Automatic calibration of a functional-structural wheat model using an adaptive design and a metamodelling approach

Emmanuelle Blanc[1*], Jérôme Enjalbert[1], Pierre Barbillon[2*]

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

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

*Corresponding authors

Emmanuelle Blanc

emmanuelle.blanc@universite-paris-saclay.fr

Pierre Barbillon

pierre.barbillon@agroparistech.fr

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

Abstract

- Background and Aims

  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. 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.

- Methods

  We used a Bayesian calibration method to estimate the values of 5 parameters of the WALTer model by fitting the model outputs to tillering dynamics data. The method presented in this paper is based on the Efficient Global Optimisation algorithm. It involves the use of Gaussian process metamodels to generate fast approximations of the model outputs. To account for the uncertainty associated with the metamodels approximations, an adaptive design was used. The efficacy of the method was first assessed using simulated data. The calibration was then applied to experimental data.

- Key Results

  The method presented here performed well on both simulated and experimental data. In particular, the use of an adaptive design proved to be a very efficient method to improve the quality of the metamodels predictions, especially by reducing the uncertainty in areas of the parameter space that were of interest for the fitting. Moreover, we showed the necessity to have a diversity of field data in order to be able to calibrate the parameters.

- Conclusions

  The method presented in this paper, based on an adaptive design and Gaussian process metamodels, is an efficient approach for the calibration of WALTer and could be of interest for the calibration of other functional-structural plant models.

INTRODUCTION

Modelling 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, modelling 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 in 3D 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).

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. Calibration can be done either manually 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). Furthermore, automatic calibration may allow the modeller to look for a distribution of likely values of the parameters instead of looking for an optimal parameter value. 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. The experimental data and the FSPM are linked in a statistical model from which a likelihood is then derived. 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, MCMC algorithms are run to obtain a sample in the posterior distribution. However, these algorithms resort to many runs of the model, which is why, for complex models, including FSPM, calibration is often done manually (see Lecarpentier et al., 2019 and Gauthier et al., 2020 for example). 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. The solution proposed in the context of Bayesian calibration is to make recourse to a Gaussian Process (Currin et al., 1991; Sacks et al., 1989) metamodel, also known as Kriging metamodel, which

3

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 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 can be used for the calibration of the WALTer FSPM.

MATERIALS AND METHODS

The efficiency of an automated calibration algorithm to estimate 5 critical parameters of the FSPM WALTer was assessed on both simulated 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). The regulation of tillering in the model is based on three simple rules. (i) Tillers emerge according to empirically fixed probabilities. 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). During tiller regression, there is a protection period (Δ_{prot}) between the deaths of two successive tillers on a plant.

Based on this formalism, 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, the orientation of the organs, the duration before plant emergence and the emergence of tillers are partly random. 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 = 11.6 would have 40% of plants with 11 leaves on their main stem and 60% of plants with 12 leaves.

Since its publication in 2019 (Lecarpentier et al.), 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., in prep). In particular, the new version of WALTer integrates: 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 from the 3D representation of plants.

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

Parameter estimation

Five parameters of the model (Table 1) were estimated by fitting the tillering dynamics. These five parameters were selected based on the results of a global sensitivity analysis of WALTer [**Supplementary Information**] 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 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

6

'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.

In order to perform the calibration of WALTer, we assumed a statistical model linking WALTer to the observed data. We denote by $y_{(d,t)}$ an observation of the number of axes per m² for density d at time t and by $f^{(d,t)}(x)$ the evaluation of WALTer for the same density and time with the input parameters set to x. Thus, the assumed statistical model is:

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

where the distribution of the noise is given by $\varepsilon_{(d,t)} \sim N(0, \sigma^2_{(d,t)})$ and all the ε are assumed to be independent.

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

(1) First, an initial numerical design D_{init} must be selected to sample the multidimensional parameter space. To ensure a good exploration of the parameter space, a five dimensions maximin Latin Hypercube (LH) design (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 detailed in Table 1.

