



HAL
open science

Efficient Bayesian automatic calibration of a functional-structural wheat model using an adaptive design and a metamodeling approach

Emmanuelle Blanc, Jérôme Enjalbert, Timothée Flutre, Pierre Barbillon

► To cite this version:

Emmanuelle Blanc, Jérôme Enjalbert, Timothée Flutre, Pierre Barbillon. Efficient Bayesian automatic calibration of a functional-structural wheat model using an adaptive design and a metamodeling approach. *Journal of Experimental Botany*, In press, 10.1093/jxb/erad339 . hal-04190551

HAL Id: hal-04190551

<https://hal.science/hal-04190551>

Submitted on 29 Aug 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Type of article: Research paper

Title: Efficient Bayesian automatic calibration of a functional-structural wheat model using an adaptive design and a metamodeling approach

Authors: Emmanuelle Blanc^{1*}, Jérôme Enjalbert¹, Timothée Flutre¹, Pierre Barbillon^{2*}

¹Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France.

²Université Paris-Saclay, AgroParisTech, INRAE, UMR MIA-Paris, 75005, Paris, France

Running title: Calibration of a 3D wheat model using an adaptive design and metamodels

Number of words: 9404

Number of tables: 1

Number of figures: 7

Corresponding author

Pierre Barbillon

pierre.barbillon@agroparistech.fr

Highlight

To estimate efficiently the unknown parameters of a costly functional structural plant model (FSPM) from experimental data, we propose an original algorithm that improves the Bayesian calibration of such models by using sequential evaluations.

Abstract

Functional-structural plant models are increasingly being used by plant scientists to address a wide variety of questions. However, the calibration of these complex models is often challenging, mainly because of their high computational cost, and, as a result, error propagation is usually ignored. In this paper, we applied an automatic method to the calibration of WALTer: a functional-structural wheat model that simulates the plasticity of tillering in response to competition for light. We used a Bayesian calibration method to jointly estimate the values of 5 parameters and quantify their uncertainty by fitting the model outputs to tillering dynamics data. We made recourse to Gaussian process metamodels in order to alleviate the computational cost of WALTer. These metamodels are built from an adaptive design that consists of successive runs of WALTer chosen by an Efficient Global Optimisation algorithm specifically adapted to this particular calibration task. The method presented here performed well on both synthetic and experimental data. It is an efficient approach for the calibration of WALTer and should be of interest for the calibration of other functional-structural plant models.

Keywords: adaptive design, Bayesian calibration, Efficient Global Optimisation, Functional Structural Plant Model, Gaussian process metamodel, Kriging metamodel, tillering, *Triticum aestivum*, wheat, uncertainty quantification

INTRODUCTION

Modeling is a powerful tool to study complex systems. Numerical experiments are often used when real experiments are too difficult or too expensive to carry out. Furthermore, computer models allow to strictly control the conditions of the experiments and to explore conditions beyond what can be covered experimentally. The advantages of computer models have made them essential tools in many disciplines. For example, modeling is widely used in plant sciences and agronomy and has been identified as a promising tool to tackle some of the challenges associated with major issues such as food security (Christensen et al., 2018). In particular, functional-structural plant models (FSPM; Godin and Sinoquet, 2005; Vos et al., 2010) have been more and more used since the 1990s, and they have become a major research subject (Guo et al., 2011; Sievänen et al., 2014; Evers et al., 2018). These individual-based models explicitly represent the plant architecture and integrate knowledge in ecophysiology and developmental biology. Thus, FSPM take into account the complex interactions between environmental factors, plant development, plant architecture and the underlying physiological processes. These models allow the integration of different scales (from gene to community) to study complex systems in plant science (Louarn and Song, 2020). Thereby, FSPM have been used to study a wide variety of species, both wild and cultivated, and a broad range of questions. Examples of application include prediction of fire behaviour in tree crowns (Parsons et al., 2011); study of the evolutionary emergence of life history strategies along gradients of stress intensity and disturbance frequency (Bornhofen et al., 2011); and study of the impact of plant height on the control of rain-borne diseases in wheat cultivar mixtures (Vidal et al., 2018). Along with the diversity of their objects of study, FSPM also exhibit significant diversity in their construction and in the modeling strategy. These models often explicitly represent the 3D architecture of the simulated plants (Vos et al., 2010; Evers et al.,

2019), but some authors have developed less detailed approaches, with significantly lower computational cost. Such is, for example, the case of the Ecomeristem model (Luquet et al., 2006), or of the mathematical FSPM GreenLab (de Reffye et al., 2021). These differences in computational cost have a major impact on the use of the models.

In order to obtain reliable results from FSPM, or from any model, their input parameters must be accurately set. Indeed, the quality of the parameter estimation is critical for the quality of the model predictions. Some parameters are known biological traits that can be directly fixed based on the literature or on experimental data. However, some other parameters cannot be directly measured and are not available in the literature. The values of these parameters have to be indirectly estimated via calibration, which means that the overall model has to be fitted to experimental data. The calibration of process-based models, a category to which the FSPM belong along with crop models, is known to be challenging, and many modeling groups rely on different strategies as analysed by Wallach et al (2021). Calibration can be done either manually (i.e., trial and error) or automatically. The latter method relies on the use of a search algorithm to identify the “optimal” values of the specified parameters based on the minimization of the distance between the model outputs and the observed data. Automatic calibration has a number of advantages over manual calibration. Indeed, automatic calibration is less subjective than manual calibration, as its success is less dependent on the experience of the modeller (Muleta and Nicklow, 2005), manual calibration hence is not recommended in the community developing crop models (Wallach et al, 2021). Among practitioners using automatic calibration, most of them target a point-estimate of the parameters (Wallach et al, 2021, for crop models; Mathieu et al, 2018, for FSPMs). Still, when one is interested in comparing parameters fitted on different data sets, uncertainty quantification is needed, similarly for uncertainty propagation (see Chen & Cournède, 2014, in the context of crop models). Such questions require a distribution of likely values of the parameters rather than

looking for point-estimates. Bayesian calibration (Kennedy and O’Hagan, 2001; Higdon et al., 2004; Bayarri et al., 2007) is an automatic calibration method which also provides the modeller with a distribution of values of the parameters that are likely to reproduce the experimental data. Most importantly, beyond providing point estimates of the parameters, it allows to quantify their uncertainty and propagate this uncertainty to predictions. Automatic calibration in a frequentist context (where no distribution is set on the parameters to calibrate) can also provide uncertainty quantification on the parameter values through bootstrap in an additional step (Trevezas & Cournède, 2013; Wong et al, 2017). In both cases, the experimental data and the FSPM are linked in a statistical model from which a likelihood is then derived. In the Bayesian approach, a prior distribution, which may encode some expert or literature information on the parameter values, is chosen. Since the corresponding posterior distribution is not tractable, algorithms such as Markov Chain Monte Carlo (MCMC) are run to obtain a sample from the posterior distribution. However, these algorithms as well as the ones maximizing the likelihood in a frequentist context resort to many runs of the model, which is why, for complex models, including such as FSPM, calibration is often done manually (see Lecarpentier et al., 2019 and Gauthier et al., 2020 for example), and for slightly less complex models such as crop models, calibration most often consists in finding point estimates (Wallach et al, 2021). Indeed, FSPM runs are usually time-consuming and parallel computing of each simulation is complicated because of the interactions between plants in these individual-based models. Some studies have focused on the development of calibration methods for the models GreenLab (Cournède et al., 2011) and Ecomeristem (Larue et al., 2019). However, for 3D FSPM with a significantly higher computational cost, these methods cannot be applied, because they still require too many runs of the model, and new developments are expected (Ribeyre et al., 2018). A solution proposed in the context of Bayesian calibration with a costly computer model is to make recourse to a Gaussian Process

metamodel (Currin et al., 1991; Sacks et al., 1989), also known as Kriging metamodel, which provides not only a fast approximation of the FSPM but also a measure of uncertainty concerning the quality of the approximation of the FSPM. Metamodels are widely used to approximate complex models and have sometimes been used with FSPM (but not for a calibration purpose) to allow analyses that would require an important number of simulations (Perez et al., 2018; Da Silva et al., 2014). However, as metamodels are an approximation of the model, the calibration can be less precise and less efficient than with the actual FSPM. To address this problem, a sequential method aiming at improving the Kriging metamodel precision and based on the EGO algorithm (Jones et al., 1998) can be implemented with respect to the calibration goal, as proposed in Damblin et al. (2018). This method has been used for the calibration of complex models (see Carmassi et al., 2019 for example) but has not yet, to the extent of our knowledge, been applied to FSPM. Moreover, the application of this type of method to the FSPM context is particular since it can require to fit several metamodels in parallel to take into account the complex and / or dynamic outputs of the model.

Here, we present how Kriging metamodels and an adaptive design, based on the EGO algorithm, can be used for the time-efficient and accurate calibration of the WALTer FSPM through an MCMC algorithm in a Bayesian context.

MATERIALS AND METHODS

The efficiency of an automated calibration algorithm to estimate 5 critical parameters of the FSPM WALTer was assessed on both synthetic and experimental datasets.

