PARAMETER IDENTIFICATION OF THE STICS CROP MODEL, USING AN ACCELERATED FORMAL MCMC APPROACH

Dumont B.[1]*, Leemans V.[1], Mansouri M.[1], Bodson B.[2], Destain J.-P.[3], Destain M.-F.[1]

1, ULg (GxABT), Unité de Mécanique et construction

2, ULg (GxABT), Unité de Phytotechnie des régions tempérées

3, ULg (GxABT), Centre wallon de Recherches agronomiques (CRA-W)

* Corresponding author: 2 Passage des Déportés, 5030 Gembloux, Belgium

@: benjamin.dumont@ulg.ac.be

Tel.: +32(0)81/62.21.63, fax: +32(0)81/62.21.67

Keywords: Crop model – Parameter estimation – Bayes – STICS – DREAM

Abstract

This study presents a Bayesian approach for the parameters' identification of the STICS crop model based on the recently developed Differential Evolution Adaptive Metropolis (DREAM) algorithm. The posterior distributions of nine specific crop parameters of the STICS model were sampled with the aim to improve the growth simulations of a winter wheat (Triticum aestivum L.) culture. The results obtained with the DREAM algorithm were initially compared to those obtained with a Nelder-Mead Simplex algorithm embedded within the OptimiSTICS package. Then, three types of likelihood functions implemented within the DREAM algorithm were compared, namely the standard least square, the weighted least square, and a transformed likelihood function that makes explicit use of the coefficient of variation (CV). The results showed that the proposed CV likelihood function allowed taking into account both noise on measurements and heteroscedasticity which are regularly encountered in crop modelling.

1. Introduction

In recent decades, the number of dynamic crop models developed for estimating crop performance based on the interactions between environment and agricultural management has greatly increased. There are two types of models: specific and generic. The former are process-oriented models capable of simulating water balance, nitrogen balance, growth and the development of a given crop, while maintaining reasonable input requirements. For example, the CERES-Wheat model simulates the growth, development and yield of wheat (*Triticum aestivum* L.), taking account of the effects of weather, genetics, soil (water, carbon and nitrogen), planting, irrigation and nitrogen fertilizer management (Ritchie and Otter, 1984; Singh et al., 2008). Generic models are based on physiological principles for growth and development processes that are common across many crops. They use a modular code for crop modelling, providing easy ways of comparing modelling approaches without the need to change the code. They also provide a way to interpret data from field experiments in various environments (Monteith, 1996) and to analyse the processes at the plant component level (Confalonieri and Bechini, 2004). Well-known generic models that are able to simulate the growth and development of various crops (wheat, maize, sorghum, etc.) are EPIC (William et al., 1989), WOFOST (Van Diepen et al., 1989), DAISY (Hansen et al., 1990), STICS (Brisson et al., 1998) and SALUS (Basso and Ritchie, 2005).

The number of parameters required by generic models is higher than for specific models. The STICS model used in this study (Brisson et al., 1998; Brisson et al., 2003; Brisson et al., 2009) is characterized by its ability to adapt to a wide range of agro-environmental issues and its adaptability to various crops : e.g. wheat, sugarbeet, sugarcane, rice. It implies that the number of parameters involved is high: more than 200 parameters are arranged in three main groups related to (i) soil, (ii) plant characteristics (species or genotype) and (iii) management techniques. The soil properties can be determined from pedotransfer functions but these give the mean soil properties for rather broadly defined soil textures classes and therefore provide limited site-specific information (Wösten et al., 1999). The soil properties can also be measured directly on site, but this is very costly and time consuming. Management techniques are usually known as they reflect the farmer’s decisions. The

parameters related to plant growth and development are determined from the literature, from experiments conducted on specific processes included in the model (*e.g.* mineralization rate, critical nitrogen dilution curve) or from calibrations based on large experimental databases (Launay et al., 2005; Flenet et al., 2004). In all cases, the propagation of uncertainty about the parameters could lead to a model that does not accurately describe responses observed in the field.