(2) The model must then be run over the initial design. All the simulations of the LHD were run with WALTer for $n_d = 6$ contrasted sowing densities (25, 50, 100, 200, 400 and 800 plants/m²), leading to a total of 600 simulations. For each simulation, the mean tillering

7

dynamics of the field was computed by WALTer and a subset of $n_t = 13$ dates spread across the dynamics was extracted.

(3) The initial design must then be used to build a Kriging metamodel which approximates WALTer in the parameter space. For each sowing density and for each of the 13 dates (78 states), a specific metamodel was fitted by using the DiceKriging package (Roustant et al., 2012). For each metamodel, the homogeneous nugget effect 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). The accuracy of each Kriging metamodel was tested through cross validation by leave-one-out. We denote for a sowing density d and a date 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 $k_{m,D_{init}}^{(d,t)}$ and the variance of which is denoted by $k_{v,D_{init}}^{(d,t)}$.

(4) The initial design D_{init} could then be enriched by the addition of new points. The additional points are selected based on 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 WALTer to the available experimental data. The difference with Damblin et al., 2018 lies in the fact that several Kriging metamodelds (one for each combination of sowing density and date) 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 LH sampling. 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 $k+1^{th}$ point is:

$$\sum_{d=1}^{n_d} \sum_{t=1}^{n_t} \mathrm{E}\left(\max \left\{ s_k^{(d,t)} - \left(y_{(d,t)} - F_{D_k}^{(d,t)}(x) \right)^2 \Big/ \sigma_{obs,(d,t)}^2, 0 \right\} \right) \quad \text{(Equation 2)}$$

where the expectations are computed with respect to the distributions of the Kriging metamodels ($F_{D_k}^{(d,t)}$) given the current design of experiments (D_k) and $s_k^{(d,t)}$ is the current minimal value of the function $x \mapsto \left(y_{(d,t)} - f^{(d,t)}(x) \right)^2 \Big/ \sigma_{obs,(d,t)}^2$ over the current design D_k.

(5) A Bayesian calibration method can then be applied to the model. For each input parameter, the posterior distribution was estimated using a Markov Chain Monte Carlo (MCMC) sampling. More precisely, a random walk Metropolis algorithm was used, as implemented in the MCMCpack package (Martin et al., 2011). From Eq.1, the log-likelihood is then computed for any value of the parameters x as:

$$\log\left(p(x|\mathbf{y}) \right) = -\frac{1}{2} \left[n_t \cdot n_d \log(2\pi) + \sum_{d,t} \left(\log\left(\sigma_{obs,(d,t)}^2 + k_{v,d}^{(d,t)}(x) \right) + \frac{\left(y_{(d,t)} - k_{m,D}^{(d,t)}(x) \right)^2}{\sigma_{obs,(d,t)}^2 + k_{v,d}^{(d,t)}(x)} \right) \right]$$

(Equation 3)

where we denote by $\mathbf{y} = \left(y_{(d,t)} \right)_{1 \le d \le n_d, 1 \le t \le n_t}$ the vector of all observations and D the numerical design of experiments that is finally used for performing the Bayesian calibration.

The observational variance of each output and the uncertainty of the Kriging metamodel were taken into account to avoid overfitting of the model.

(6) The goodness of fit can then be evaluated through cross-validation. For this step, we calibrated the model using only the data from one (200 plants/m²) sowing density and used the whole data set for the validation.

Application to synthetic data

At first, we applied our methodology to a simulated dataset. Using 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, the observational variance of each date was set to the value of WALTer's variance for the corresponding date at the same sowing density.

Application to experimental data

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 take a very long time. Furthermore, this sowing density represents extreme conditions that would rarely be observed and is thus

10

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. 1), 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 synthetic data

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). 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

11

is greatly improved by the use of the sequential design, even with an addition of only 15 points. 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 to select a set of parameters that reproduces accurately the simulated tillering dynamics for all sowing densities.

However, even when using the design enriched with 15 simulations, 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. This resulted in some of the simulated data (especially the end of the tillering dynamics at densities 50 and 100 plants/m²) being outside the 95% credibility range of the Kriging metamodel when using this single density to select the 'optimal' set of parameters (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.

