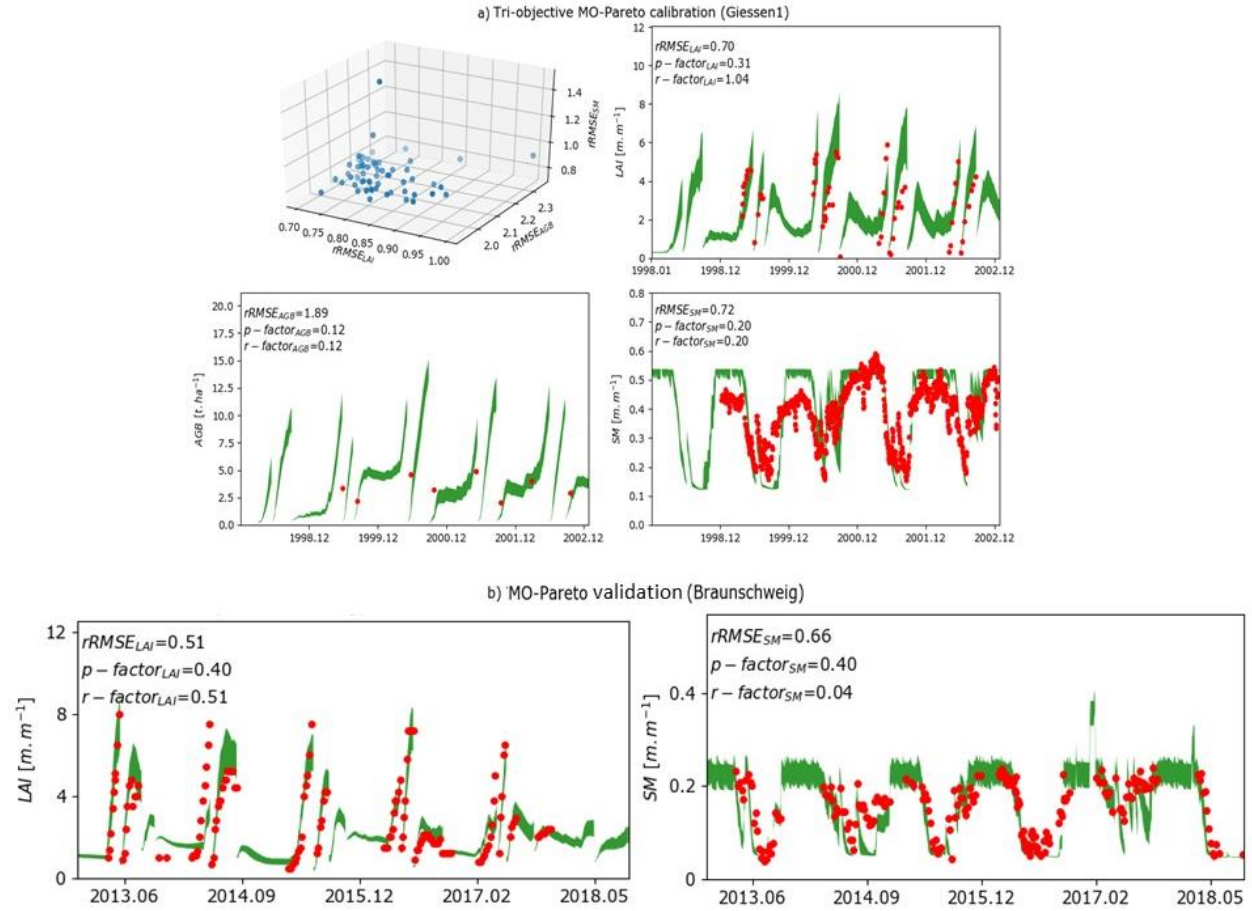


Fig. 6. MONICA simulations of multiple target variables, after the model had been simultaneously calibrated using the MO-Pareto calibration algorithm: a) calibration in MO-Pareto and b) validation in Braunschweig. The relative root mean square error ($rRMSE$), the p -factor and the r -factor are given as performance criteria.

533
534

Table 2. Comparing performance indicators of the $rRMSE$, the p -factor, and the r -factor based on utilizing different combinations of $rRMSE_{LAI}$, $rRMSE_{AGB}$, and $rRMSE_{SM}$ as objective functions for model calibration (“min” here means “minimization”)

Grassland site with two/three cuts (calibration: Gießen1)									
Objective function	$rRMSE_{LAI}$	$rRMSE_{AGB}$	$rRMSE_{SM}$	p -factor _{LAI}	p -factor _{AGB}	p -factor _{SM}	r -factor _{LAI}	r -factor _{AGB}	r -factor _{SM}
min [$rRMSE_{LAI}$]	0.67	2.42	0.76	0.65	0.25	0.27	1.64	1.55	0.05
min [$rRMSE_{AGB}$]	1.30	1.01	0.81	0.42	0.43	0.38	1.91	1.20	0.03
min [$rRMSE_{SM}$]	1.22	1.81	0.66	0.41	0.35	0.44	1.81	1.33	0.40
min [$rRMSE_{LAI} + rRMSE_{AGB} + rRMSE_{SM}$]	0.72	2.2	0.78	0.55	0.40	0.40	1.7	1.25	0.23
Grassland site with two/three cuts (validation: Braunschweig)									
min [$rRMSE_{LAI}$]	0.70	-	0.73	0.64	-	0.35	1.63	-	0.06
min [$rRMSE_{AGB}$]	-	-	-	-	-	-	-	-	-
min [$rRMSE_{SM}$]	0.81	-	0.63	0.51	-	0.51	1.50	-	0.90
min [$rRMSE_{LAI} + rRMSE_{AGB} + rRMSE_{SM}$]	0.75	-	0.8	0.51	-	0.38	1.45	-	0.89
Grassland with four/five cuts (calibration: Rollesbroich)									
min [$rRMSE_{LAI}$]	0.61	1.7	0.70	0.61	0.12	0.62	1.42	0.72	0.05
min [$rRMSE_{AGB}$]	1.02	1.5	0.75	0.48	0.51	0.55	1.66	1.02	0.08
min [$rRMSE_{SM}$]	0.88	1.91	0.61	0.31	0.12	0.64	1.78	0.99	0.30
min [$rRMSE_{LAI} + rRMSE_{AGB} + rRMSE_{SM}$]	0.65	1.63	0.75	0.44	0.23	0.66	1.70	0.88	0.23
Grassland with four/five cuts (validation: Bad Lärchstadt1)									
min [$rRMSE_{LAI}$]	0.77	3.64		0.50	0.005		1.63	0.89	
min [$rRMSE_{AGB}$]	0.88	1.33		0.49	0.22		1.23	0.90	
min [$rRMSE_{SM}$]	0.87	2.98		0.51	0.23		1.22	1.11	
min [$rRMSE_{LAI} + rRMSE_{AGB} + rRMSE_{SM}$]	0.89	1.40		0.50	0.10		1.30	1.12	

535

3.3.2. Uncertainty-based multi-objective SUFI-2 (MO-SUFI2)

After the MO-SUFI2 calibration procedure was performed, based on the ranking of Pareto fronts, the results were compared with the outcome from SO-SUFI2 calibrated using $\min(rRMSE_{LAI} + rRMSE_{AGB} + rRMSE_{SM})$ as an objective function as well as MO-Pareto algorithms. As shown in Fig. 7, the $rRMSE$ values obtained from both MO algorithms outperformed SO-SUFI2 for all state variables at both types of grassland sites i.e. two/three cuts and four/five cuts. For example, the $rRMSE_{LAI}$ in Braunschweig fell from 0.7 to 0.51 in MO-Pareto (Fig. 6) and 0.5 in MO-SUFI2 (Fig. 7a). Similarly, the $rRMSE_{LAI}$ values in Bad Lauchstädt1 fell from 0.89 in the MO-Pareto calibration to 0.55 in MO-SUFI2. The MO-SUFI2 also performed satisfactorily in terms of the two objective functions designed to capture the uncertainty bound, i.e. the p -factor and the r -factor. The p -factors for all three state variables at the calibration and the validation stage and for both in grasslands with two/three and four/five cutting events were above 0.5; the r -factor varied between 0.85 and 1.3.