All the analyses were done using the R software (R Core Team, 2017).

WALTer: a 3D wheat model

WALTer (Lecarpentier et al., 2019) is a functional structural plant model (FSPM) (Vos et al., 2010; Godin & Sinoquet, 2005). This individual-based model simulates the development of the aerial architecture of winter wheat (*Triticum aestivum* L.) from sowing to maturity with a daily time step. In the model, the vegetative development of the plants follows a thermal time schedule and is based on a formalism derived from ADEL-Wheat (Fournier et al., 2003). Thanks to a radiative model (CARIBU: Chelle et al., 1998), WALTer simulates the competition for light between plants and the resulting plasticity of tillering (i.e. the branching ability of grasses): according to competition for light, plants adjust their number of tillers (branching process in monocotyledones). Indeed, the regulation of tillering in the model is based on three simple rules. (i) Each tiller emerges according to an empirically fixed probability (98%). Then, (ii) an early neighbour perception controls the cessation of tillering: plants stop emitting tillers when the surrounding Green Area Index (GAI: ratio of photosynthetic surface to ground area:) reaches a critical value (GAI_c). Finally, (iii) some of the tillers that were emitted regress: a tiller regresses if the amount of photosynthetically active radiation (PAR) it intercepts falls below a threshold (PAR_t). Therefore the 1st rule is introducing a slight stochasticity in tiller emission, independent of the environment, while the 2nd rule makes the tillering cessation earlier when the mean plant and leaf density is higher in the neighborhood of the focal plant. Light interception only plays a role in the 3rd rule, taking into account the shading of organs by the neighboring plants at a fine-scale. During tiller regression, there is a protection period (Δ_{prot}) between the deaths of two successive tillers on a plant. See Table 1 for units of these parameters. These few parameters allow a realistic

simulation of tillering dynamics according to plant density (Lecarpentier et al. 2019) or competition between mixed genotypes (Blanc et al. 2021). A higher plant density results in higher leaf density, acting first on tillering cessation (GAIp), followed by a more intense tiller regression, both contributing to a lower number of mature tillers (i.e. spikes) per plant. Note that despite this regulation, spike density (spike number per m²) increases with sowing density (Lecarpentier et al. 2019).

Table 1. List of the 5 WALTer parameters to be estimated, definition, ranges of variation explored for the calibration, values used to generate the synthetic test-dataset and units.

Parameter	Description	Range	Value for the synthetic data	Unit
GAI_c	Green Area Index threshold above which the emission of tillers stops	0.25 - 1.25	0.75	-
PAR_t	PAR threshold below which a tiller does not survive	$10^5 - 10^6$	3×10^5	$\mu\text{mol} \cdot \text{cm}^{-2} \cdot ^\circ\text{Cd}^{-1}$
Δ_{prot}	Thermal time interval during which two tillers of the same plant cannot die	10 - 100	25	$^\circ\text{Cd}$
N_{MS}^B	Final number of leaves on the main stem	8 - 16	12.5	-
L_{max}^B	Final length of the longest blade of the main stem	8 - 35	16.6	cm

Based on this formalism, and according thermal and light daily variations (meteorological sequences) as well as a fully designable sowing scheme, WALTer produces useful outputs, such as the tillering dynamics of each plant (i.e. the number of axes on a plant for each day of the simulation). The model has shown its ability to accurately simulate the different patterns of tillering dynamics resulting from variations in sowing density (Lecarpentier et al., 2019).

WALTer is not a completely deterministic model, as several elements in the model integrate some stochasticity. Indeed, the position of the plants and the duration before plant emergence are drawn from a Gaussian distribution centred on reference values, while the plant orientation is drawn in a uniform distribution (Lecarpentier et al., 2019). Furthermore, the final number of leaves on the main stem (N_{MS}^B) is defined at the scale of the field: it is a

decimal number and the decimal part indicates the proportion of plants carrying round ($N_{MS}^B + 0.5$) leaves on the main stem. For example, a field with $\mu = 11.6$ would have 40% of plants with 11 leaves on their main stem and 60% of plants with 12 leaves.

Since its publication (Lecarpentier et al., 2019), WALTer has undergone some changes aiming at reducing the computational cost of the simulations, enhancing its realism and improving its ability to simulate mixtures of varieties (Blanc et al., 2021). In particular, the WALTer V2.0 version integrates: (i) a representation of curved leaves (Dornbusch et al., 2009; Fournier and Pradal, 2012; Perez et al., 2016); (ii) an improved discretization of the sky (den Dulk, 1989; Alinea.ASTK, version 2.1.0, 2019); (iii) the possibility to simulate an infinite periodic canopy allowing to discard border effects; (iv) removal of non-visible organs to improve the 3D representation of plants. For a full list of WALTer parameters, see Blanc et al. (2021).

WALTer is available as an open source Python package on the OpenAlea platform (<https://github.com/openalea/walter>).

Parameter estimation

First, we generically detail the method to calibrate a costly FSPM, and then we detail how it is applied to WALTer. In Supplementary Protocol S2, we provide a fully-reproducible calibration of a toy model to illustrate the genericity of the approach and facilitate the use of our method. The associated code is available as an R package also provided in Supplementary Protocol S2.

Sequential Bayesian calibration method.

In order to perform a statistical calibration of a numerical model, we posit a statistical model that links the observed data to the numerical model. This numerical model is denoted by

$f^{(d,t)}(\theta)$ where θ is the vector of parameters to calibrate, d the experimental or environmental conditions and t the time of the observation. The field data are denoted by $y_{(d,t)}$ where the indices d and t have the same meaning as in the numerical model. We assume that we have n_{exp} different experimental conditions and the number of time points at which measurements were conducted is denoted by T . Therefore the set of field data used for calibration is $Y = (y_{(d,t)})_{1 \leq i \leq n_{exp}, 1 \leq t \leq T}$.

The assumed statistical model is then:

$$y_{(d,t)} = f^{(d,t)}(\theta) + \varepsilon_{(d,t)} \quad (\text{Equation 1})$$

where the distribution of the noise is given by $\varepsilon_{(d,t)} \sim N(0, \sigma_{obs,(d,t)}^2)$ and all the ε are assumed to be independent. We allow the variances $\sigma_{obs,(d,t)}^2$ to depend on both the time of observation t and the conditions, a more flexible and likely assumption than assuming homoscedasticity. These variances aggregate uncertainties on the accuracy of the numerical model, its stochasticity and the observational noise (linked to the precision of the field measurements). They are fixed in the calibration process we propose.

The proposed calibration method is based on the Efficient Global Optimisation (EGO) algorithm (Jones et al., 1998) and involves the six following steps:

(1) **Build initial designs of numerical experiments.** A numerical design D_{init} must be selected to sample the multidimensional parameter space (θ). To ensure a good exploration of the parameter space, a maximin Latin Hypercube Design (LHD: McKay et al., 1979; Johnson et al., 1990) of 100 parameter-sets was generated with the DiceDesign package (Dupuy et al., 2015) using the ranges provided by expert knowledge.

(2) **Run f .** The numerical model must then be run over the initial design in combination with all the conditions d that are encountered in the field data. Then, the outputs corresponding to the times of observations are extracted.

(3) **Build Kriging metamodels.** The initial designs must then be used to build Kriging metamodels which approximate f in the parameter space. For each d_i and each t ($n_{exp} * T$ states), a specific metamodel is fitted. This was done in this study by using the DiceKriging package (Roustant et al., 2012). If f is stochastic, a nugget effect can be set to an estimation of the variance based on replicated runs of f with the same input. This nugget can be specific to each combination (d_i, t) . The accuracy of each Kriging metamodel can be tested through leave-one-out cross validation. We denote for each combination (d_i, t) given the design of numerical experiments D_{init} , $F_{D_{init}}^{(d,t)}$ the corresponding Kriging metamodel, the mean of which is denoted by $k_{m, D_{init}}^{(d_i, t)}$ and the variance of which is denoted by $k_{v, D_{init}}^{(d_i, t)}$.