Application to experimental data

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 most influence on the tillering dynamics according to the global sensitivity analysis [**Supplementary Information**], only include intermediate values of the parameters and have a rather low variance. 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

12

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. 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 except 800 plants/m², the model fails to reproduce the number of axes per m² observed experimentally for the first 2 dates of the dynamics.

DISCUSSION

The calibration of complex models, such as FSPM, involves several challenges, due in part to the high computational cost of the simulations, as well as the numerous parameters to consider. This study illustrates the interest of using a sequential design with a metamodelling approach to calibrate the WALTer FSPM for 5 critical parameters, based on data from the tillering dynamics. The method presented here performed well, especially when applied to synthetic data. The Kriging metamodel efficiently approximated the mean number of axes per m² simulated by WALTer for 13 dates in pure stands. 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. 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 enriched design with only 15 additional simulations (115 simulations total), obtained by the sequential method, allowed for a satisfactory fit of the

13

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. 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 rigorous 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 Latin Hypercube Sampling (LHS) as the initial design for the sequential method can be discussed. Indeed, Zhang et al. (2019) argue that other designs outperform maximin LHS 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 assumed to be rather high, but a different observational variance may have 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 has an impact on the quality of the fitting and on its robustness. This highlights the importance of the quality of the data used for the fitting. 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 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

14

longest blade (L_{max}^{B}). However, 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. 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 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 other FSPM. Interestingly, by reducing the computational cost of parameter space exploration, this approach would make it possible to both calibrate FSPM for a large number

15

of genotypes or conditions, and help design the experiments needed to collect the necessary data for a reliable parameter estimation.

SUPPLEMENTARY INFORMATION

Supplementary information consist of the following. SI1_WALTer_sensitivity_analysis: information regarding the sensitivity analysis of WALTer.

ACKNOWLEDGEMENTS

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. (in prep). Functional-structural plant modeling highlights how diversity in leaf dimensions and tillering capability could promote the efficiency of wheat cultivar mixtures.

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.

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.

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.

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.

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.

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.

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.

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.

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. URL http://www.jstatsoft.org/v42/i09/.

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/.

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.

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.

Zhang, B., Cole, D. A., & Gramacy, R. B. (2019). Distance-distributed design for Gaussian process surrogates. *Technometrics*, 1-13.

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

Parameter	Description	Range	Value for the simulated 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	$\mu mol.cm^{-2}.°Cd^{-1}$
Δ_{prot}	Thermal time interval during which two tillers of the same plant cannot die	10 - 100	25	°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