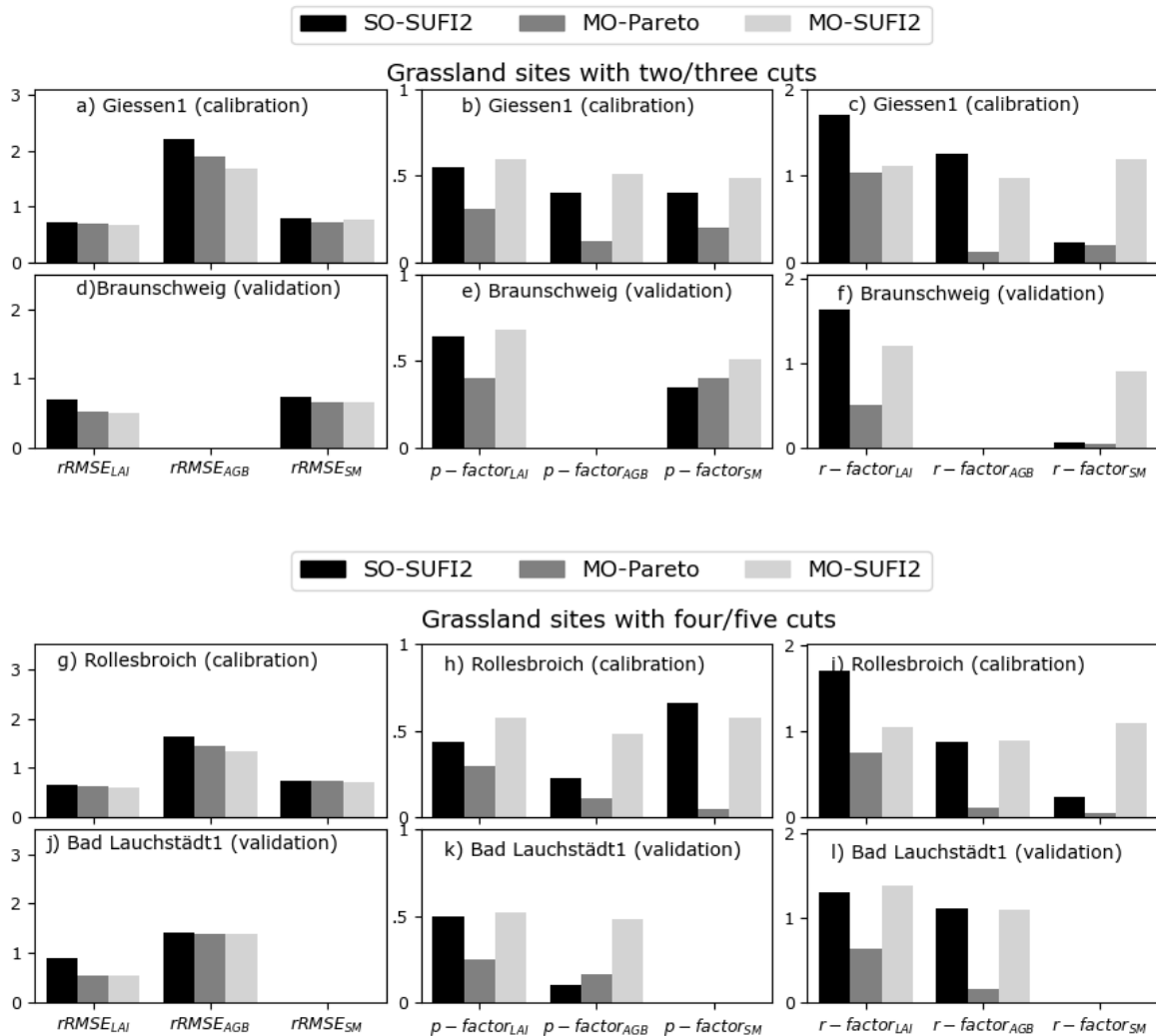


Fig. 7. Comparing the results of SO-SUFI2 with MO-Pareto and MO-SUFI2 calibration at extensively managed (a-f) and intensively managed (g-l) grassland sites and at the calibration and validation sites.

Fig. 8 shows the final ranges of parameters obtained from MO-SUFI2 during calibration. For some parameters, such as *AssimilationPartitionCoeff* and *OrganSenescenceRate*, the ranges differed depending on the organ (root, shoot, leaf, and stolon) and growth stages (Fig. 8). The *SenescenceRate* values of root organs with values between 0.04 and 0.06 had the highest values compared with all other organs. Fig. 8 shows that parameter uncertainty is higher for *AssimilationPartitionCoeff* and *MaximumAssimilationRate* compared with other parameters. Comparing the uncertainty ranges of *AssimilationPartitionCoeff* for different organs presented wider ranges for stolon compared with other organs for all six growth stages. This shows that reducing model uncertainty depends on detailed information on partitioning the assimilated carbon among different organs.