Parameter estimation is not straightforward in generic crop models. Most of the equations are non-linear, coupled and hierarchical; the number of parameters to optimize is important; and field spatial variability and climatic temporal fluctuations are high. Several methods have been proposed for parameter estimation, based on frequentist or Bayesian approaches (Beven, 1989; Wallach et al. 2006). In the first category are sensitivity analyses (Wallach et al., 2001; Ruget et al., 2002; Bechini et al., 2006; Makowski et al., 2006; Monod et al., 2006; Campolongo et al., 2007; Lamboni et al., 2009) and stepwise regression methods (Wallach et al., 2001, 2006). Recently, Wallach et al. (2009, 2011) developed a software package suited to the STICS crop model (OptimiSTICS) that used the Extended Fast algorithm (also used by Varella et al., 2010a, 2011) to analyse the sensitivity indices.

The Bayesian approaches (Gilks et al., 1996; Jansen and Hagennars, 2004 ; Makowski et al. 2002) are becoming increasingly popular for estimating model outputs and parameters distributions in different types of complex models, like the simulation of biological processes (Minunno et al., 2013), environmental (Dietzel and Reichert, 2012; Rasmussen and Hamilton, 2012), hydrological (Jeremiah et al., 2012; Laloy et al., 2010; Vrugt et al., 2003; Wu and Liu, 2012) or crop modelling (Makowski et al. 2006 ; Varella et al., 2010b). In these approaches, the parameters are considered as stochastic variables defined by the prior distribution of probability. The process aims to sample the posterior distribution of the parameters leading to the statistically most relevant simulations.

Traditionally, it has been difficult to estimate the posterior distribution of parameter estimates and/or the model output predictions, but the use of Markov Chain Monte Carlo (MCMC) simulations (Metropolis et al., 1953; Vrugt et al., 2009b) has made this task easier. The basis of these methods is a Markov chain, which generates a random walk through the search space and iteratively visits solutions with stable frequencies. To do this, an MCMC algorithm generates trial moves from a current position in the parameter space, defined by the actual position in the Markov chain, to a new position in the

parameter space. The earliest and most widely used MCMC approach is the Random Walk Metropolis (RWM) algorithm (Metropolis et al., 1953). One of the particularities of the algorithm lies in the use of the Metropolis acceptance probability ratio (Metropolis et al., 1953) as a selection rule to decide whether or not the candidate parameter set could replace its parents. The result of the algorithm is a Markov chain that, for the values that are sufficiently far from the starting point, has a unique stationary distribution with stable frequencies stemming from the underlying probability density function (pdf).

In 1970, Hastings extended the original MCMC to include non-symmetrical proposal distribution. Called the Metropolis Hastings (MH) algorithm, this extension became the basic building block of many existing MCMC sampling schemes. In the 1990s, much research was devoted to Markov chain sampling (e.g., Gilks et al. 1996; Gelman et al., 1997; Brooks, 1998). Although this research improved the efficiency of MCMC algorithms, they remained inefficient when confronted with posteriors with very heavy tails and with posterior model output prediction surfaces that contained multiple local optima. Recognizing the limitations of previous MCMC schemes, ter Braak (2006) developed the Differential Evolution-Markov Chain (DE-MC) method, which can run simultaneously and in parallel with several Markov chains and uses a genetic algorithm for estimating parameter evolution. DE-MC solves the RWM practical problem of choosing an appropriate scale and orientation for the jumping distribution. Vrugt et al. (2008a, 2009a) proposed a new MCMC sampler called the Differential Evolution Adaptive Metropolis (DREAM) algorithm. DREAM is a follow-up of the DE-MC method and an adaptation of the Shuffled Complex Evolution Metropolis (SCEM-UA) global optimization algorithm (Vrugt et al., 2003). The authors showed how using self-adaptive randomised subspace sampling, with explicit consideration of aberrant trajectories, could still enhance, sometimes considerably, the efficiency of the DE-MC algorithm. Vrugt et al. (2009a) demonstrated that there was an optimal choice for the multiple of the difference of two randomly chosen members from remaining chains used in the genetic algorithm. The advantages of DREAM are summarised here. First, DREAM solves two important problems. One is the automatic selection of an appropriate scale and orientation of the proposal distribution during evolution towards the posterior distribution (i.e., self-adaptive randomized subspace sampling). The second one is the efficient accommodation of

heavy-tailed and multimodal target. Unlike the SCEM-UA algorithm, DREAM can maintain a detailed balance and ergodicity while showing good efficiency for complex and highly non-linear and multimodal target distributions (Vrugt et al., 2009a). DREAM also solves limitations such as the need to choose the starting values and the unlimited number of parameters that could be optimized at the same time (Makowski et al.,2002). Finally, and most recently, Vrugt et al. (2011) have shown how DREAM could be enhanced using parameter sampling from past states of the genetic evolutionary chains, leading to the DREAM-ZS algorithms (Vrugt et al., 2011; Laloy et al., 2012). Let's also mention that in the recent years, another suitable solutions emerged which consist to consider simultaneously parameter optimization and data assimilation (Vrugt et al., 2006, Mansouri et al., 2013).