(4) **Augment the designs of numerical experiments.** The initial design D_{init} could then be enriched by the addition of new points. The additional points are added sequentially one by one and selected on the basis of the Expected Improvement (EI) criterion adapted to the calibration goal as in Damblin et al. (2018). This criterion aims to improve the precision of the surrogate model, especially for values of the input parameters which are likely to make a good fit of f to the available experimental data. Indeed, for a given metamodel built on a limited number of runs of the FSPM, it aims to target a value of the parameters θ where the likelihood is likely to be the highest (or equivalently where the sum of weighted squares is likely to be the lowest) according to not only the predictions provided by the metamodel but also their relative uncertainties. Then, the FPSM is run for this new θ and the metamodel is updated. The difference between our method and Damblin et al. (2018) lies in the fact that several Kriging metamodels (one for each combination (d_i, t)) are combined. For this step, we

decided to enrich the design by the addition of a single point at a time, corresponding to the highest EI criterion among a set of 10 000 points selected from the parameter space by an LHD. We denote by D_k the current design of numerical experiments after that k points were already added (with the notation $D_0 = D_{init}$). The EI criterion before adding the $k+1^{\text{th}}$ point is:

$$\sum_{i=1}^{n_{\text{exp}}} \sum_{t=1}^T E \left(\max \left\{ s_k^{(d_i,t)} - \left(y_{(d_i,t)} - F_{D_k}^{(d_i,t)}(\theta) \right)^2 / \sigma_{\text{obs},(d_i,t)}^2, 0 \right\} \right) \quad (\text{Equation 2})$$

where the max notation corresponds to take the maximum between the two terms in the brackets, the expectations are computed with respect to the distributions of the Kriging metamodels ($F_{D_k}^{(d_i,t)}$) given the current design of experiments (D_k) and for each (d_i, t) , $s_k^{(d_i,t)}$ is the current minimal value of the function $\theta \mapsto \left(y_{(d_i,t)} - f^{(d_i,t)}(\theta) \right)^2 / \sigma_{\text{obs},(d_i,t)}^2$ over the current design D_k . The EI criterion computes the sums of the expectations of the improvement of the current minimal values $s_k^{(d_i,t)}$ for any value of θ . Maximizing the EI criterion will then provide a value of θ which may improve the likelihood used in Bayesian calibration. This sequential procedure is repeated until a fixed number of points has been added to the design of experiments or until an adaptive stopping criterion has been met. This adaptive criterion can consist in stopping adding new points when both the last added point and the value of the EI criterion are close to the ones obtained at the previous iteration.

(5) **Run Bayesian calibration.** A Bayesian calibration method can then be applied to the statistical model where the calls to f are replaced with calls to the metamodels and the additional uncertainty relative to this replacement is taken into account. More precisely, the statistical model used in the Bayesian calibration is based on the following expression of the logarithm of the likelihood:

$$\log(p(\theta \vee y)) = -\frac{1}{2} \left(T \cdot n_{\text{exp}} \log(2\pi) + \sum_{i,t} \left(\log(\sigma_{\text{obs},(d_i,t)}^2 + k_{v,D}^{(d_i,t)}(\theta)) + \frac{(y_{(d_i,t)} - k_{m,D}^{(d_i,t)}(\theta))^2}{\sigma_{\text{obs},(d_i,t)}^2 + k_{v,D}^{(d_i,t)}(\theta)} \right) \right) \quad (\text{Equation 3})$$

3)

where we denote by $y = (y_{(d_i,t)})_{1 \leq d_i \leq n_{\text{exp}}, 1 \leq t \leq T}$ the vector of all observations and D the numerical design of experiments that is finally used for performing the Bayesian calibration.

The posterior distribution of the parameter θ is estimated using a Markov Chain Monte Carlo (MCMC) sampling. In this study, we used a random walk Metropolis algorithm as implemented in the MCMCpack package (Martin et al., 2011). The behavior of the MCMC method can be checked by considering the value of the acceptance rate. Values between 0.2 and 0.5 are considered as reasonable (Hoff, 2011). The convergence of the chains can be checked by computing the Brooks-Gelman-Rubin statistic (Brooks and Gelman, 1998).

(6) Evaluate the calibration quality. The goodness of fit can then be evaluated through cross-validation. For this step, we can calibrate the model using only a subset of the data and use the whole data set for the validation. In order to produce probabilistic predictions of y for different observation dates and conditions, samples of parameters from the posterior distribution are used, then draws from the distributions of the Kriging metamodels (Gaussian distribution with given means and variances) are taken for these inputs. This results in sampled values for y at the different observation dates and conditions. The mean and quantiles can be derived to produce pointwise predictions and credibility intervals.

The calibration procedure we propose consists then in iterating these steps as described in Figure 1.

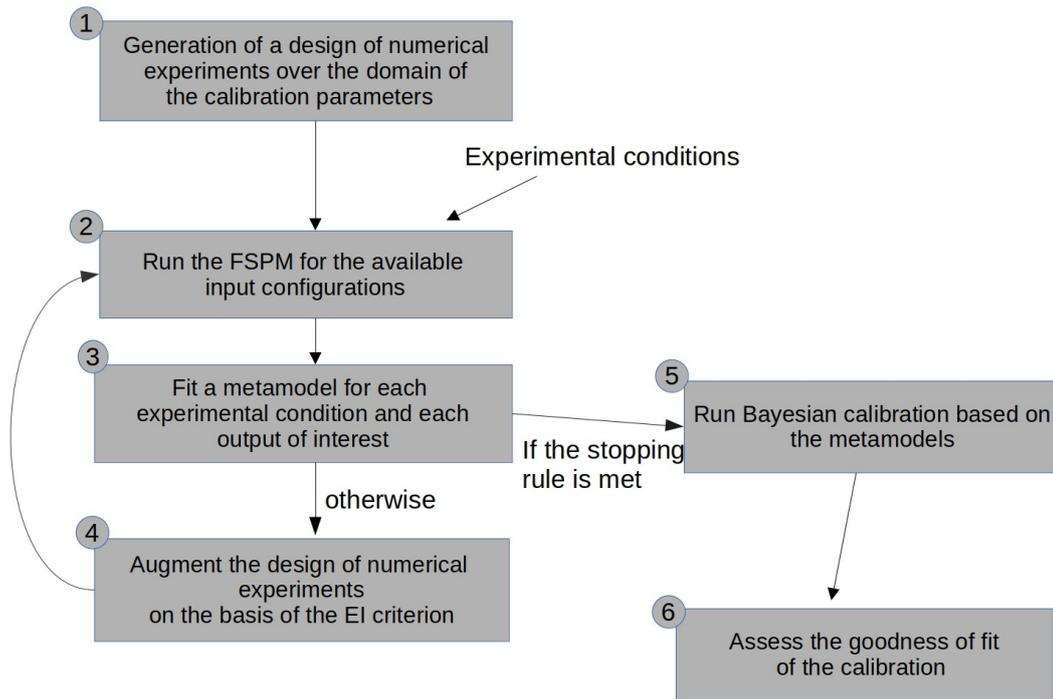


Fig. 1. Diagrammatic representation of the proposed calibration method. Step 1 consists in providing a first exploratory design of numerical experiments for the parameters to calibrate. Then, they are combined with the experimental conditions in order to be inputted in the FSPM f in step 2. From the outputs of interest that are extracted, a metamodel is fitted for each output of interest and each experimental condition in step 3. Step 4 consists in finding a new set of calibration parameters that could augment the design of numerical experiments. Step 2 and 3 are repeated in order to update the metamodels for each new added point. This process of design augmentation is stopped when a chosen stopping criterion is met. Then, step 5 runs the Bayesian calibration, the quality of which is assessed in step 6.

Specific details for calibrating WALTer. This calibration procedure is applied to calibrate five parameters of WALTer (Table 1) on the basis of observed tillering dynamics. These five parameters were selected on the basis of the results of a global sensitivity analysis of WALTer

[see its description in Supplementary Protocol S1 with the associated parameters in Supplementary Table S1 and the results in Supplementary Fig S1] because they had an important impact on the variance of the tillering dynamics and/or because they could not be measured or estimated otherwise. The ranges used for the calibration are given in Table (1).

The parameters that were not estimated here were set according to the bibliography and the manual calibration described in Lecarpentier *et al.* (2019) for the winter wheat cultivar 'Maxwell'. For meteorological data, the original sequence of PAR available in Darwinkel (1978) for Lelystad (The Netherlands) and a sequence of daily temperatures obtained by averaging Lelystad data from 2004 to 2014 were used for the simulations. The configurations denoted by d_i correspond to the sowing densities (number of plants/m²) and the sequences of PAR and temperatures. The data denoted by $y = (y_{(d_i,t)})_{1 \leq i \leq n_{\text{exp}}, 1 \leq t \leq T}$ consists of the number of axes observed at some specific times and for the different configurations.

The initial designs of experiments used to build the metamodels have size 100 over the 5 parameters to calibrate and 14 additional points are added according to the EI criterion. The number of additional points is chosen empirically. We stop when we notice that the new added points are close to the previous one and that the EI criterion seems to have reached a convergence. Since WALTer is stochastic the homogeneous nugget effect for the Kriging metamodels was set to WALTer's variance for the corresponding date and sowing density. WALTer's variance was computed thanks to 10 replicates of a reference simulation (Table 1). Note that the metamodels were built from runs of WALTer that took two days of computations. Once the metamodels are built, no more calls to WALTer are needed to run the Bayesian calibration. It takes around ten minutes to obtain posterior samples of size 20 000 while relying on direct calls to WALTer in the calibration algorithm would have taken two years of computation.

Application to a synthetic dataset

The synthetic dataset is obtained by running WALTer. We simulated tillering dynamics at six contrasted sowing densities (25, 50, 100, 200, 400 and 800 plants/m²) with a fixed set of parameters (Table 1). To build the synthetic dataset, we then extracted 13 dates spread across the tillering dynamics, in accordance with the experimental data described hereafter. Finally, a Gaussian white noise was added to the resulting tillering dynamics. For each date and at each sowing density, the noise had a standard deviation corresponding to WALTer's standard deviation between replicates at the same date and for the same sowing density. For each sowing density and each date, the observational variance σ^2_{obs} was set to the value of WALTer's variances for the corresponding dates and sowing densities.