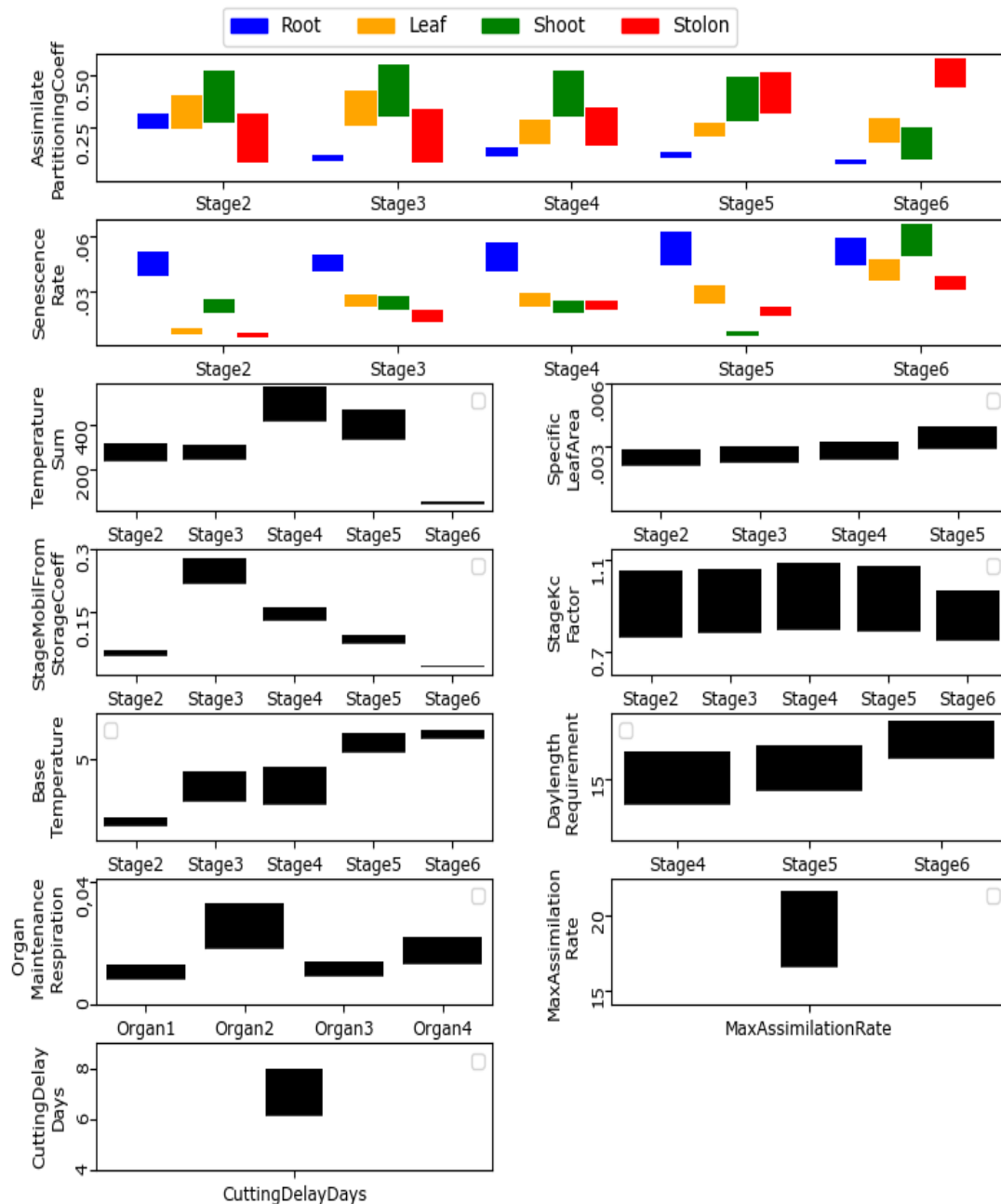


Fig. 8. Final ranges of parameters obtained from the uncertainty-based multi-objective calibration, MO-SUFI2.

3.4. Simulation of grasslands across Germany

Figs 9a-d show the spatial distribution of annual *AGB* productivity (the sum of fresh *AGB* t ha^{-1} from cutting events within the growing season) under different cutting regimes across Germany. The annual productivity rates depicted below were obtained from the average *AGB* from 25 simulations executed with 25 parameter sets sampled within the ranges obtained from 95PPU (Fig. 9). The spatial variability showed a discernable high magnitude of productivity in northwest Germany, with values above $20 \text{ t ha}^{-1} \text{ yr}^{-1}$, and low values in central Germany, with values lower than $12\text{--}15 \text{ t ha}^{-1} \text{ yr}^{-1}$ for the four cutting regimes. The difference in grassland productivity levels of different cutting regimes were well captured by the model, as grasslands with a larger number of cutting events per growing season exhibited higher productivity values in all parts of the country (Fig. 9a–d). Annual grassland productivity varied from 9 t ha^{-1} in central and southern Germany during regimes with two cutting events (the lowest amount) to 18 t ha^{-1} in the northwest of Germany with four cutting events (the highest amount). Eastern Germany, with productivity between 12 and 17 t ha^{-1} , falls somewhere in between. Comparing the productivity of two with three cutting events revealed no remarkable spatial difference between the two types in terms of magnitude or spatial variability. The three-cut regime generated slightly higher productivity, mostly in the northwestern part of Germany. Similarly, the comparison of four with five-cut regimes exhibited a spatial difference mostly in the northwest of Germany (Figs 9 c and d).

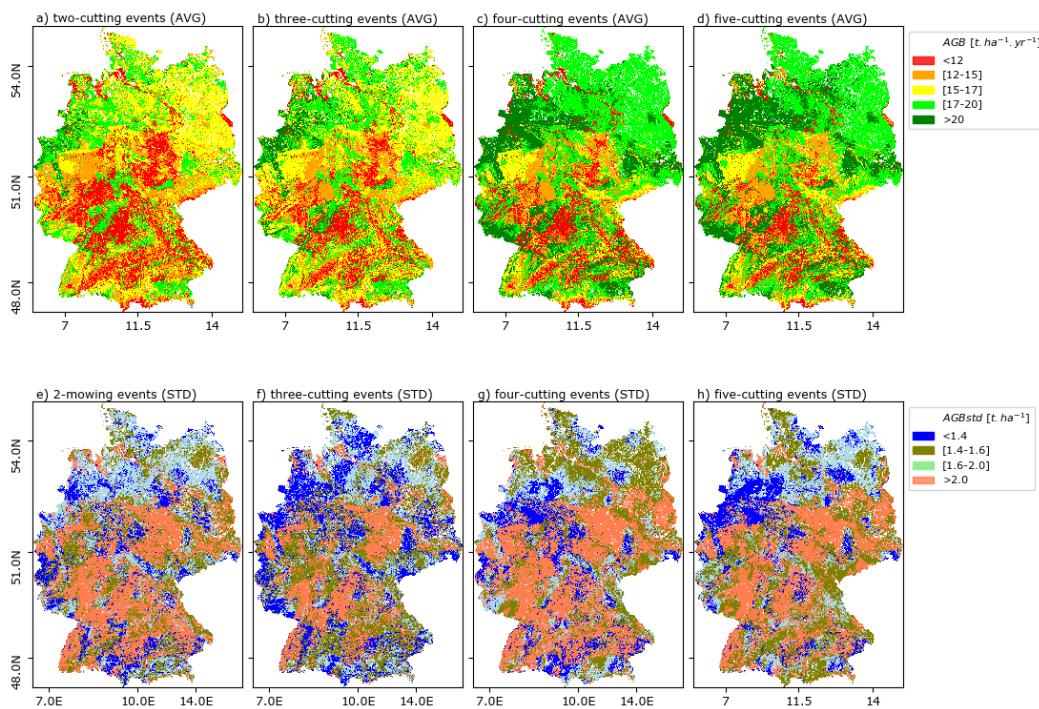


Fig. 9: Spatial distribution of average (AVG) grassland productivity during the period 1990–2018 for a) two, b) three, c) four, and d) five cutting events per growing seasons. The average was obtained from 25 parameter sets within 95PPU uncertainty ranges; (e–h) spatial distribution of the standard deviation (STD) of yearly grassland productivity showing temporal variability for e) two, f) three, g) four, and h) five cutting events per growing seasons

The temporal dynamics of grassland productivity showed considerable variation from one year to the other (Figs. S.5-8). For all cutting regimes, the lowest productivity occurred in the years 1991, 1992, 2010, and 2013, when the seasonal amount of precipitation also was lowest (Fig. S.4). The highest productivity was seen between 2000 and 2002, mostly in western and southern Germany, as well as in 2007. The standard deviation (STD) of the time series over the 27 years of the study period (1991–2018) was calculated at the grid level to quantify the temporal variability of these grasslands' productivity. The results also showed a higher STD in intensively managed grasslands compared to extensively managed grasslands. The STDs of grassland productivity were lower in western ($<1.4 \text{ t ha}^{-1}$) compared with eastern Germany ($>2.0 \text{ t ha}^{-1}$).