In recent years, the debate has focused on the use of a formal or informal approach for specifying the likelihood function (Beven et al., 2008; Schoups and Vrugt, 2010; Vrugt et al. 2008b, 2009b). Informal likelihood functions have been proposed as a pragmatic approach to uncertainty estimation in the presence of complex residual error structures. Importance sampling algorithms, such as the Generalised Likelihood Uncertainty Estimation (GLUE) method (Beven and Binley, 1992), are becoming very popular because they have the potential to deal with estimation uncertainty problems where simple theoretical likelihood assumptions are not appropriate (Beven and Binley, 1992; Beven, 2008; Vrugt et al., 2009b). For example, Varella et al. (20010b, 2011) investigated characterizing soil properties in agricultural fields by inverting the STICS dynamic crop model, using the observations conducted in those fields by remote sensing or yield monitoring. This method, however, involves discretisating the parameter space in order to perform optimization, and such an approach could lead to an inaccurate representation of the posterior parameter distribution when the model parameters are numerous (Makowski et al., 2002).

Alternatively, the formal approach starts from an assumed statistical model for the residual errors (Joseph and Guillaume, 2013 ; Laloy et al., 2010 ; Vrugt et al., 2009b). This model, which is specified *a priori*, is then used to derive the appropriate form for the likelihood function that links the model output with the real-life measurements and that should therefore correctly sample the high-probability density region of the parameter space. MCMC simulations then allow behavioural

137 solutions to be separated from non-behavioural ones, using a threshold based on the sampled
138 probability mass. Typically, the residual error assumptions can be classified into three groups relating
139 to (*i*) error variance, (*ii*) error distribution and (*iii*) error correlation. The advantage of the formal
140 approach is that error model hypotheses are stated explicitly and their validity can be verified *a*
141 *posteriori* (e.g., Schoups and Vrugt, 2010). The formal approach, however, has been criticised for
142 relying too heavily on residual error assumptions that do not reflect reality in many applications
143 (Beven et al., 2008). For example, considering that the errors are independent and identically
144 distributed, following a normal distribution with zero mean and constant variance σ^2, the statistical
145 error model would result in the standard least squares (SLS) approach (Box and Tiao, 1973). In many
146 cases, however, and especially in agricultural research, the errors are correlated, non-stationary and
147 non-Gaussian. Correlations between model residuals often arise when several measurements are
148 performed at different dates in a given site-year. Site-year characteristics have a strong influence on
149 observations and, as only a part of the between site-year variability can be predicted by crop models,
150 model residuals obtained in a given site-year have different variances and are often correlated
151 (Wallach et al., 2006).

152 The main objective of this paper is to extend the available parameter estimation tools of the
153 STICS soil-crop model. Currently, DREAM and DREAM-ZS are probably among the most optimized
154 MCMC algorithms than can offer genericity and robustness in the parameter sampling process. On the
155 other hand, the STICS model is widely used and its ability to simulate contrasted situations and to
156 adapt to new species is well recognised; to date, however, parameter optimization is rarely obtained
157 using MH algorithms. The first aim of this paper is therefore to extend the parameter estimation
158 techniques available for the STICS model by using the DREAM-ZS scheme and to assess the coupling
159 of both algorithms.

160 At another level, in-field measurement errors associated with crop modelling experiments is
161 not a trivial problem. To improve the computational efficiency of the sampling MCMC algorithms, the
162 expert knowledge could be expressed at the process initialisation stage through a more appropriate
163 definition (e.g., tightening) of the parameters' *prior* distribution. In our opinion, however, it should
164 also advantageously appear in the likelihood function, making it possible to take account of systematic

error measurements. In this context, and using a formal representation of error assumptions, a new version of the likelihood function was derived that makes explicit use of the coefficient of variation (CV) of the measurements and which should be able to account for heteroscedastic error cases.