Application to an experimental dataset

The same methodology was then applied to estimate the five parameter values that would allow the best fitting to experimental data. We used data from an experiment described by Darwinkel (1978), in which the winter wheat cultivar 'Lely' was sown in plots of 1 m² at seven contrasted sowing densities (5, 25, 50, 100, 200, 400 and 800 plants/m²) in Lelystad (The Netherlands). For each density, the number of shoots per m² was measured at 13 dates spread across the development period of the crop, thus giving a good estimation of the total tillering dynamics.

We decided not to use the data for the lowest sowing density (5 plants/m²) because simulations with WALTer for this sowing density would have taken a very long time. Furthermore, this sowing density represents extreme conditions that would rarely be observed and is thus not the most interesting density to study. Therefore, we focused our calibration on the 6 other sowing densities (25, 50, 100, 200, 400 and 800 plants/m²).

For each sowing density, the observational variance, necessary to compute the likelihood (Eq. 3), was set to achieve a coefficient of variation of 20% for the first two dates, of 5% for the last two dates and of 10% for the other dates. These coefficients of variation were chosen to take into account the uncertainty associated with the experimental measurements. A more important coefficient of variation was selected for the first two dates to account for the relatively higher difficulty of measurement at the beginning of the tillering dynamics and for a possible temporal shift due to the discrepancy between the approximated temperature sequence used for the simulations and the real experimental one, not available. The coefficient of variation for the last two dates of the tillering dynamics was set to a smaller value to account for the relative simplicity of measurement on mature plants and for the importance of the fit for these dates as they represent an important component of the yield (number of spikes/m²).

RESULTS

Application to the synthetic dataset

The accuracy of the calibration is low when it is performed using the Kriging metamodel constructed from the initial design of 100 simulations (Fig 2, top row) although the diagnostic tools for the MCMC showed satisfactory results. Indeed, the posterior distributions of Δ_{prot}

and L_{\max}^B do not include the actual values used to generate the synthetic data. The posterior density for PAR_t has a small variance but is shifted from the actual parameter value and the posterior densities of GAI_c and N_{MS}^B include the actual parameter values but have rather important variances. On the other hand, the accuracy of the calibration is greatly improved by the use of the sequential design, even with an addition of only 14 simulations. Using this enriched design, the posterior distributions of all input parameters have small variances and include the values used to generate the synthetic data (Fig 2, bottom row). Thus, as shown in Figure 3, the enriched design allows one to select a set of parameters that accurately reproduces the simulated tillering dynamics for all sowing densities.

However, even when using the augmented design, the accuracy of the calibration is greatly deteriorated when it is carried out only with the data of a single sowing density (Fig 4). Indeed, the posterior distributions obtained using only the data at 200 plants/m² all include the actual parameter values but their variances are very large especially for Δ_{prot} for which the true value is on the verge of the corresponding posterior distribution. This resulted in some of the synthetic data (especially the end of the tillering dynamics at densities 50 and 100 plants/m²) being outside the 95% credibility range of the posterior predictions which include both the Kriging and the parameter uncertainties (Fig 5). However, the fit to the synthetic data still seems reasonably good, as the 95% credibility range includes the vast majority of the data.

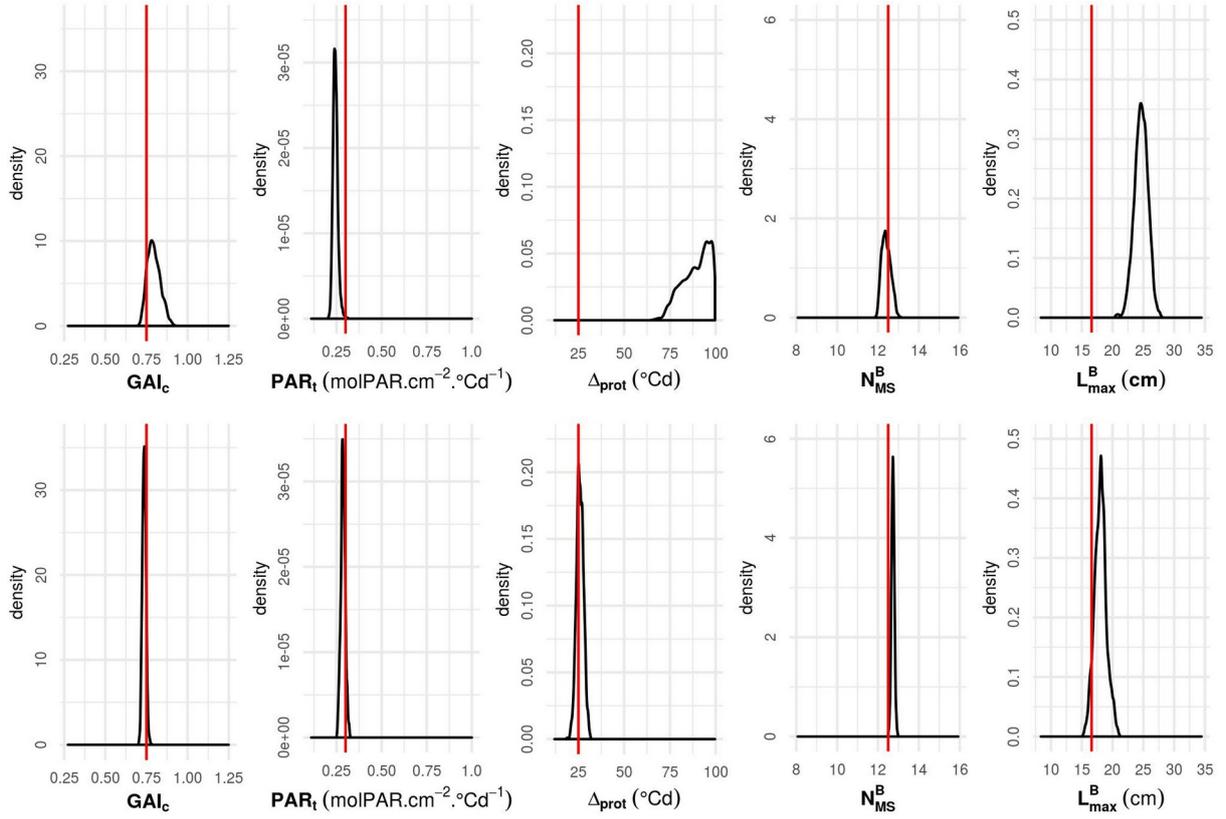


Fig. 2. Calibration results using the complete synthetic dataset: posterior distributions of WALTER parameters (GAI_c , PAR_t , Δ_{prot} , N_{MS}^B and L_{max}^B) using the synthetic data from all densities and the initial LHD design of 100 simulations (top row) or the enriched design after 14 loops of EGO (bottom row). Parameter values used to simulate the data are shown with a red vertical line.