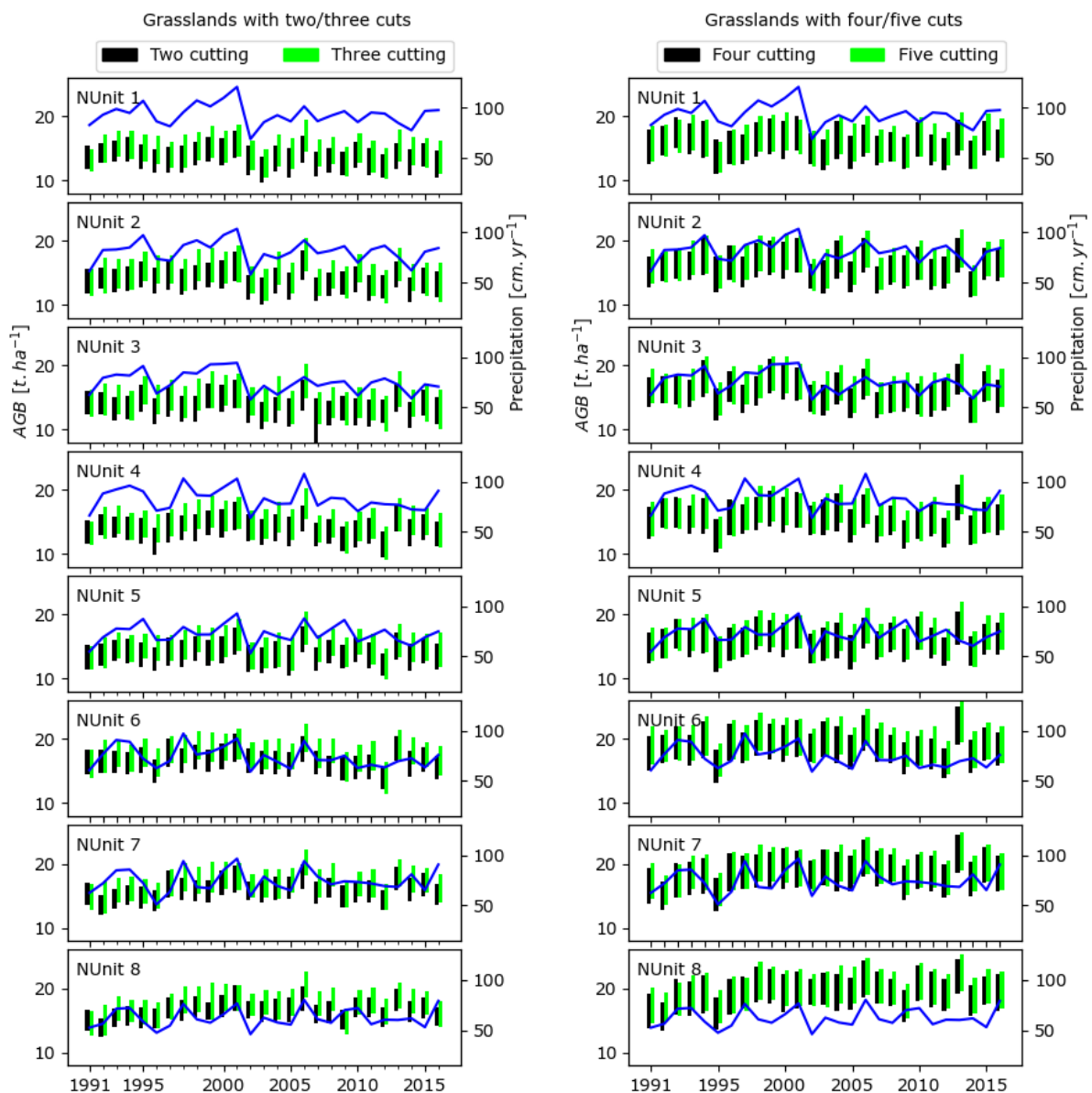


Fig. 10. Comparison of the simulated 95PPU intervals of grassland productivity in eight nature units (Fig. 1) during the period 1990–2017. The band interval shows the uncertainty derived from the 95PPU calculations. The left (right) panels show the results for grassland with two/three (four/five) cuts. The blue line shows the yearly precipitation.

We also quantified the yearly 95PPU ranges of the grass productivity of different types of grasslands with different numbers of cutting regims in eight nature units (NUnits 1–8) across Germany (Fig. 10). Comparing productivity across these units revealed that NUnit 6 and NUnit 8 had the highest values, and NUnit 3 and NUnit 8 had the lowest compared with the others. We also found that the uncertainty of simulated grassland productivity (i.e. the upper and lower values of *AGB* obtained from the 95PPU uncertainty band) was remarkably larger in grasslands with four/five cutting events compared with in grasslands with two/three cutting events. The difference between the lower and upper uncertainty bounds of grassland productivity increased to 4.0 t ha⁻¹ in grasslands with four/five cutting events, whereas this number was a maximum of 2.0 t ha⁻¹ in extensively managed grasslands.

4. Discussion

4.1. Effects of the sensitivity analysis

In this paper, we applied a two-step sensitivity analysis procedure, i.e. screening and time-varying ranking, which the literature has recommended as a systematic approach to sensitivity analysis (Pianosi et al., 2016). The time-varying ranking of parameters provided useful information for the prioritization of parameters to be estimated. Sensitivity analysis is a function of boundary conditions as well as time (Paleari and Confalonieri, 2016). Ignoring this fact can result in the misinterpretation of parameters. This is an important finding, which, although not new, still receives too little attention in many studies (Herman et al., 2013; Zhan et al., 2013), especially in grassland modeling studies. The temporal discretization of sensitive indices allowed for tackling the parameters which were influential for certain periods within the growing season and provided deeper insights on the model's temporal behavior. Defining the most sensitive parameters of *AGB* or *SM* at only one certain stage of the growth season (e.g. the end of the season or the vegetation stage) may result in misleading results, because the sensitivity of some parameters may vary over the course of a growing season (Lamboni et al., 2009). In our case, we found that *MaximumAssimilatioRate* and *OrganScenesenceRate* played key roles at the end of every growing season.

Our results correspond with previous investigations from Specka et al. (2015) for the same model, which showed that *SpecificLeafAreaIndex* was the most influential parameter for *AGB* and *LAI*; the leaf is the key organ for light interception and energy absorption (Wang et al., 2013). However, their ranking of common parameters in the study of Specka et al. (2015) and ours differed for two reasons. First, unlike their study, which focused primarily on annual crops, we focused on perennial grasslands. Moreover, our study included additional processes that described grassland dynamics in the new version of MONICA, as well as parameters such as *StageMobilFromStorageOrgan*. Therefore, our results provided base information on model parameterization for the future application of MONICA when simulating grasslands.