2. Material and methods

2.1. Case study

The data used in this paper derive from an experiment designed to study wheat growth response (*Triticum aestivum* L., cultivar Julius) under different nitrogen fertilization levels. The experimental blocks were prepared on two soil types (loamy and sandy loam), corresponding to the agro-environmental conditions of the Hesbaye region in Belgium. The measurements were the results of four repetitions by date, nitrogen level, soil type and crop season. Each repetition was performed on a small block (2 m × 6 m) within the original experiment as a complete randomised block distribution, spread over the field within each soil type, to ensure measurement independence. A wireless microsensor network was used to continuously characterize the soil (water content, suction, temperature at two depths: 30 and 50 cm) and the atmosphere (radiation, temperature, relative humidity) within the vegetation. Pluviometry data were also acquired in the experimental field. Biomass and soil nitrogen content were regularly measured manually.

This paper focuses on the biomass growth, described by the MASEC output within the STICS model, over three years (crop seasons 2008-09 to 2010-11). Two fertilization levels were considered in this study: crop growth (i) without nitrogen feeding and (ii) under a nitrogen level of 180 $kgN.ha^{-1}$ applied in three fractions and according to three equivalent doses, respectively at the tillering (Zadoks stage 23), redress (Zadoks stage 30), and last-leaf stages (Zadoks stage 39). The above ground biomass measurements were performed at a bi-weekly interval from mid-February (about Julian day 410) until harvest. The above ground biomass was defined here as the sum of straw and grain yields. The measurements were performed on dried samples, corresponding to the sampling of three adjacent 50cm rows.

Table 1 summarizes the different identified cultural situations (CS) according to the cropping seasons and the stresses events. For each of the two crop cycles of the first season (CS 1 and 2), ten measurements were performed. Nine aboveground biomass measurements were made for each nitrogen level of the season 2009-10 (CS 3 and 4) while five biomass samples were taken during the last season (CS 5 and 6).

Table 1: The different cultural situations (CS) and the stress effects

Stress effect	No nitrogen stress 180kgN.ha^{-1}	Nitrogen stress 0kgN.ha^{-1}
No water stress Season 2008-09	CS 1 Calibration dataset	CS 2 Calibration dataset
Water stress #1 Season 2009-10	CS 3 Calibration dataset	CS 4 Calibration dataset
Water stress #2 Season 2010-11	CS 5 Validation dataset	CS 6 Validation dataset

2.1.1. Calibration dataset

The first two years of experiments were used to calibrate the model. The 2008-2009 crop was sown in late October (Julian day 297) and harvested in mid-August (Julian day 593). The yields were quite high and close to the optimum of the cultivar, mainly because of the good weather conditions and the sufficient nitrogen nutrition level. In the 2009-2010 season, the crop was sown in early November (Julian day 323) and harvested a bit later than in first year (Julian day 598), due to the poor aestival conditions. This season was characterised by significant water stress that occurred at the early season (February) and in the early summer (July).

2.1.2. Validation dataset

The last year of experiments was used to perform the model validation. During the season 2010-11, the measured yields were close to the ones observed in 2009-10. However, a lower number of tillers and fewer grains per ear were observed. This was a consequence of strong climate-induced stresses, namely an important water deficit and high temperatures at spring (from the middle of March till the end of May). Owing to the return of rain at early summer, the grains have been correctly filled but the straw yield has remained really poor.

2.2. Model description

2.2.1. The STICS crop model

The STICS crop growth model (INRA, France) used in this study has been described in several papers (Brisson et al., 1998; Brisson et al., 2003; Brisson et al., 2009). STICS is a generic soil-crop model that can simulate a broad range of crops. It simulates the water, carbon and N dynamics in the soil-plant-atmosphere system on a day-by-day basis. It allows to take into account the effect of

water and nutrient stress on development rate (Palosuo et al., 2011). It requires daily weather data inputs (i.e., minimum and maximum temperatures, total radiation and total rainfall, vapour pressure and wind speed).