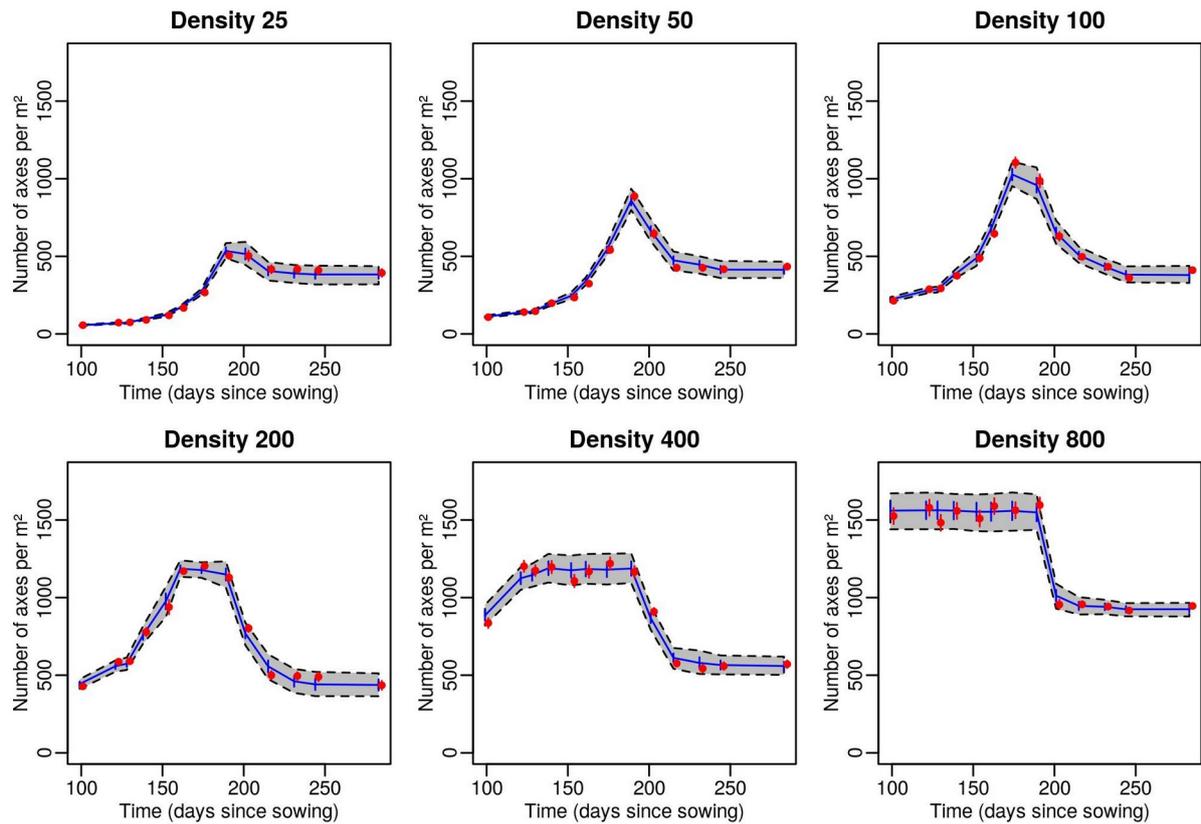


Fig 3. Validation of the calibration using the complete synthetic dataset: number of axes per m² vs. time since sowing at six sowing densities (25, 50, 100, 200, 400 and 800 plants.m⁻²). Red dots represent the synthetic data and red segments are the associated observational variance; blue lines represent the mean Kriging prediction for the set of input parameters with the best loglikelihood and blue segments are the associated Kriging standard deviation. The grey area delimited by dotted lines is the 95% credibility interval taking into account the Kriging standard deviation and uncertainty on the calibration parameters. Red points and segments are shifted on the x-axis to avoid overlapping. The set with the best loglikelihood was selected by MCMC using the enriched design after 14 loops of EGO and synthetic data from all sowing densities.

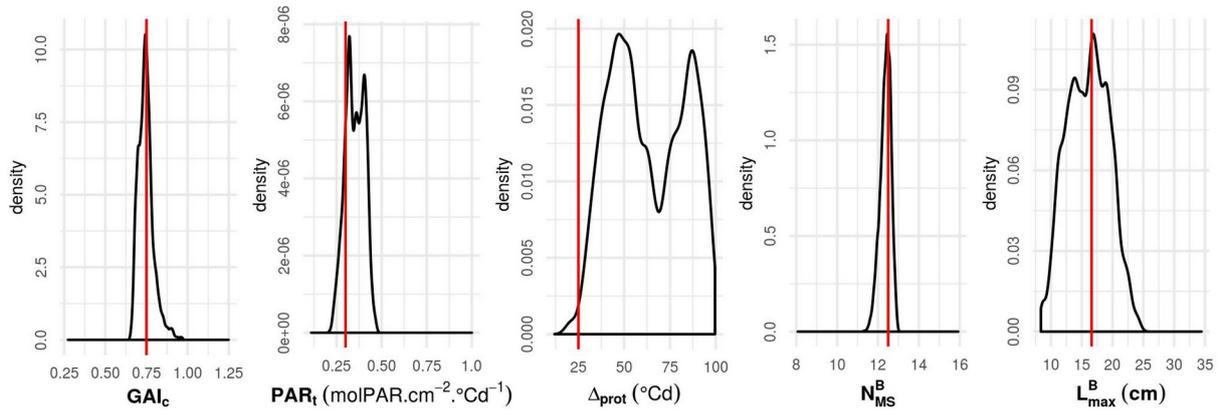


Fig. 4. Calibration results using a subset of the synthetic dataset: posterior distributions of WALTER parameters (GAI_c , PAR_t , Δ_{prot} , N_{MS}^B and L_{max}^B) using the enriched design after 14 loops of EGO and synthetic data from density 200 plants/m². Parameter values used to simulate the data are shown with a red vertical line.

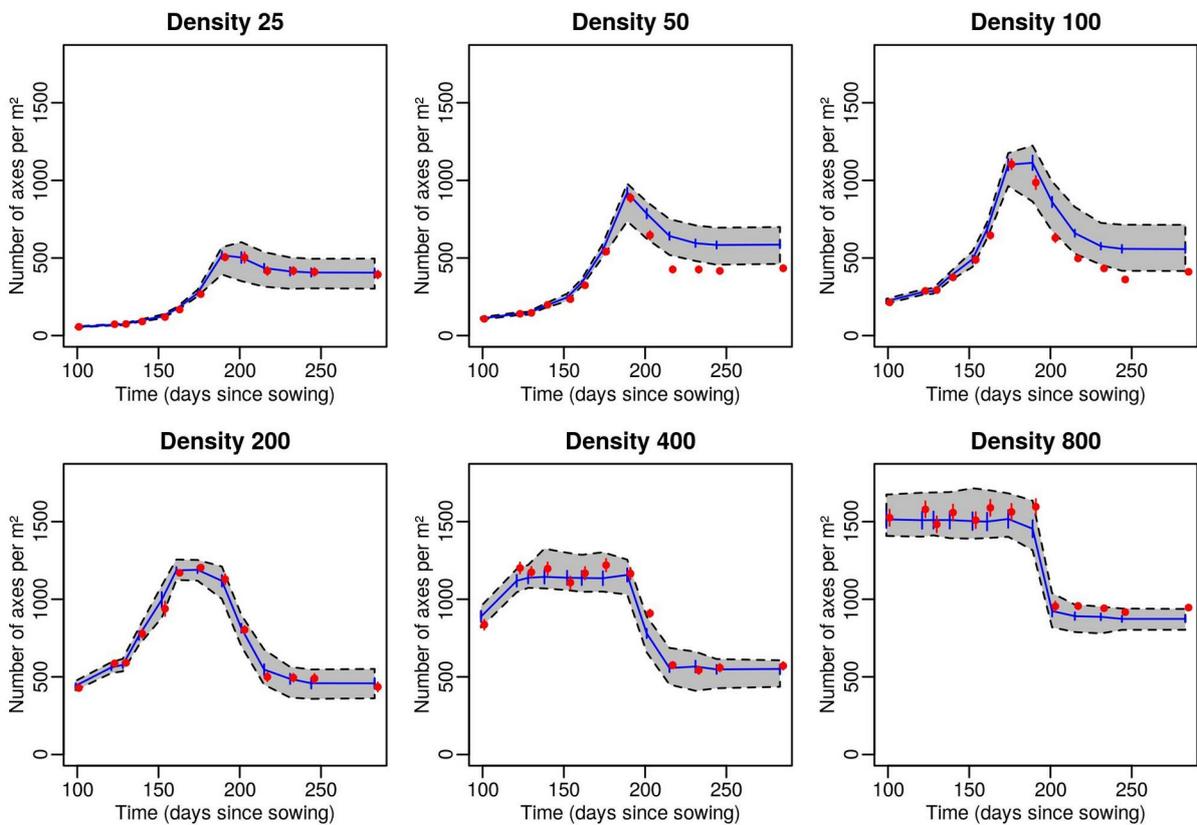


Fig 5. Validation of the calibration using a subset of the synthetic dataset: number of axes per m^2 vs. time since sowing at six sowing densities (25, 50, 100, 200, 400 and 800 plants. m^{-2}). Red dots represent the synthetic data and red segments are the associated observational variance; blue lines represent the mean Kriging prediction for the set of input parameters with the best loglikelihood and blue segments are the associated Kriging standard deviation. The grey area delimited by dotted lines is the 95% credibility interval taking into account the Kriging standard deviation and uncertainty on the calibration parameters. Red points and segments are shifted on the x-axis to avoid overlapping. The set with the best loglikelihood was selected by MCMC using the enriched design after 14 loops of EGO and synthetic data from density 200 plants/ m^2 .

Application to the experimental dataset

When the enriched design is used to calibrate WALTER on the experimental data from Darwinkel (1978), the posterior distributions (Fig 6) show contrasting results depending on the input parameter considered. The posterior density for N_{MS}^B and GAI_c , which are the 2 parameters with the largest influence on the tillering dynamics according to the global sensitivity analysis [Supplementary Fig. S1], only include intermediate values of the parameters and have a rather low variance. These two parameters are effectively critical as N_{MS}^B determines in a factorial way the number of meristems potentially producing tillers, while GAI_c provokes tillering cessation according to leaf density around the focal plant (Lecarpentier et al. 2019). On the other hand, the posterior density for Δ_{prot} has a very large variance and includes almost all the range of values explored. As for the posterior density of PAR_t , its variance is low, but it only includes values that are close to the lower bound of the variation range. PAR_t modulates tillers' regression, initiated when leaf density is high enough to shade recent tillers, so that the PAR they receive is lower than the PAR_t value (Lecarpentier

et al. 2019). The lower the density, the longer the time needed for reaching this leaf density and shading level as seen on Figure 3. Similarly, the posterior density of L_{\max}^B is clearly shifted towards the upper bound of the variation range, even though its variance is quite large.

With the 'optimal' set of parameters selected by MCMC, there is a rather good fit between the simulated tillering dynamics and the experimental dynamics of Darwinkel (1978) (Fig 7). However, some of the experimental data is outside the 95% credibility range of the Kriging metamodel, even when the observational variance is considered. In particular, for all sowing densities, the model fails to reproduce the number of axes per m^2 observed experimentally for the first 2 dates of the dynamics, with various putative origins already discussed in Lecarpentier et al. (2019) (especially the lack of knowledge on the architecture of the cultivar used by Darwinkel, as well as the true climatic sequence for this experiment). More details on the parameters' impacts on tillering or LAI dynamics and their biological meanings can be found in Lecarpentier et al. (2019) and Blanc et al. (2021).