20
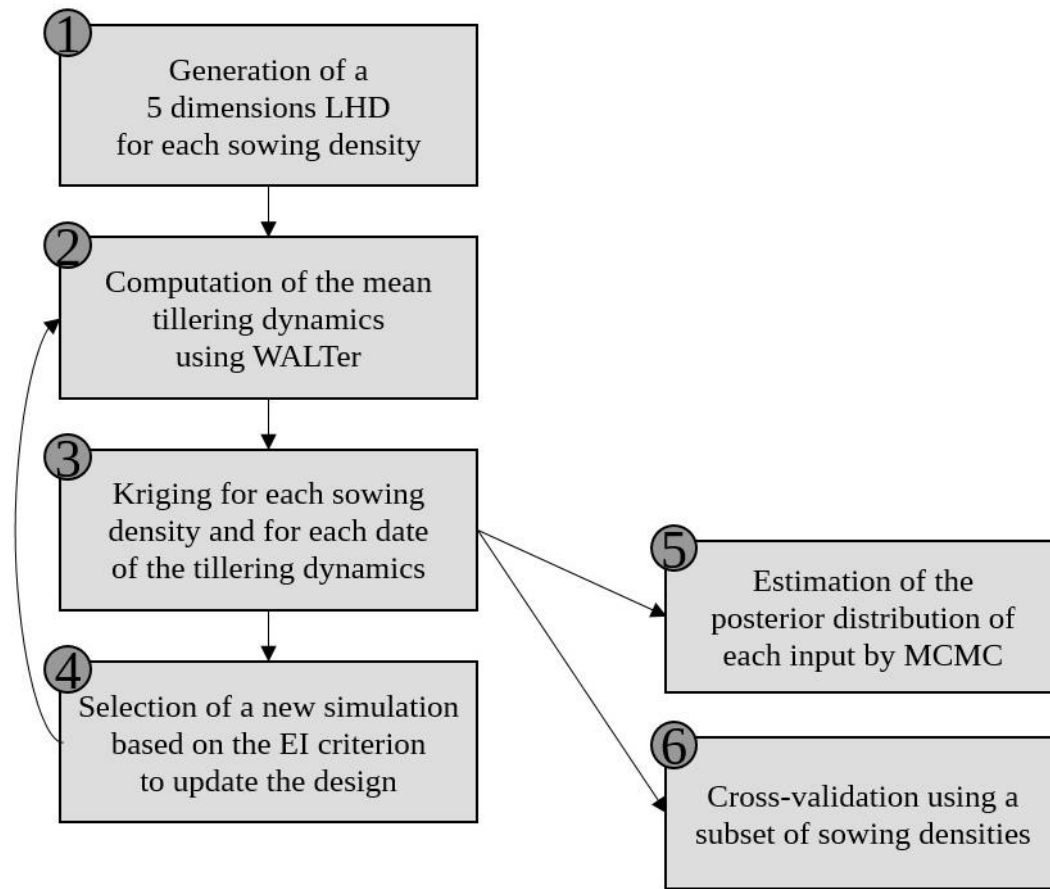
Fig. 1. Diagrammatic representation of the proposed calibration method
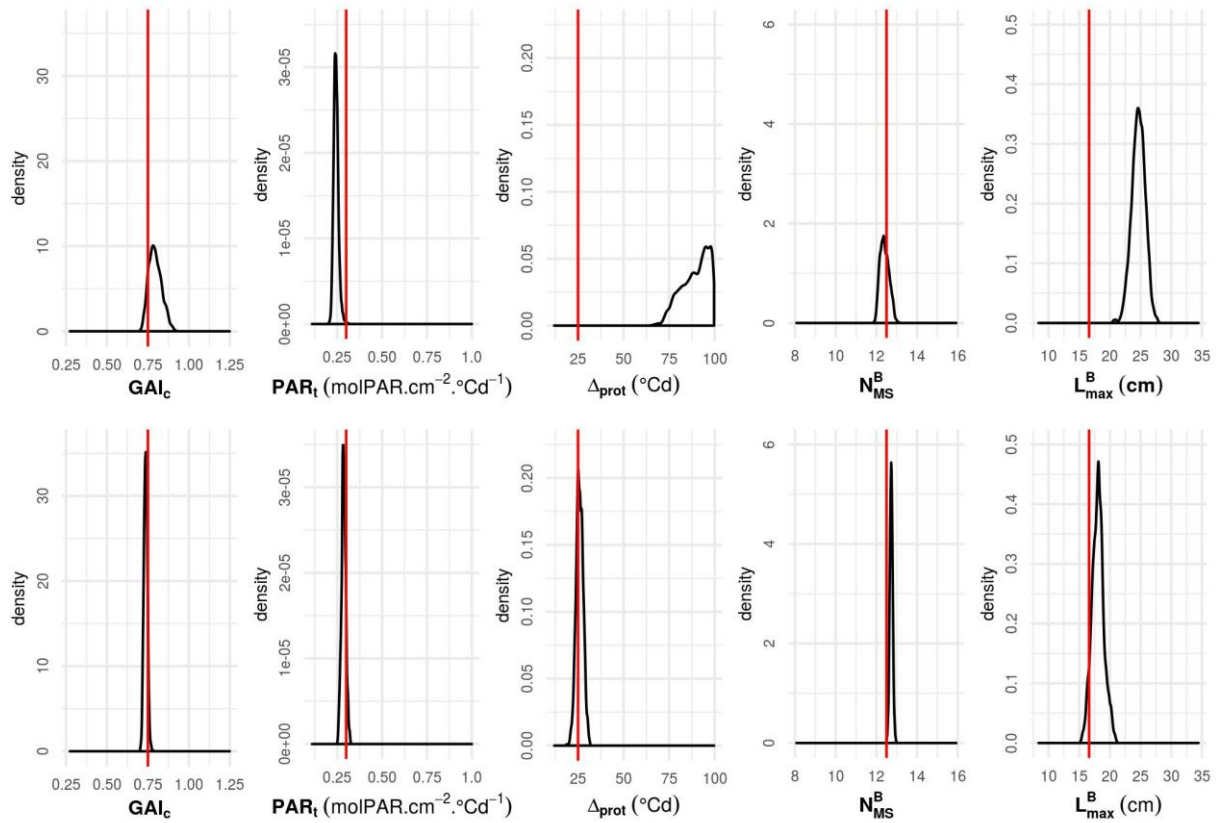
Fig. 2. Calibration results using the complete simulated dataset: posterior distributions of WALTer parameters (GAI$_c$, PAR$_t$, Δ_{prot}, N^{B}_{MS} and L^{B}_{max}) using the simulated data from all densities and the initial LH design of 100 simulations (top row) or the enriched design after 15 loops of EGO (bottom row). Parameter values used to simulate the data are shown with a red vertical line.