Within STICS, the eco-physiology of aboveground growth is driven by a classic carbon balance : the leaf development allows the interception of the solar radiation, which is converted into biomass and later oriented towards harvestable organs. The whole plant phenology of aboveground growth is driven by the degree-day thermal index [°C-day].

2.2.2. Parameter assumptions

Nine parameters involved in the aboveground biomass growth simulation were selected to be optimised. However, in order to avoid over-parameterization (Varella, 2011; Varella et al., 2010b), the selected parameters were chosen as not being all directly linked to the formalism of the simulated variable (MASEC) : we considered parameters involved in the phenology (*stlevamf*, *stamflax*), the leaf area development (*dlaimaxbrut*, *durvieF*), parameters directly related to biomass growth (*efcroijuv*, *efcroirepro*, *efcroiveg*) and finally related to water and nitrogen stresses (*psisto*, *INNmin*). The remaining parameters of the species were fixed at the suggested default values (Brisson et al., 1998; 2003).

Table 2 summarizes the studied parameters, their initial value and their prior distribution. In this table, the ILEV, IAMF and ILAX stages correspond respectively to the stage of emergence, the day when the leaf growth rate is maximal (AMF stage), and the day when the maximal leaf area index (LAI) is reached. The complete senescence of the crop, conducted by the *durvieF* parameter is reached a few days before maturity of the crop. The radiation use efficiency is known to be different during plant growth. It is lower during the juvenile phase, which extends between emergence (ILEV) and AMF stage (IAMF). It is higher during the vegetative stage, which occurs between AMF stage and flowering, and during the reproductive phase. As an illustration, the Figure 1 shows the biomass measurements performed during the crop season 2008-2009 with the corresponding standard deviation.

Table 2 : Initial parameters values and prior distribution

Parameter	θ^{init}	*Prior values*	*Unit*	Definition
dlaimaxbrut	4.5E-4	[0 - 4E-3]	$m^2_{leaf}.(plant)^{-1}.(°C\text{-}day)^{-1}$	Maximum rate of LAI daily increase
stlevamf	255	[0 - 400]	°C-day	Duration between ILEV and IAMF stages
stamflax	350	[0 - 500]	°C-day	Duration between IAMF and ILAX stages
durvieF	220	[0 - 500]	°C-day	Maximal lifespan of an adult leaf
efcroijuv	1.8	[0 - 4.5]	$g.MJ^{-1}$	Radiation use efficiency during juvenile phase
efcroiveg	4.25	[0 - 10]	$g.MJ^{-1}$	Radiation use efficiency during vegetative stage
efcroirepro	4.25	[0 - 9]	$g.MJ^{-1}$	Radiation use efficiency during grain filling phase
INNmin	0.360	[0 - 1]	/	Minimum value of Nitrogen Nutrition Index allowed
psisto	15	[1 - 20]	bar	Absolute value of the potential of stomatal closing

The lower and upper boundaries of the *prior* parameter distribution were slightly modified compared with the original OptimiSTICS package. They were reduced in order to ensure faster convergence, but they were kept wide enough to produce a sufficiently high parameter space.

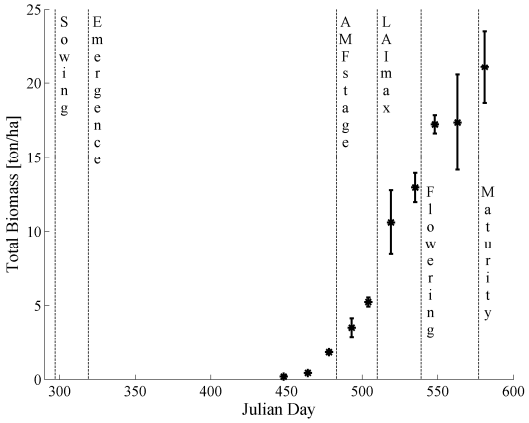


Figure 1: Biomass measurements (mean values and standard deviations), and principal phenological stages of the crop during the cultural season 2008-09.

The parameters were sampled/optimized on the first four contrasted cultural situations, *i.e.* corresponding to the climatic input data of season S.2008-09 and S.2009-10, and to the nitrogen level 0 and $180 kgN.ha^{-1}$ (CS 1-4 in table 1). A total of 38 biomass measurements were used to identify the nine parameters. Once the parameters sampled, the model was then evaluated on the crop season 2010-11 (CS 5 and 6).