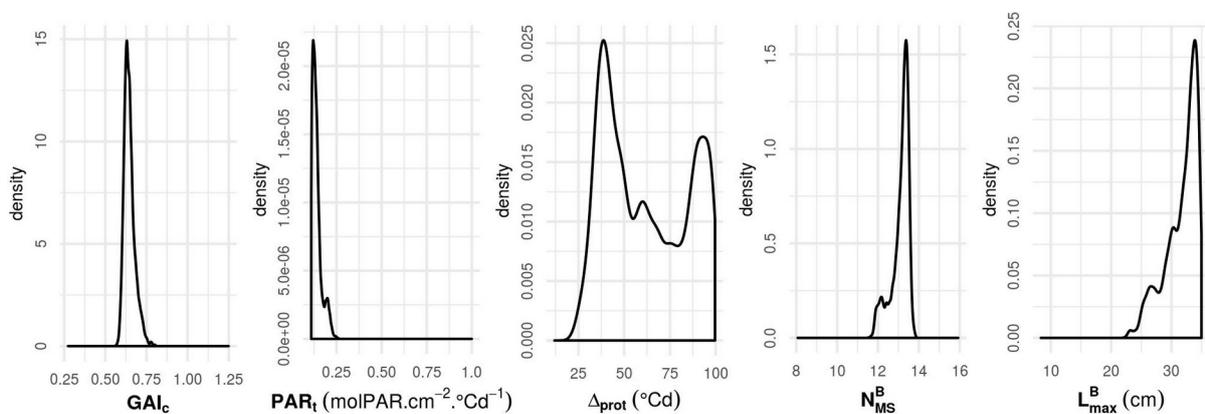


Fig. 6. Calibration results using the experimental dataset: posterior distributions of WALTER parameters (GAI_c , PAR_t , Δ_{prot} , N_{MS}^B and L_{\max}^B) using the experimental data from Darwinkel (1978) at all densities and the enriched design after 14 loops of EGO.

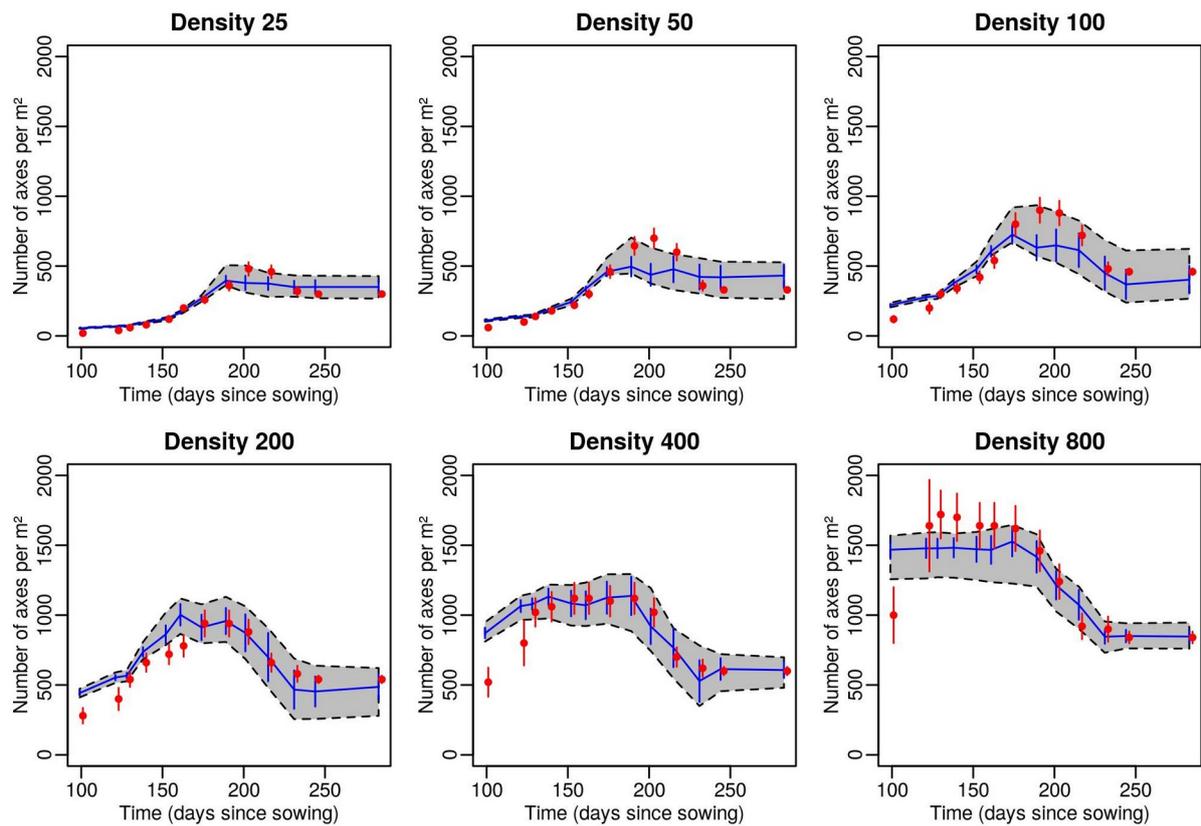


Fig 7. Validation of the calibration using the experimental dataset: number of axes per m^2 vs. time since sowing at six sowing densities (25, 50, 100, 200, 400 and 800 plants.m^{-2}). Red dots represent the experimental data from Darwinkel (1978) and red segments are the associated observational variance; blue lines represent the mean Kriging prediction for the set of input parameters with the best loglikelihood and blue segments are the associated Kriging standard deviation. The grey area delimited by dotted lines is the 95% credibility interval taking into account the Kriging standard deviation and uncertainty on the calibration parameters. Red points and segments are shifted on the x-axis to avoid overlapping. The set with the best loglikelihood was selected by MCMC using the enriched design after 14 loops of EGO and experimental data of Darwinkel (1978) from all sowing densities.

DISCUSSION

The calibration of complex models involves several challenges, due in part to the high computational cost of the simulations, as well as the numerous parameters to consider. This situation is well known for crop models (Wallach et al, 2021) and is only exacerbated for FSPM as, contrary to crop models, they explicitly model individual plants along with interactions between them (Sievanen et al, 2014). Few automatic methods have been proposed for FSPM calibration. Some methods focus on providing point estimates only (e.g., Mathieu et al, 2018) and most FSPMs were calibrated parameter-by-parameter (e.g., Colbach et al, 2021) or without full uncertainty quantification (e.g., Ding et al, 2016; Faverjon et al, 2019; Gauthier et al, 2020; Li et al, 2021). Our study, aimed at alleviating these drawbacks, illustrates the interest of using a sequential design with a metamodeling approach to calibrate the WALTer FSPM with uncertainty quantification for 5 critical parameters, based on data from the tillering dynamics. The method presented here performed well, especially when applied to synthetic data. On real data, our approach efficiently approximated the mean number of axes per m² simulated by WALTer for 13 observation dates in pure stands and 6 sowing densities. This highlights the interest of using a Kriging metamodel with WALTer to limit the computational cost and thus allow to use methods that require a large number of runs, such as the MCMC: two days were necessary to calibrate the model, instead of 2 years (single core computation)! Importantly, a set of 100 simulations was not sufficient to ensure a quality of the metamodel that was high enough for the calibration. However, an augmented design with only 14 additional simulations (114 simulations total), obtained by the sequential method, allowed for a satisfactory fit of the metamodel outputs to both synthetic and experimental data. The use of an adaptive design thus proved to be a very efficient method to improve the quality of the metamodel, reducing the uncertainty in areas of the parameter

space that are of interest for the fitting. Although the objective function in the EI criterion corresponds to the maximization of the likelihood, its optimization under the uncertainties due to the metamodel approximation leads to add points in the design of simulations that make an interesting trade-off between the value of the likelihood and the metamodel residual uncertainty at these points (see Damblin et al, 2018). Contrary to the manual calibration previously applied to WALTer (Lecarpentier et al., 2019), the method presented here provides the practitioner with a distribution of likely values for each parameter considered. Moreover, the Bayesian method presented allows for a more thorough exploration of the parameter space than the previous manual calibration. However, the method presented here could be improved. For example, the choice of a maximin LHD as the initial design for the sequential method can be discussed. Indeed, Zhang et al. (2019) argue that other designs outperform maximin LHD in both static and sequential settings. Furthermore, the good performance of the method for the calibration on the experimental data of Darwinkel (1978) relies on assumptions regarding the observational variance. Indeed, the experimental dataset consisted only of the mean tillering dynamics of the plots and no variance was provided. The uncertainty associated with the experimental measurements was thus conservatively assumed to be rather high, but a different observational variance may have differently impacted the results of the fitting. It is also important to mention the potential divergence between the sequence of daily temperature used in this work and the real one, not available, that may have an impact on the quality of the fitting and on its robustness. This was not illustrated properly in the present study but earlier results suggest that WALTer is sensitive to this data (Lecarpentier et al, 2019). For FSPM, parameter estimation often requires a lot of data at various scales and the issue of the limited availability of such appropriate datasets is a major concern (Louarn and Song, 2020). Interestingly, our study also provides information regarding the type of data necessary for the calibration of WALTer. First of all, our results suggest that data on the tillering dynamics are