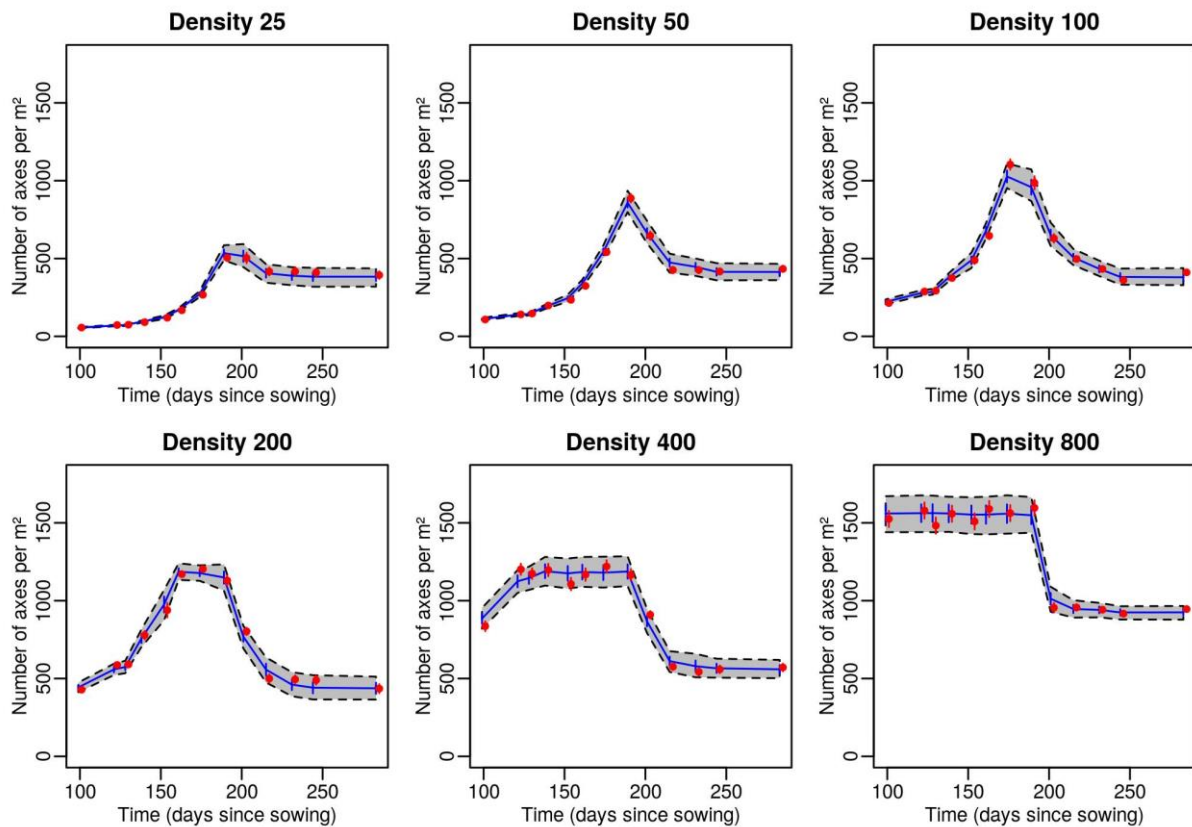
Fig 3. Validation of the calibration using the complete simulated dataset: number of axes per m² vs. time since sowing at six sowing densities (25, 50, 100, 200, 400 and 800 plants.m⁻²). Red dots represent the synthetic data and red segments are the associated observational variance; blue lines represent 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 parameter. 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 15 loops of EGO and simulated data from all sowing densities.
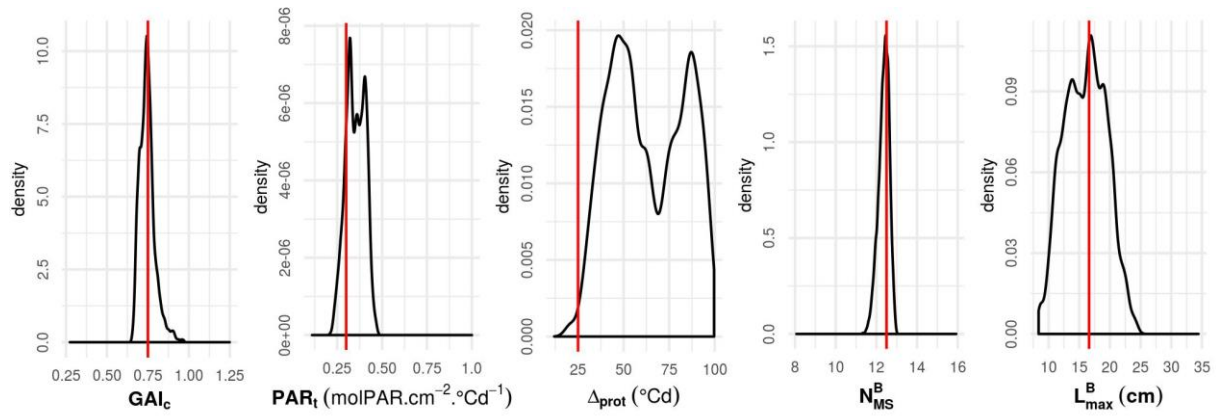