4.2. Insights concerning multi-objective calibration in the context of uncertainty

The performance of grassland models in terms of simulating different variables have been investigated in literature in which the SO calibration approach has been applied for parameter estimation (Höglind et al., 2016; Korhonen et al., 2018). However, these studies did not take into account the trade-off of grassland model performances in terms of simulating multiple variables (i.e. objectives) together with uncertainty analysis (Höglind et al., 2020; Persson et al., 2019), which is an essential concept in uncertainty-based MO calibration. Those studies which presented results in terms of simulating multiple variables lacked additional information on the uncertainty ranges of model output variables. This study is one of the first attempt to consider both aspects for simulating grasslands. Our results for example revealed that, when taking an SO calibration approach, improving one variable will occur at the expense of compromising another variable. This means that the SO calibration has a limited capacity to represent the multi-response dynamics in conceptual agro-ecosystem models; model calibration systems that take into consideration multiple components of interconnecting soil, plant, and water processes are required. This corroborates the findings of Groh et al. (2018), who found similar behaviors when identifying soil hydraulic parameters based on various observations of soil-related state variables. The traditional MO-Pareto calibration improved model performance in terms of the three target variables, but also highlighted the weakness of the method in terms of quantifying the uncertainty ranges of parameters. Therefore, parameter optimization procedures should incorporate uncertainty into a model's parameters (Tao et al., 2018), which is important when simulating natural processes to measure things such as grassland productivity over a large range of soil–climate situations – as our study did for Germany.

We therefore proposed a novel uncertainty-based multi-objective calibration approach i.e. MO-SUF12, which not only accounted for multiple variables, but also measured the associated uncertainty for each of them. The method searches parameter space in two directions. The first direction (the Pareto-along search in Fig. 3b) found non-dominated solutions located on the Pareto front, efficiently exploring the search space in light of different target variables. The second search, which was orthogonal to the first, ranked the previously discovered Pareto fronts and exploited those fronts which fulfilled the defined thresholds for the *p-factor* and the *r-factor*. Optimization methods based on ranking Pareto fronts have been used in other fields such as control systems (Ma and Qiu, 2014; Mittal and Singh, 2015), but to the best of our knowledge, have never been applied in the context of environmental studies. Here, we used this novel insight to calibrate agro-ecosystem models that work with water and vegetation variables; our goal is to improve the modeling results of future researchers and nudge the parameters of such models in a more meaningful direction. Overall, the comparison of the two MO approaches showed that they do not necessarily conflict with each other, but MO-SUF12 provides complementary information that facilitates a better estimation of model parameters' uncertainty.

Using our novel MO calibration approach, we found the reasonable parameter ranges for the most sensitive parameters. The higher value of *SenescenceRate* for root organs value compared with all other organs is consistent with the results generated by Malik et al. (2018), who adopted a constant 5%

SenescenceRate for root organ throughout the growing season. Other organs (root, leaf, and stolon) all had a lower *SenescenceRate* (<0.02) during Stage 2 to Stage 5. Only Stage 6 (which is defined as the stage after maturity) showed high values for all organs. Overall, the ranges obtained for *SenescenceRate* during each stage were consistent with the natural growth of grass, demonstrating that MO-SUFI2 performed satisfactorily in terms of capturing parameter uncertainty.

4.3. Implications of uncertainty-based multi-objective calibration for Germany-wide simulations

Information on grassland productivity or related environmental impacts is not currently available at high spatiotemporal resolution for Germany. Process-based simulation models can be used to fill this gap and deliver such information. However, since the required input for driving models at this scale is still fragmented, a range of assumptions need to be made prior to the simulation; this introduces uncertainty into the simulation's results. One of the main sources of uncertainty comes from the simulation model itself. Process-based simulation models such as MONICA have been designed so as to reproduce the complex interactions of ecosystem processes, but their reliability depends on estimating a large number of parameters that describe different processes embedded in the model. With a limited set of observations available to parameterize and calibrate a model for all the different shades of managed and semi-natural grassland ecosystems that contribute to grass production for fodder or biomass, the calibration process requires special attention. So far, only a few detailed experiments are available. In most cases, experiments including observations on both water and nutrient cycles are even rarer, which makes it challenging to set up a model in a way that supports our understanding of how these unique ecosystems respond to changes in management and environmental conditions.

Despite all these existing limitations, however, quantifying model uncertainty is an essential element of more robust and reliable decision-making regarding grasslands. Resorting to methods that separately quantify different sources of model uncertainty, such as climate data (Parkes et al., 2019), soil data (Folberth et al., 2016), model structure (Tao et al., 2018), or spatial scale (Ojeda et al., 2020), is unwieldy, and sometimes completely infeasible at a large scale due to the lack of sufficient data for model validation. The advantage of this study is that the MO-SUFI2 method accounts for all sources of uncertainty expressed in terms of the *p-factor* and the *r-factor*, and can ultimately find the ranges of these parameters. The ranges the model obtained helps to characterize the uncertainty of key outputs of the model, e.g. *AGB*, under various cutting regimes.

A comparison of the four cutting regimes shows relatively similar spatial patterns, with a discernably high level of productivity in northwest Germany and low values in central Germany. This corroborates the results of Smit et al. (2008), where the largest amount of production was seen in the northwest part of Germany. It also demonstrates that MONICA is able to capture the spatial dynamics of grassland productivity. However, the magnitudes of our productivity estimates are not directly comparable with the study of Smit et al. (2008), as their analysis was based on a combination of statistics and remote

sensing-based variables. This is not comparable with our maps, which are based on simulations of biophysical processes. It is worth mentioning that our simulations did not take into account the spatial conditions of management treatments, such as cutting or fertilization time, amount of fertilizer, or grassland species sown in the field, due to a lack of data at such resolution. Given the fact that similar management treatments are assigned to all grids in our simulation, some discrepancies with the literature and measurement values are expected.

Among the many factors which might have contributed to this remarkable degree of uncertainty, one important fact is related to the lack of information on accurate cutting events in each grid, which varied from one year to the next. Due to a lack of data at such detailed spatial and temporal levels for our simulations, we assumed the same cutting dates for every year. This has a considerable impact on model simulations. The more frequent cutting events in four/five cuts grasslands resulted in larger uncertainty in the model prediction. Our paper has provided an initial model set-up as a baseline for considering multiple variables simultaneously during model calibration.

4. Conclusion and future prospect

We have presented the first simulations of grasslands across with number of cutting events per growing season the whole area of Germany. To the best of our knowledge, there is no study available that compares grassland productivity in light of different numbers of cutting regimes over such a large territory. We have concluded that, to obtain reliable estimates on grassland productivity, model calibration based on both vegetation and soil-related components is essential, and should be more prominently used in future studies. Further improvement of grassland simulations across large areas requires (a) detailed information on cutting or grazing regimes and their spatial distribution and (b) information on grassland species composition, which both could potentially be provided through remote sensing (Griffiths et al., 2020; Lyu et al., 2020). Also here, strong interdependencies prevail, as environmental conditions govern the emergence of different species communities, each of which then inheres specific productivity levels and response patterns. Multi-objective calibration as presented here will therefore also be essential for grassland models that include species composition dynamics. Combining process-based grassland models for productivity and related water and nutrient fluxes with species composition dynamics will enable addressing a larger variety of grassland production systems, and related environmental responses in simulations, offering an even more realistic representation of regional grassland production.