2.3. Bayesian theorem, error assumptions and adapted likelihood function

2.3.1. The Bayes theorem

According to the Bayes theorem, the posterior probability density function (pdf) $\pi(\theta|Y)$ is given by following equation:

$$\pi(\theta|Y) = \frac{\pi(Y|\theta)\pi(\theta)}{\pi(Y)} \quad (1)$$

In this expression, θ and Y represent the vectors of the parameters and the measurements, respectively, and $\pi(\theta|Y)$ represents the pdf of the parameters given the observed data and/or measurements. This probability constitutes the *posterior probability* of the estimated parameters. $\pi(\theta)$ is the probability distribution of the parameters to be estimated. This constitutes the *prior probability*, referring to the *prior* knowledge existing about the parameters. It usually consists of a uniform distribution limited by realistic lower and upper bound parameter values. $\pi(Y)$ is the probability distribution of the observed data. It is a constant determined by the requirement that the integral of the posterior distribution $\pi(\theta|Y)$ over the parameter space must equal 1. $\pi(Y|\theta)$ is the probability distribution of the measurements given the parameters and is referred to as the *likelihood function*. Its value is determined from the probability distribution of the error ε_i between modelled and observed data :

$$\varepsilon_i(\theta|Y, X) = \hat{y}_i(X, \theta) - y_i, \qquad i = 1,...,n \quad (2)$$

where n is the total number of observations, $\hat{y}_i(\theta,X)$ is the i^{th} modelled value, according to model inputs X and model parameters θ and y_i is the corresponding observation.

The problem lies in estimating the likelihood function. Assuming that errors, also called residuals (Equation 2), are uncorrelated and Gaussian-distributed (Equation 3),

$$\varepsilon_i \approx N(0, \sigma_i^2) \quad (3)$$

the likelihood function can be simplified, taking the following form (Equation 4; Box and Tiao, 1973):

$$\pi(Y|\theta) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left\{ -\frac{[\hat{y}_i(\theta, X) - y_i]^2}{2\sigma_i^2} \right\} \quad (4)$$

where σ^2_i is the error variance on measurement i.

Finally, for reasons of algebraic simplicity, numerical stability and algorithm implementation, Vrugt et al. (2009b) proposed using the logarithm transformation of the likelihood function:

$$\pi_{\log}(Y|\theta) = -\frac{n}{2}.\ln(2\pi) - \sum_{i=1}^{n}\ln(\sigma_i) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\hat{y}_i(\theta, X) - y_i}{\sigma_i}\right)^2 \quad (5)$$

The actual form of Equation 5 is known as the weighted least square (WLS) function. Instead of the σ_i^2 value, a constant value for the error variance σ^2 could be hypothesised. Such an assumption would consist to consider a constant error variance whatever the measurement dates and their absolute values. In that way, it allows simplifications to be made in Equation 5, which results in the standard least square (SLS) form of the equation. These error assumptions (SLS and WLS), however, are both quite strong and can be unrealistic in crop modelling (*e.g.*, when the measurements are performed at a same location throughout the season).

2.3.2. Experimental design and residual assumptions

As noted above, the SLS and WLS approaches made the assumption that the errors were uncorrelated and (identically or proportionally) Gaussian-distributed. The experimental design was adapted to meet part of this assumption and the original experiment was implemented as a *complete randomised block distribution*.

Applying Fisher's three principles (Preece, 1990) – *replication*, *randomization* and *local control* – allows the error variances to be estimated while increasing the precision of the experiment (diminution of error). More precisely, *randomisation* allows an unbiased estimation of the residual variance to be obtained, whereas *local control* (sometimes called *blocking*) increases the precision of the experiment. The main objectives of the *complete randomised block distribution*, especially its *randomisation* component, is to create experimental units that are as similar as possible in order to reduce, within the blocks, the heterogeneity of the experimental conditions. This allows the spatial correlation to be reduced and, at a lower measure, the temporal correlation between the measurements, which then correspond to an average over replicates.

Each of the y_i-values and the corresponding standard σ_i, deviations needed for the likelihood function calculation therefore resulted from four replicates randomly spread over the experimental field.