Fig. 4. Calibration results using a subset of the simulated dataset: posterior distributions of WALTer parameters (GAI$_c$, PAR$_t$, Δ_{prot}, N^B_{MS} and L^B_{max}) using the enriched design after 15 loops of EGO and simulated data from density 200 plants/m². Parameter values used to simulate the data are shown with a red vertical line.
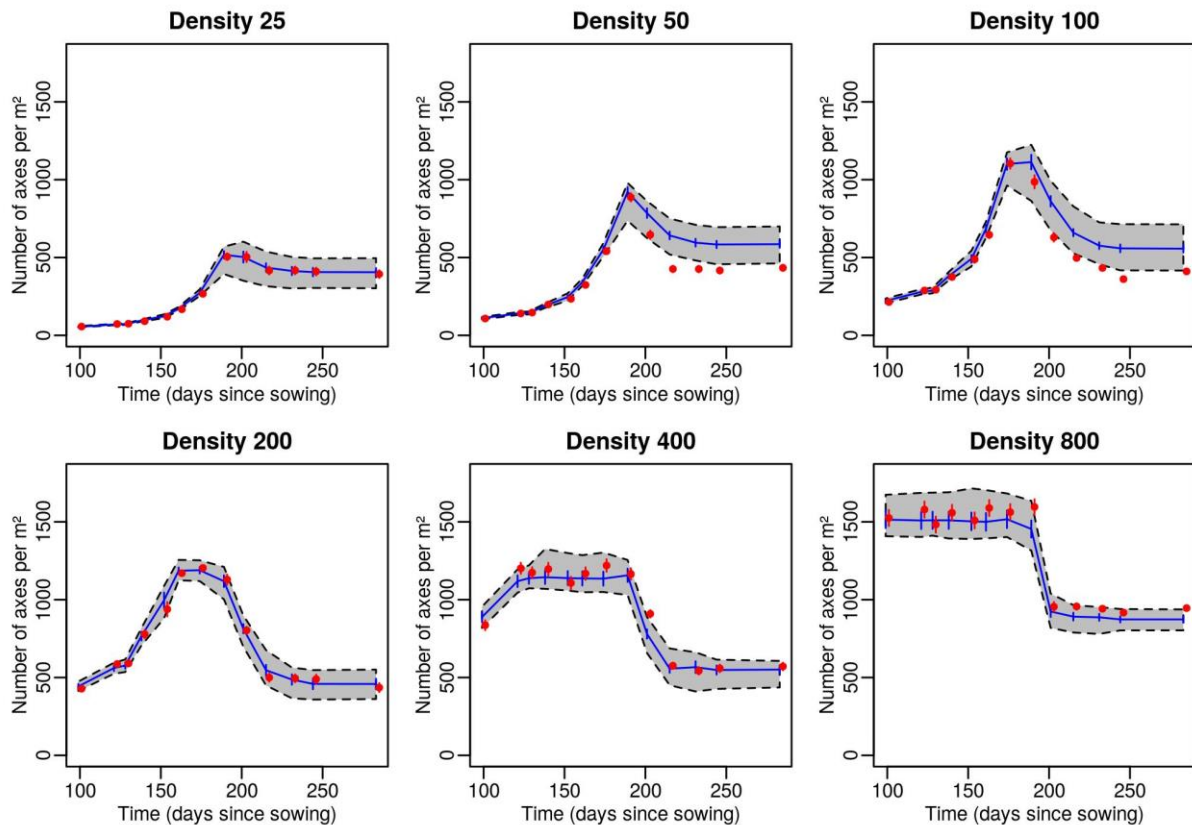
Fig 5. Validation of the calibration using a subset of the simulated dataset: number of axes per m² vs. time since sowing at six sowing densities (25, 50, 100, 200, 400 and 800 plants.m⁻²). Red dots