5. Acknowledgements

This work was supported by the innovation funding program of the German Federal Ministry of Food and Agriculture (BMEL; Project No. 2818300716).

6. References

- Abbaspour, K.C., 2015. SWAT-CUP; SWAT calibration and uncertainty programs - A user manual.
- Abbaspour, K.C., Rouholahnejad, E., Vaghefi, S., Srinivasan, R., Yang, H., Klove, B., 2015. A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. *J Hydrol.* 524, 733-752. <https://doi.org/10.1016/j.jhydrol.2015.03.027>.
- Agrometeorological Research Center (ZAMF), Germany, https://www.dwd.de/EN/aboutus/locations/branch_offices/braunschweig.html.
- Archontoulis, S.V., Castellano, M.J., Licht, M.A., Nichols, V., Baum, M., Huber, I., Martinez-Feria, R., Puntel, L., Ordonez, R.A., Iqbal, J., Wright, E.E., Dietzel, R.N., Helmers, M., Vanloocke, A., Liebman, M., Hatfield, J.L., Herzmann, D., Cordova, S.C., Edmonds, P., Togliatti, K., Kessler, A., Danalatos, G., Pasley, H., Pederson, C., Lamkey, K.R., 2020. Predicting crop yields and soil-plant nitrogen dynamics in the US Corn Belt. *Crop Sci.* 60(2), 721-738. <https://doi.org/10.1002/csc2.20039>.
- BFN, 2014. GRÜNLAND-Report. Alles im Grünen Bereich? Federal Agency for Nature Conservation: Grassland conservation in Germany. <https://www.bfn.de/en/activities/agriculture/grassland-conservation-in-germany.html>.
- BGR, 1995. Institute for Geosciences and Natural Resources, <https://www.bgr.bund.de/DE/Themen/Boden/Projekte/Informationsgrundlagen-laufend/BUEK200/BUEK200.html>.
- Chang, J.F., Ciais, P., Viovy, N., Vuichard, N., Sultan, B., Soussana, J.F., 2015a. The greenhouse gas balance of European grasslands. *Global Change Biol.* 21(10), 3748-3761. <https://doi.org/10.1111/gcb.12998>.
- Chang, J.F., Viovy, N., Vuichard, N., Ciais, P., Campioli, M., Klumpp, K., Martin, R., Leip, A., Soussana, J.F., 2015b. Modeled changes in potential grassland productivity and in grass-fed ruminant livestock density in Europe over 1961-2010. *Plos One.* 10(5). <https://doi.org/10.1371/journal.pone.0127554>.
- Confalonieri, R., Bellocchi, G., Bregaglio, S., Donatelli, M., Acutis, M., 2010. Comparison of sensitivity analysis techniques: A case study with the rice model WARM. *Ecological Modelling.* 221(16), 1897-1906. <https://doi.org/10.1016/j.ecolmodel.2010.04.021>.
- Dass, P., Houlton, B.Z., Wang, Y.P., Warlind, D., 2018. Grasslands may be more reliable carbon sinks than forests in California. *Environ Res Lett.* 13(7). <https://doi.org/ARTN07402710.1088/1748-9326/aacb39>.
- DWD, 1994. German Weather Service: <https://www.dwd.de/>.
- Efstratiadis, A., Koutsoyiannis, D., 2010. One decade of multi-objective calibration approaches in hydrological modelling: a review. *Hydrolog Sci J.* 55(1), 58-78. <https://doi.org/10.1080/02626660903526292>.

789 Fenicia, F., Savenije, H.H.G., Matgen, P., Pfister, L., 2007. A comparison of alternative multiobjective
790 calibration strategies for hydrological modeling. *Water Resour Res.* 43(3).
791 <https://doi.org/W0343410.1029/2006wr005098>.

792 Folberth, C., Skalsky, R., Moltchanova, E., Balkovic, J., Azevedo, L.B., Obersteiner, M., van der
793 Velde, M., 2016. Uncertainty in soil data can outweigh climate impact signals in global crop yield
794 simulations. *Nat Commun.* 7. <https://doi.org/ARTN1187210.1038/ncomms11872>.

795 Fonseca, C.M., Fleming, P.J., 1993. Multiobjective genetic algorithms, IEE colloquium on ‘Genetic
796 Algorithms for Control Systems Engineering. IEE, London, UK

797 Gomara, I., Bellocchi, G., Martin, R., Rodriguez-Fonseca, B., Ruiz-Ramos, M., 2020. Influence of
798 climate variability on the potential forage production of a mown permanent grassland in the French
799 Massif Central. *Agr Forest Meteorol.* 280.
800 <https://doi.org/UNSP10776810.1016/j.agrformet.2019.107768>.

801 Graux, A.I., Gaurut, M., Agabriel, J., Baumont, R., Delagarde, R., Delaby, L., Soussana, J.F., 2011.
802 Development of the Pasture Simulation Model for assessing livestock production under climate
803 change. *Agr Ecosyst Environ.* 144(1), 69-91. <https://doi.org/10.1016/j.agee.2011.07.001>.

804 Griffiths, P., Nendel, C., Pickert, J., Hostert, P., 2020. Towards national-scale characterization of
805 grassland use intensity from integrated Sentinel-2 and Landsat time series. *Remote Sensing of*
806 *Environment.* 238. <https://doi.org/ARTN1112410.1016/j.rse.2019.03.017>.

807 Groh, J., Stumpp, C., Lücke, A., Pütz, T., Vanderborght, J., Vereecken, H., 2018. Inverse estimation of
808 soil hydraulic and transport parameters of layered soils from water stable Isotope and lysimeter data.
809 *Vadose Zone J.* 17(1). <https://doi.org/UNSP17016810.2136/vzj2017.09.0168>.

810 Herman, J.D., Reed, P.M., Wagener, T., 2013. Time-varying sensitivity analysis clarifies the effects of
811 watershed model formulation on model behavior. *Water Resour Res.* 49(3), 1400-1414.
812 <https://doi.org/10.1002/wrcr.20124>.

813 Höglind, M., Cameron, D., Persson, T., Huang, X., van Oijen, M., 2020. BASGRA_N: A model for
814 grassland productivity, quality and greenhouse gas balance. *Ecological Modelling.* 417.

815 Höglind, M., Van Oijen, M., Cameron, D., Persson, T., 2016. Process-based simulation of growth and
816 overwintering of grassland using the BASGRA model. *Ecological Modelling.* 335, 1-15.
817 <https://doi.org/10.1016/j.ecolmodel.2016.04.024>.

818 Houska, T., Kraft, P., Liebermann, R., Klatt, S., Kraus, D., Haas, E., Santabarbara, I., Kiese, R.,
819 Butterbach-Bahl, K., Müller, C., Breuer, L., 2017. Rejecting hydro-biogeochemical model structures
820 by multi-criteria evaluation. *Environ Modell Softw.* 93, 1-12.

821 Jones, M.B., Donnelly, A., 2004. Carbon sequestration in temperate grassland ecosystems and the
822 influence of management, climate and elevated CO₂. *New Phytol.* 164(3), 423-439.
823 <https://doi.org/10.1111/j.1469-8137.2004.01201.x>.