sufficient to estimate the value of parameters controlling the regulation of tillering (GAI_c , PAR_t and Δ_{prot}), but also to estimate the final number of leaves on the main stem (N_{MS}^B) and the final length of the longest blade (L_{max}^B). However, data for several densities are needed to estimate precisely the parameters, especially Δ_{prot} and L_{max}^B . Moreover, it is noteworthy that the two architectural parameters (N_{MS}^B and L_{max}^B) can be measured experimentally. It might thus be possible to estimate the values of the parameters controlling the tillering dynamics (GAI_c , PAR_t and Δ_{prot}) with a dataset containing fewer measurement dates, provided that N_{MS}^B and/or L_{max}^B are measured experimentally. Moreover, the value of the five parameters were estimated with a set of 13 dates spread across the development period of the crop. However, since the parameters estimated here do not impact the rate of emission of the tillers, it is possible that the calibration could be done with a slightly lower measurement effort by reducing the number of dates at the beginning of the dynamics. In line with these last considerations, a second interest of the proposed approach is to allow the exploration of alternative experimental schemes at a low computational cost, thanks to the use of metamodels. Our study has for example illustrated the importance of collecting data for several sowing densities for the calibration of WALTer, as the estimation of the parameters is deteriorated when it is based only on data collected at a 200 plants/m² sowing density (especially for Δ_{prot} and L_{max}^B). Thanks to our simulations, it would thus be possible to identify more precisely the sowing densities that should be used to generate experimental data for the calibration of WALTer. In particular, it would be possible to identify which parameters require data at low or high sowing densities. In future work, we will explore these possibilities based on the synthetic data generated here, to identify the minimal dataset necessary to achieve a satisfactory calibration.

The automatic Bayesian calibration has room for improvement in particular with respect to providing an automatic and adaptive stopping criterion. In our study, we chose a fixed budget for the augmentation of the design of numerical experiments and we checked empirically that the last added points were close to the ones already within the design of numerical experiments. Proposing a relevant threshold to assess the proximity of the EI values or the proximity of the added points is indeed challenging since their orders of magnitude could heavily depend on the case study. Another solution to derive an automatic stopping criterion could rely on a comparison between the posterior distributions obtained by using either the new point or not when fitting the metamodels. By doing so, the stopping criterion will clearly target a stabilization of the calibration which is what we aim for. However, it will result in a more computationally demanding procedure since for each point added to the design of numerical experiments, an MCMC sampling has to be run.

One of interest of an FSPM is to produce predictions for experimental conditions that were not yet tested in a field. In this case, the FSPM is calibrated on the available field data and an extrapolation to other experimental conditions is done by running the FSPM under these conditions with the calibrated values of the parameters. If only a pointwise prediction is looked for, the maximum a posteriori value for the parameters can be chosen. However, if a credibility interval is needed, the FSPM has to be run many times with the parameters sampled in their posterior distribution which is prohibitive when the FSPM is expensive. In this case, specific metamodels have to be fitted under the new experimental conditions in order to alleviate the computational cost. The choice of the design of numerical experiments should be done with respect to the posterior distribution of the calibrated parameters.

The method presented here, based on an adaptive design and a Kriging metamodel, is an efficient approach for the calibration of WALTER and could be of interest for the calibration of

any complex model, for instance process-based plant models such as crop models and most notably other FSPM. Indeed, the method is not specific to the WALTer model, as illustrated by Carmassi et al. (2019), who applied a similar approach to a model simulating the power of a photovoltaic plant. Importantly, the method presented here was specifically adapted for the FSPM context: several metamodels were used to take into account the complex, dynamic outputs of this type of model. The use of several metamodels also allows to integrate different measurement conditions (in our case, different sowing densities), which makes the calibration method easily adaptable for non-continuous conditions, as often found in agronomic experiments. Interestingly, by reducing the computational cost of parameter space exploration, this approach would make it possible to both calibrate a given FSPM for a large number of genotypes or conditions, as well as help design the experiments needed to collect the necessary data for a reliable parameter estimation. More specifically, the method we proposed can be used to calibrate up to about ten parameters, knowing that a sensitivity analysis is a first important step in order to limit the number of parameters to actually calibrate.

SUPPLEMENTARY INFORMATION

- Supplementary Protocol S1 describes the sensitivity analysis of WALTer that was conducted.
- Supplementary Table S1 contains the list of parameters and the associated ranges considered in the sensitivity analysis.
- Supplementary Fig. S1 provides the results of the sensitivity analysis.
- Supplementary Protocol S2 implements as an R package the approach and provides a tutorial (vignette) illustrating it on a toy model.

- Supplementary Dataset S1 is a zipped file that contains WALTer inputs and outputs used in calibration on synthetic and experimental datasets.

DATA AVAILABILITY

The field data used to calibrate WALTer were manually extracted from Darwinkel (1978) and are available as supplementary material (see SI3). The input-output combinations for WALTer used in the calibration on synthetic and experimental datasets as well as the synthetic dataset are also provided in SI3.

ACKNOWLEDGMENTS

The authors thank Christian Fournier, Christophe Lecarpentier and Christophe Pradal for their valuable help in the development of the new version of WALTer.

FUNDING

This work is supported by public grants overseen by the French National research Agency (ANR) as part of the “Investissement d’Avenir” program, through the “IDI 2017” project funded by the IDEX Paris-Saclay, ANR-11-IDEX-0003-02 and the “MoBiDiv” project, ANR-20-PCPA-0006.

AUTHOR CONTRIBUTIONS

EB, JE and PB conceived the study. PB designed the method and conceived the statistical framework with input from EB and JE. EB wrote the R-scripts for the calibration, the analyses and the figures with input from JE and PB. PB and TF wrote the R package. EB wrote the first draft of the manuscript and all authors reviewed and approved the final manuscript.

CONFLICT OF INTEREST

The authors have no conflicts to declare.

LITERATURE CITED

Alinea.astk [package python]. (2019). URL: <https://github.com/openalea-incubator/astk.git>

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., ... & Tu, J. (2007). A framework for validation of computer models. *Technometrics*, 49(2), 138-154.

Blanc, E., Barbillon, P., Fournier, C., Lecarpentier, C., Pradal, C., Enjalbert, J. (2021). Functional-structural plant modeling highlights how diversity in leaf dimensions and tillering capability could promote the efficiency of wheat cultivar mixtures. *Frontiers in Plant Science*, 12, 734056.

Bornhofen, S., Barot, S., & Lattaud, C. (2011). The evolution of CSR life-history strategies in a plant model with explicit physiology and architecture. *Ecological Modelling*, 222(1), 1-10.

Brooks, S. P., & Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, 7(4), 434-455.

Carmassi, M., Barbillon, P., Chiodetti, M., Keller, M., & Parent, É. (2019). Bayesian calibration of a numerical code for prediction. *Journal de la société française de statistique*, 160(1), 1-30.

Chelle, M., Andrieu, B., & Bouatouch, K. (1998). Nested radiosity for plant canopies. *The Visual Computer*, 14(3), 109-125.

Chen, Y., & Cournède, P. H. (2014). Data assimilation to reduce uncertainty of crop model prediction with convolution particle filtering. *Ecological Modelling*, 290, 165-177

Christensen, A. J., Srinivasan, V., Hart, J. C., & Marshall-Colon, A. (2018). Use of computational modeling combined with advanced visualization to develop strategies for the design of crop ideotypes to address food security. *Nutrition reviews*, 76(5), 332-347.

Colbach, N., Colas, F., Cordeau, S., Maillot, T., Queyrel, W., Villerd, J., & Moreau, D. (2021). The FLORSYS crop-weed canopy model, a tool to investigate and promote agroecological weed management. *Field Crops Research*, 261, 108006.

Cournede, P. H., Letort, V., Mathieu, A., Kang, M. Z., Lemaire, S., Trevezas, S., ... & De Reffye, P. (2011). Some parameter estimation issues in functional-structural plant modelling. *Mathematical Modelling of Natural Phenomena*, 6(2), 133-159.

Currin, C., Mitchell, T., Morris, M., & Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416), 953-963.