represent the synthetic data and red segments are the associated observational variance; blue lines represent 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 parameter. 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 15 loops of EGO and simulated data from density 200 plants/m².
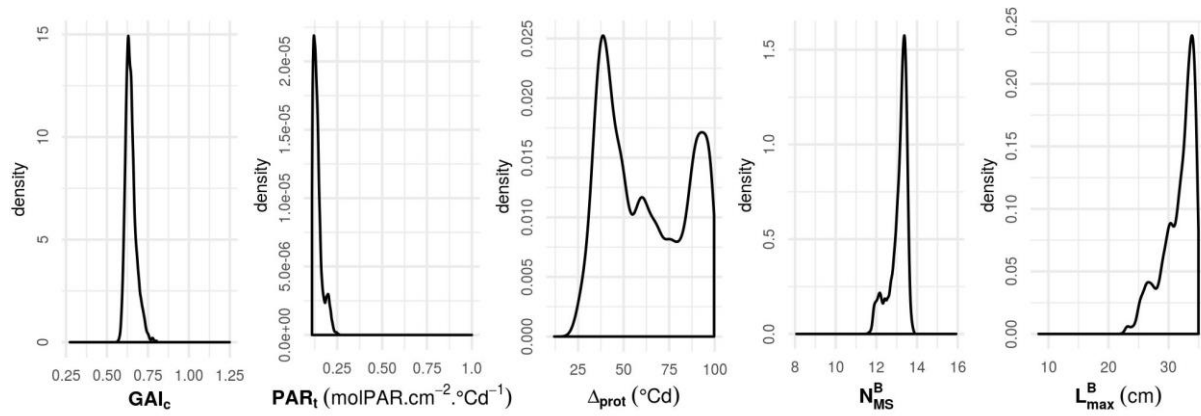
Fig. 6. Calibration results using the experimental dataset: posterior distributions of WALTer parameters (GAI$_c$, PAR$_t$, Δ_{prot}, N^B_{MS} and L^B_{max}) using the experimental data from Darwinkel (1978) at all densities and the enriched design after 15 loops of EGO.