824 Kamali, B., Abbaspour, K.C., Lehmann, A., Wehrli, B., Yang, H., 2018. Uncertainty-based auto-
825 calibration for crop yield – the EPIC⁺ procedure for a case study in Sub-Saharan Africa. *Eur J Agron.*
826 19, 57–72. <https://doi.org/10.1016/j.eja.2017.10.012>.

827 Kamali, B., Mousavi, S.J., Abbaspour, K.C., 2013. Automatic calibration of HEC-HMS using single-
828 objective and multi-objective PSO algorithms. *Hydrol Process.* 27(26), 4028-4042.
829 <https://doi.org/10.1002/hyp.9510>.

830 Kennedy, J., Eberhart, R., 1995. Particle swarm optimization, in: Springer (Ed.) In Proceedings of
831 IEEE international conference on neural network IV Piscataway, pp. 1942–1948.

832 Kersebaum, K.C., 2007. Modelling nitrogen dynamics in soil-crop systems with HERMES. *Nutr Cycl*
833 *Agroecosys.* 77(1), 39-52. <https://doi.org/10.1007/s10705-006-9044-8>.

834 Korhonen, P., Palosuo, T., Persson, T., Höglind, M., Jégo, G., Van Oijen, M., Gustavsson, A.-M.,
835 Bélanger, G., Virkajärvi, P., 2018. Modelling grass yields in northern climates – a comparison of three
836 growth models for timothy. *Field Crops Research.* 224, 37-47.
837 <https://doi.org/https://doi.org/10.1016/j.fcr.2018.04.014>.

838 Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. *Science.*
839 304(5677), 1623-1627. <https://doi.org/10.1126/science.1097396>.

840 Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., Monod, H., 2009. Multivariate global
841 sensitivity analysis for dynamic crop models. *Field Crops Research.* 113(3), 312-320.
842 <https://doi.org/10.1016/j.fcr.2009.06.007>.

843 Lei, T.J., Pang, Z.G., Wang, X.Y., Li, L., Fu, J., Kan, G.Y., Zhang, X.L., Ding, L.Q., Li, J.R., Huang,
844 S.F., Shao, C.L., 2016. Drought and carbon cycling of grassland ecosystems under global change: A
845 review. *Water-Sui.* 8(10). <https://doi.org/46010.3390/w8100460>.

846 Lyu, X., Li, X., Dang, D., Dou, H., Xuan, X., Liu, S., Li, M., Gong, J., 2020. A new method for
847 grassland degradation monitoring by vegetation species composition using hyperspectral remote
848 sensing. *Ecol Indic.* 114. <https://doi.org/https://doi.org/10.1016/j.ecolind.2020.106310>.

849 Ma, Y.Q., Qiu, H.X., 2014. Integrated Multiobjective Optimal Design for Active Control System
850 Based on Genetic Algorithm. *Math Probl Eng.* 2014. <https://doi.org/74823710.1155/2014/748237>.

851 Malik, W., Boote, K.J., Hoogenboom, G., Cavero, J., Dechmi, F., 2018. Adapting the CROPGRO
852 model to simulate Alfalfa growth and yield. *Agron J.* 110(5), 1777-1790.
853 <https://doi.org/10.2134/agronj2017.12.0680>.

854 Mitchell, R.A.C., Lawlor, D.W., Mitchell, V.J., Gibbard, C.L., White, E.M., Porter, J.R., 1995. Effects
855 of elevated CO₂ concentration and increased temperature on winter-wheat - test of Arcwheat1
856 simulation-model. *Plant Cell Environ.* 18(7), 736-748. <https://doi.org/10.1111/j.1365-3040.1995.tb00576.x>.

858 Mittal, E., Singh, S., 2015. Multiobjective Gain-Impedance Optimization of Yagi-Uda Antenna
859 Design Using Different BBO Migration Variants. *Appl Artif Intell.* 29(1), 33-48.
860 <https://doi.org/10.1080/08839514.2014.962280>.

861 Morris, M.D., 1991. Factorial sampling plans for preliminary computational experiments.
862 *Technometrics.* 33(2), 161-174. <https://doi.org/10.2307/1269043>.

863 Nendel, C., Berg, M., Kersebaum, K.C., Mirschel, W., Specka, X., Wegehenkel, M., Wenkel, K.O.,
864 Wieland, R., 2011. The MONICA model: Testing predictability for crop growth, soil moisture and
865 nitrogen dynamics. *Ecological Modelling.* 222(9), 1614-1625.
866 <https://doi.org/10.1016/j.ecolmodel.2011.02.018>.

867 Nendel, C., Kersebaum, K.C., 2004. A simple model approach to simulate nitrogen dynamics in
868 vineyard soils. *Ecological Modelling.* 177(1-2), 1-15.
869 <https://doi.org/10.1016/j.ecolmodel.2004.01.014>.

870 Nendel, C., Kersebaum, K.C., Mirschel, W., Wenkel, K.O., 2014. Testing farm management options
871 as climate change adaptation strategies using the MONICA model. *Eur J Agron.* 52, 47-56.
872 <https://doi.org/10.1016/j.eja.2012.09.005>.

873 Obermeier, W.A., Lehnert, L.W., Kammann, C.I., Muller, C., Grunhage, L., Luterbacher, J., Erbs, M.,
874 Moser, G., Seibert, R., Yuan, N., Bendix, J., 2017. Reduced CO₂ fertilization effect in temperate C3
875 grasslands under more extreme weather conditions. *Nat Clim Change.* 7(2), 137-141.
876 <https://doi.org/10.1038/Nclimate3191>.

877 Ojeda, J.J., Rezaei, E.E., Remenyi, T.A., Webb, M.A., Webber, H.A., Kamali, B., Harris, R.M.B.,
878 Brown, J.N., Kidd, D.B., Mohammed, C.L., Siebert, S., Ewert, F., Meinke, H., 2020. Effects of soil-
879 and climate data aggregation on simulated potato yield and irrigation water requirement. *Science of*
880 *the Total Environment.* 710. <https://doi.org/10.1016/j.scitotenv.2019.135589>.

881 Paleari, L., Confalonieri, R., 2016. Sensitivity analysis of a sensitivity analysis: We are likely
882 overlooking the impact of distributional assumptions. *Ecological Modelling.* 340, 57-63.
883 <https://doi.org/10.1016/j.ecolmodel.2016.09.008>.

884 Parkes, B., Higginbottom, T.P., Hufkens, K., Ceballos, F., Kramer, B., Foster, T., 2019. Weather
885 dataset choice introduces uncertainty to estimates of crop yield responses to climate variability and
886 change. *Environ Res Lett.* 14(12). <https://doi.org/ARTN12408910.1088/1748-9326/ab5ebb>.

887 Persson, T., Höglind, M., Van Oijen, M., Korhonen, P., Palosuo, T., Jégo, G., Virkajärvi, P., Bélanger,
888 G., Gustavsson, A.-M., 2019. Simulation of timothy nutritive value: A comparison of three process-
889 based models. *Field Crops Research.* 231.

890 Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B., Wagener, T., 2016.
891 Sensitivity analysis of environmental models: A systematic review with practical workflow. *Environ*
892 *Modell Softw.* 79, 214-232. <https://doi.org/10.1016/j.envsoft.2016.02.008>.