2.3.3. Non-stationary and correlation error assumptions

From these in-field observations, it appeared that the averages and the standard deviations of the total biomass measurements increased throughout the seasons, transducing a non-stationarity of the residuals. As the first part of the results section shows, however, the CVs, expressed as the ratio between the standard deviation and the measure (Equation 6), exhibited stationary values:

$$CV = \frac{\sigma_i}{y_i} \tag{6}$$

We therefore decided to introduce CV explicitly into Equation 4 and, after log-transformation, a revised likelihood function was obtained (Equation 7), referred here-after as CV likelihood function:

$$\pi_{\log}\left(Y|\theta\right) = -\frac{n}{2}.\ln\left(2\pi.CV^2\right) - \sum_{i=1}^{n}\ln\left(y_i\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\left(\hat{y}_i\left(\theta,X\right)/y_i - 1\right)}{CV}\right)^2 \tag{7}$$

Typically, crop growth is known to be a heteroscedastic phenomenon. In that way, if the CV is stationary over the seasons and over the years, the proposed formula will offer important advantages. On one hand, if too few measurements are available for practical reasons (such as financial constraints or storm events), the use of the proposed likelihood function would allow the computation of a CV relevant for the whole crop growth cycle, which will increase the efficiency of the parameters sampling process.

On another hand, ideally, the CV value should correspond exclusively to the expression of the crop natural genetic variability. However, for practical reasons, it involves measurement errors, *i.a.* linked to inadequate measurements sampling process or non-adapted equipment. Such errors will be added to the natural variability and may conduct to overestimated CV values. Next to the *prior* definition, the definition of a realistic CV value will thus also allow to express the expert's knowledge at each step of the parameter sampling process.

2.4. Parameter identification and model output uncertainty

2.4.1. The OptimiSTICS parameter optimisation package

The OptimiSTICS package was used as a reference in this study to assess the performance of

the DREAM algorithm. A brief description of OptimiSTICS is given here beneath, a full description can be found in Wallach et al. (2011). OptimiSTICS calculates the parameter values that optimize the goodness-of-fit criterion (for example that minimize a sum of squared errors). OptimiSTICS uses the Nelder-Mead simplex algorithm which can be used for multidimensional minimization for any function. The simplex algorithm used is the Matlab function "*fminsearchbnd*".

However, crop models are complex functions of the parameters and there is no assurance that local optimization techniques will converge to the global optimum. To overcome this problem, in OptimiSTICS, the simplex algorithm is run with several different starting points. The more numerous starting points used, the less the risk of missing the global optimum.

It is worth mentioning that the OptimiSTICS package proposes different options. The software can treat the case where some parameters are genotype specific while others are common to all genotypes. It can also automatically do several sequential stages of parameter estimation. Finally, the software offers the possibility to consider different model errors, including the WLS case (Wallach et al., 2011).

2.4.2. The DREAM algorithm and the associated parameter uncertainty

The origins and developments that led to DREAM were depicted in details in the introduction section. The present section and the following are focused on the advantages offered by DREAM in terms of post-data treatment.

Assessing the posterior distribution of the model parameters using MCMC simulations, performed with DREAM or DREAM-ZS, led to several chains that contained all the necessary information about model parameterization.

The first step in obtaining parameter estimates is to select, among the chains, the parameter set that offers the optimal solution (θ^{opt}), *i.e.* the one that optimises the convergence criterion. However, provided convergence has achieved a stationary distribution, from a statistical/methodological point of view, the information contained in each chain has the same relevance. In a second step, the marginal posterior pdfs were thus evaluated, with the concatenated information contained in each chain (e.g., drawing their histograms). This insight should offer primal information about the quality of sampling,

depending on whether the histograms exhibit a pronounced mode, are bimodal or close to the prior distribution. An interesting discussion about such observations was reported by Laloy et al. (2010).

When designing decision-support tools, it seems necessary for the modeller to summarize the marginal posterior pdf in one parameter estimate. An initial step in assessing the most probable parameter value involves calculating the posterior means (Equation 8), the corresponding standard deviation, and eventually the correlation coefficients between the generated parameter samples.

$$\theta^{mean} = \frac{1}{n \times 2 \times d} \sum_{i=1}^{n \times 2 \times d} \theta_i \quad (8)$$

In this equation, d is the number of sampled parameters and $2 \times d$ is the number of chains, n is the number of last elements in a chain of the sampling process, when each chain exhibits a stable posterior parameter distribution, and θ_i is one of the numerous probable values for the parameters. The number of chains was fixed as two times the number of parameters ($2 \times d$). In this study, the last $n=1000$ elements of each chain were compiled in order to calculate the mean of each parameter value.