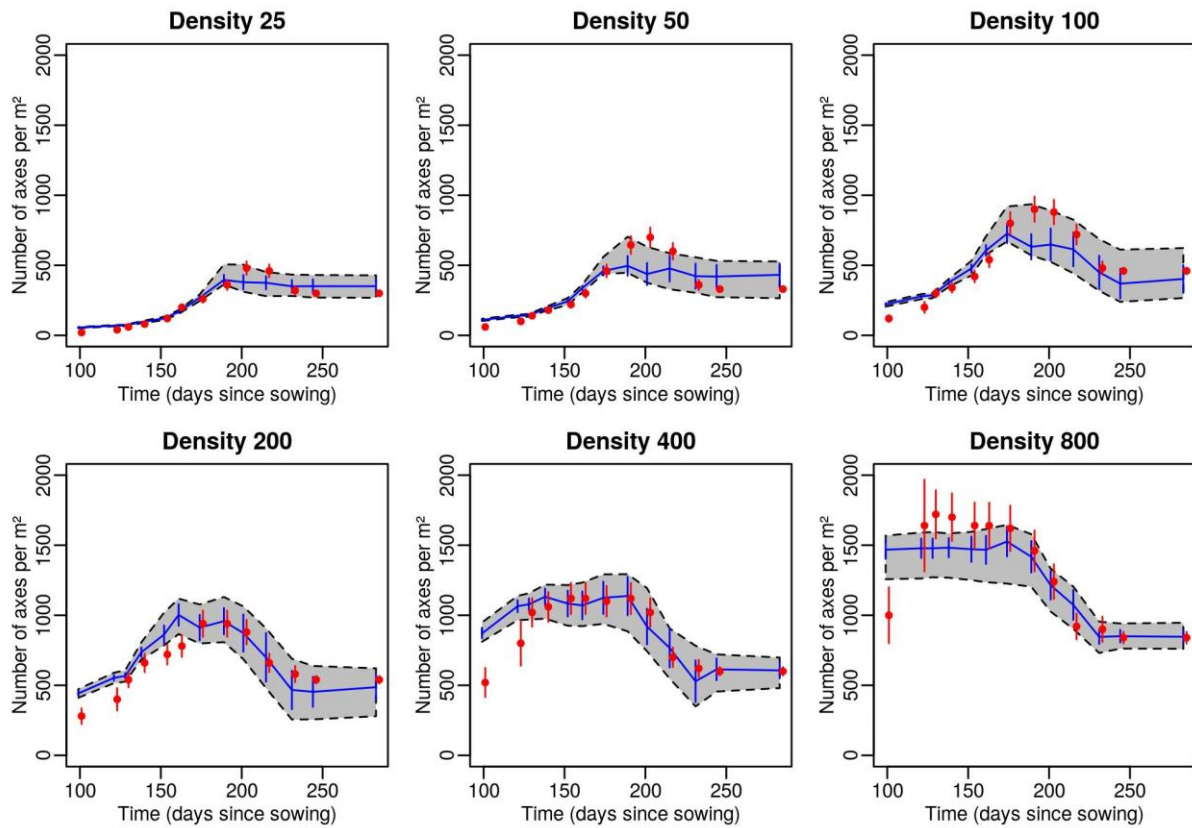
Fig 7. Validation of the calibration using the experimental dataset: number of axes per m² vs. time since sowing at six sowing densities (25, 50, 100, 200, 400 and 800 plants.m⁻²). Red dots represent the experimental data from Darwinkel (1978) and red segments are the associated observational variance; 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. 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 15 loops of EGO and experimental data of Darwinkel (1978) from all sowing densities.