893 Proximedia, October 8, 2018. Pareto Front, www.cenaero.be.

894 Pütz, T., Kiese, R., Wollschläger, U., Groh, J., Rupp, H., Zacharias, S., Priesack, E., Gerke, H.H.,
895 Gasche, R., Bens, O., Borg, E., Baessler, C., Kaiser, K., Herbrich, M., Munch, J.C., Sommer, M.,
896 Vogel, H.J., Vanderborght, J., Vereecken, H., 2016. TERENO-SOILCan: a lysimeter-network in
897 Germany observing soil processes and plant diversity influenced by climate change. *Environ Earth*
898 *Sci.* 75(18), 1242. <https://doi.org/10.1007/s12665-016-6031-5>.

899 Sandor, R., Ehrhardt, F., Brilli, L., Carozzi, M., Recous, S., Smith, P., Snow, V., Soussana, J.F.,
900 Dorich, C.D., Fuchs, K., Fitton, N., Gongadze, K., Klumpp, K., Liebig, M., Martin, R., Merbold, L.,
901 Newton, P.C.D., Rees, R.M., Rolinski, S., Bellocchi, G., 2018. The use of biogeochemical models to
902 evaluate mitigation of greenhouse gas emissions from managed grasslands. *Science of the Total*
903 *Environment.* 642, 292-306. <https://doi.org/10.1016/j.scitotenv.2018.06.020>.

904 Schadler, M., Buscot, F., Klotz, S., Reitz, T., Durka, W., Bumberger, J., Merbach, I., Michalski, S.G.,
905 Kirsch, K., Remmler, P., Schulz, E., Auge, H., 2019. Investigating the consequences of climate change
906 under different land-use regimes: a novel experimental infrastructure. *Ecosphere.* 10(3).
907 <https://doi.org/10.1002/ecs2.2635>.

908 Smit, H.J., Metzger, M.J., Ewert, F., 2008. Spatial distribution of grassland productivity and land use
909 in Europe. *Agr Syst.* 98(3), 208-219. <https://doi.org/10.1016/j.agsy.2008.07.004>.

910 Sobol, I.M., 1967. Distribution of points in a cube and approximate evaluation of integrals. *U.S.S.R*
911 *Comput. Maths. Math. Phys.* (7), 86–112. [https://doi.org/10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9).

912 Specka, X., Nendel, C., Wieland, R., 2015. Analysing the parameter sensitivity of the agro-ecosystem
913 model MONICA for different crops. *Eur J Agron.* 71, 73-87. <https://doi.org/10.1016/j.eja.2015.08.004>.

914 Specka, X., Nendel, C., Wieland, R., 2019. Temporal sensitivity analysis of the MONICA model:
915 Application of two global approaches to analyze the dynamics of parameter sensitivity. *Agriculture-*
916 *Basel.* 9(2). <https://doi.org/ARTN3710.3390/agriculture9020037>.

917 Statistisches Bundesamt, 2019. Statistisches Jahrbuch in: Kap. 19 Land- und Forstwirtschaft (Ed.).
918 Wiesbaden, p. p.492.

919 Stella, T., Mouratiadou, I., Gaiser, T., Berg-Mohnicke, M., Wallor, E., Ewert, F., Nendel, C., 2019.
920 Estimating the contribution of crop residues to soil organic carbon conservation. *Environ Res Lett.*
921 14(9). <https://doi.org/ARTN09400810.1088/1748-9326/ab395c>.

922 Tang, Y.T., Marshall, L., Sharma, A., Ajami, H., 2018. A Bayesian alternative for multi-objective
923 ecohydrological model specification. *J Hydrol.* 556, 25-38.
924 <https://doi.org/10.1016/j.jhydrol.2017.07.040>.

925 Tao, F.L., Rotter, R.P., Palosuo, T., Diaz-Ambrona, C.G.H., Minguez, M.I., Semenov, M.A.,
926 Kersebaum, K.C., Nendel, C., Specka, X., Hoffmann, H., Ewert, F., Dambreville, A., Martre, P.,
927 Rodriguez, L., Ruiz-Ramos, M., Gaiser, T., Hohn, J.G., Salo, T., Ferrise, R., Bindi, M., Cammarano,
928 D., Schulman, A.H., 2018. Contribution of crop model structure, parameters and climate projections to
929 uncertainty in climate change impact assessments. *Global Change Biol.* 24(3), 1291-1307.
930 <https://doi.org/10.1111/gcb.14019>.

931 von Wilpert, K., Hartmann, P., Puhlmann, H., Schmidt-Walter, P., Meesenburg, H., Müller, J., Evers,
932 J., 2016. Bodenwasserhaushalt und Trockenstress. *Thünen Rep.* 43(343-386).

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

936 White, R., Murray, S., Rohweder, M., 2000. Pilot analysis of global ecosystems: grassland ecosystems
937 technical report, in: Institute, W.R. (Ed.). Washington, DC.

938 Wimberly, M.C., Janssen, L.L., Hennessy, D.A., Luri, M., Chowdhury, N.M., Feng, H.L., 2017.
939 Cropland expansion and grassland loss in the eastern Dakotas: New insights from a farm-level survey.
940 *Land Use Policy.* 63, 160-173. <https://doi.org/10.1016/j.landusepol.2017.01.026>.

941 Wöhling, T., Gayler, S., Priesack, E., Ingwersen, J., Wizemann, H.D., Högy, P., Cuntz, M., Attinger,
942 S., Wulfmeyer, V., Streck, T., 2013. Multiresponse, multiobjective calibration as a diagnostic tool to
943 compare accuracy and structural limitations of five coupled soil-plant models and CLM3.5. *Water*
944 *Resour Res.* 49(12), 8200-8221. <https://doi.org/10.1002/2013wr014536>.

945 Wöhling, T., Schöniger, A., Gayler, S., Nowak, W., 2015. Bayesian model averaging to explore the
946 worth of data for soil-plant model selection and prediction. *Water Resour Res.* 51(4), 2825-2846.
947 <https://doi.org/10.1002/2014wr016292>.

948 Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic
949 model. *J. Hydrol.* 204, 83-97. [https://doi.org/10.1016/S0022-1694\(97\)00107-8](https://doi.org/10.1016/S0022-1694(97)00107-8).

950 Zhan, C.S., Song, X.M., Xia, J., Tong, C., 2013. An efficient integrated approach for global sensitivity
951 analysis of hydrological model parameters. *Environ Modell Softw.* 41, 39-52.
952 <https://doi.org/10.1016/j.envsoft.2012.10.009>.

953 Zhao, G., Siebert, S., Enders, A., Rezaei, E.E., Yan, C.Q., Ewert, F., 2015. Demand for multi-scale
954 weather data for regional crop modeling. *Agr Forest Meteorol.* 200, 156-171.
955 <https://doi.org/10.1016/j.agrformet.2014.09.026>.

956 Zheng, Z.T., Zhu, W.Q., Zhang, Y.J., 2020. Seasonally and spatially varied controls of climatic factors
957 on net primary productivity in alpine grasslands on the Tibetan Plateau. *Glob Ecol Conserv.* 21.
958 <https://doi.org/ARTN0081410.1016/j.gecco.2019.e00814>.

959