- Damblin, G., Barbillon, P., Keller, M., Pasanisi, A., & Parent, É. (2018).** Adaptive numerical designs for the calibration of computer codes. *SIAM/ASA Journal on Uncertainty Quantification*, 6(1), 151-179.
- Darwinkel A. (1978).** Patterns of tillering and grains production of winter wheat at a wide range of plant densities. *Netherland Journal of Agricultural Science* 26: 383–398.
- Da Silva, D., Han, L., Faivre, R., & Costes, E. (2014).** Influence of the variation of geometrical and topological traits on light interception efficiency of apple trees: sensitivity analysis and metamodelling for ideotype definition. *Annals of botany*, 114(4), 739-752.
- Ding W, Xu L, Wei Y, Wu F, Zhu D, Zhang Y, Max N. (2016).** Genetic algorithm based approach to optimize phenotypical traits of virtual rice. *Journal of Theoretical Biology*. 403:59–67.
- Dornbusch, T., Watt, J., Baccar, R., Fournier, C., & Andrieu, B. (2009).** Towards a quantitative evaluation of cereal lamina shape using an empirical shape model. In 2009 Third International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (pp. 229-236). IEEE.
- den Dulk, J. A. (1989).** The interpretation of remote sensing: a feasibility study (Doctoral dissertation, Den Dulk).
- Dupuy, D., Helbert, C., Franco, J. (2015).** DiceDesign and DiceEval: Two R Packages for Design and Analysis of Computer Experiments. *Journal of Statistical Software*, 65(11), 1-38. URL <http://www.jstatsoft.org/v65/i11/>.
- Evers, J. B., Letort, V., Renton, M., & Kang, M. (2018).** Computational botany: advancing plant science through functional–structural plant modelling. *Annals of Botany*, 121(5), 767–772.
- Evers, J. B., Van Der Werf, W., Stomph, T. J., Bastiaans, L., & Anten, N. P. (2019).** Understanding and optimizing species mixtures using functional–structural plant modelling. *Journal of experimental botany*, 70(9), 2381-2388.
- Faverjon, L., Escobar-Gutiérrez, A., Litrico, I., Julier, B., & Louarn, G. (2019).** A generic individual-based model can predict yield, nitrogen content, and species abundance in experimental grassland communities. *Journal of Experimental Botany*, 70(9), 2491-2504.
- Fournier, C., Andrieu, B., Ljutovac, S., & Saint-Jean, S. (2003).** ADEL-Wheat: A 3D Architectural Model of wheat development. In *Plant Growth Modeling and Applications* (pp. 54-63). Springer Verlag.
- Fournier, C., & Pradal, C. (2012).** A plastic, dynamic and reducible 3D geometric model for simulating gramineous leaves. In 2012 IEEE 4th International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (pp. 125-132). IEEE.
- Gauthier, M., Barillot, R., Schneider, A., Chambon, C., Fournier, C., Pradal, C., ... & Andrieu, B. (2020).** A functional structural model of grass development based on metabolic regulations and coordination rules. *Journal of Experimental Botany*.

- Godin, C., Sinoquet, H. (2005).** Functional–structural plant modelling. *New phytologist*, 166(3), 705-708.
- Guo, Y., Fourcaud, T., Jaeger, M., Zhang, X., & Li, B. (2011).** Plant growth and architectural modelling and its applications. *Annals of botany*, 107(5), 723-727.
- He, D., Wang, E., Wang, J., & Robertson, M. J. (2017).** Data requirement for effective calibration of process-based crop models. *Agricultural and Forest Meteorology*, 234, 136-148.
- Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., & Ryne, R. D. (2004).** Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2), 448-466.
- Hoff, P. D. (2009).** A first course in Bayesian statistical methods (Vol. 580). New York: Springer.
- Johnson, M. E., Moore, L. M., & Ylvisaker, D. (1990).** Minimax and maximin distance designs. *Journal of statistical planning and inference*, 26(2), 131-148.
- Jones, D. R., Schonlau, M., & Welch, W. J. (1998).** Efficient global optimization of expensive black-box functions. *Journal of Global optimization*, 13(4), 455-492.
- Kennedy, M. C., & O'Hagan, A. (2001).** Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(3), 425-464.
- Larue, F., Fumey, D., Rouan, L., Soulié, J. C., Roques, S., Beurier, G., & Luquet, D. (2019).** Modelling tiller growth and mortality as a sink-driven process using EcoMeristem: implications for biomass sorghum ideotyping. *Annals of botany*, 124(4), 675-690.
- Lecarpentier, C., Barillot, R., Blanc, E., Abichou, M., Goldringer, I., Barbillon, P., ... & Andrieu, B. (2019).** WALTer: a three-dimensional wheat model to study competition for light through the prediction of tillering dynamics. *Annals of botany*, 123(6), 961-975.
- Louarn, G., & Song, Y. (2020).** Two decades of functional–structural plant modelling: now addressing fundamental questions in systems biology and predictive ecology. *Annals of Botany*, 126(4), 501-509.
- Luquet, D., Dingkuhn, M., Kim, H., Tambour, L., & Clement-Vidal, A. (2006).** EcoMeristem, a model of morphogenesis and competition among sinks in rice. 1. Concept, validation and sensitivity analysis. *Functional Plant Biology*, 33(4), 309-323.
- Martin, A. D., Quinn, K. M., & Park, J. H. (2011).** MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*. 42(9): 1-21.
- Mathieu, A., Vidal, T., Jullien, A., Wu, Q., Chambon, C., Bayol, B., Cournède, P.-H., (2018).** A new methodology based on sensitivity analysis to simplify the recalibration of functional–structural plant models in new conditions. *Ann Bot* 122, 397–408.

McKay, M. D., Beckman, R. J., & 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.

Muleta, M. K., & Nicklow, J. W. (2005). Sensitivity and uncertainty analysis coupled with automatic calibration for a distributed watershed model. *Journal of hydrology*, 306(1-4), 127-145.

Parsons, R. A., Mell, W. E., & McCauley, P. (2011). Linking 3D spatial models of fuels and fire: effects of spatial heterogeneity on fire behavior. *Ecological Modelling*, 222(3), 679-691.

Perez, R. P., Dauzat, J., Pallas, B., Lamour, J., Verley, P., Caliman, J. P., ... & Faivre, R. (2018). Designing oil palm architectural ideotypes for optimal light interception and carbon assimilation through a sensitivity analysis of leaf traits. *Annals of botany*, 121(5), 909-926.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

de Reffye, P., Hu, B., Kang, M., Letort, V., & Jaeger, M. (2021). Two decades of research with the GreenLab model in Agronomy. *Annals of botany*, 127(3), 281-295.

Ribeyre, F., Jaeger, M., Ribeyre, A., & de Reffye, P. (2018, November). StemGL, a FSPM tool dedicated to crop plants model calibration in the single stem case. In 2018 6th International Symposium on Plant Growth Modeling, Simulation, Visualization and Applications (PMA) (pp. 33-38). IEEE.

Roustant, O., Ginsbourger, D., Deville, Y. (2012). DiceKriging, DiceOptim: Two R Packages for the Analysis of Computer Experiments by Kriging-Based Metamodeling and Optimization. *Journal of Statistical Software*, 51(1), 1-55. URL <http://www.jstatsoft.org/v51/i01/>.

Sacks, J., Welch, W. J., Mitchell, T. J., & Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical science*, 409-423.

Sievänen, R., Godin, C., DeJong, T. M., & Nikinmaa, E. (2014). Functional–structural plant models: a growing paradigm for plant studies. *Annals of botany*, 114(4), 599-603.

Simpson, T. W., Poplinski, J. D., Koch, P. N., & Allen, J. K. (2001). Metamodels for computer-based engineering design: survey and recommendations. *Engineering with computers*, 17(2), 129-150.

Trevezas, S., & Cournède, P. H. (2013). A sequential Monte Carlo approach for MLE in a plant growth model. *Journal of Agricultural, Biological, and Environmental Statistics*, 18(2), 250-270.

- Vidal, T., Gigot, C., de Vallavieille-Pope, C., Huber, L., & Saint-Jean, S. (2018).** Contrasting plant height can improve the control of rain-borne diseases in wheat cultivar mixture: modelling splash dispersal in 3-D canopies. *Annals of botany*, 121(7), 1299-1308.
- Villa-Vialaneix, N., Follador, M., Ratto, M., & Leip, A. (2012).** A comparison of eight metamodeling techniques for the simulation of N₂O fluxes and N leaching from corn crops. *Environmental Modelling & Software*, 34, 51-66.
- Vos, J., Evers, J. B., Buck-Sorlin, G. H., Andrieu, B., Chelle, M., & De Visser, P. H. (2010).** Functional–structural plant modelling: a new versatile tool in crop science. *Journal of experimental Botany*, 61(8), 2101-2115.
- Wallach, D., Buis, S., Lecharpentier, P., Bourges, J., Clastre, P., Launay, M., ... & Justes, E. (2011).** A package of parameter estimation methods and implementation for the STICS crop-soil model. *Environmental Modelling & Software*, 26(4), 386-394.
- Wallach D, Palosuo T, Thorburn P, Hochman Z, Gourdain E, Andrianasolo F, Asseng S, Basso B, Buis S, Crout N, et al. (2021).** The chaos in calibrating crop models: Lessons learned from a multi-model calibration exercise. *Environmental Modelling & Software*. 145:105206.
- Wong, R. K., Storlie, C. B., & Lee, T. C. (2017).** A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(2), 635-648.
- Zhang, B., Cole, D. A., & Gramacy, R. B. (2019).** Distance-distributed design for Gaussian process surrogates. *Technometrics*, 1-13.