2.4.3. The DREAM algorithm and the output predictive uncertainty

In addition to parameter uncertainty, we were also interested in the predictive uncertainty linked to the corresponding model output. The posterior distribution of the model parameters derived with DREAM or DREAM-ZS contains all the information needed to summarize predictive uncertainty (Vrugt et al., 2009b). A common and easy approach is to evaluate the model output Y for the last P parameter sets of each chain ($2 \times d$ chains) when convergence has been achieved for a stationary distribution. The so-obtained model output set $\{Yj, j = 1,...,2 \times d \times P\}$ is summarized in the desired way, *e.g.* by computing the 2.5% and 97.5% percentiles of the model predictions, which difference corresponds to the 95% uncertainty boundaries. This predictive distribution includes only the effect of parameter uncertainty (Vrugt et al., 2009b). The wider the parameter posterior distribution, the wider the 95% boundaries. In addition, the 50% percentile simulation could also be used to evaluate model performance, and be compared with the $f(X,\theta^{mean})$ simulations.

In this case, the last 1.000 sets were no longer considered. To reduce the simulation time, the dataset was reduced to the last 30 values of each Markov chain. Since there are 18 chains

(*2×d* parameters), the parameter uncertainty evaluation in the model MASEC output was summarized in the percentile computation of $30 \times 18 = 540$ simulations.

2.4.4. The sampling process

Even if the STICS model has been widely used to study and simulate wheat growth, Belgian cultivars differ from French ones, notably by their phenology and yields. For a first evaluation of the model, the original parameters file of the wheat species remained at the suggested default values (Brisson et al., 1998; 2003) included in the STICS software. This case was referred to as the *initial case* and θ_{init} represents this initial parameter set.

As a first parameter optimisation technique, the OptimiSTICS package was used. In accordance with the requirements of the DREAM algorithm (see below), 18 starting points were used and randomly generated among the prior knowledge one owned about parameter, *i.e.* it's *a priori* distribution. When running OptimiSTICS, the residuals were considered as being independent errors, with zero expectation and the same variance, which corresponds to the same assumptions as for the SLS case run with the DREAM algorithm (see below). This case is referred later as OptimiSTICS-SLS. The selected parameter set was the one that gave the minimum error (Wallach et al., 2011), *i.e.* the one that should offer the optimal solution (θ^{opt}).

The DREAM-ZS algorithm was then used to perform parameter sampling of the STICS model. To evaluate its performance, various assumptions about the error measurements were considered and taken into account for different likelihood functions.

The first case made use of a classical sum of squared error to represent the likelihood function, in line with the frequentist approaches. Since simplification appeared in the algorithm, the constant standard deviation disappeared and the measurements were considered only by their mean value. This case was referred to as the DREAM-SLS case, and θ_{SLS} represented the corresponding optimised parameter set. The second case corresponded to the weighting, within the likelihood function computation, of residual data by the nominal standard deviation calculated on the basis of the four replicates of in-field measurements, and relied on implementing Equation 5. This case, corresponding

421 to the DREAM-WLS, was represented by θ_{WLS}. Finally, DREAM's ability to retrieve parameter values
422 was evaluated against the error measurement assumption making an explicit use of the CV (Equation
423 7). This case will be referred as DREAM-CV. Table 3 summarizes all the error measurement
424 assumptions.

425 With regard to the DREAM options, the toolbox was run a maximum of 22,500 times, which
426 corresponded to 2,500 evaluation functions multiplied by the number of parameters (d = 9). This value
427 was checked on preliminary studies to ensure convergence. The number of Markov chains was fixed at
428 18 because there were nine parameters to be estimated (MC ≥ 2d, Vrugt et al., 2009a).

429 In each cases, a single-step calibration procedure, involving all the variables (i.e. the MASEC
430 output of the 4 CS) and all the parameters to optimize, was used instead of a multiple-step
431 optimization procedure (Guillaume et al., 2011).

432

433 **Table 3: The different cases considered for measurements errors**

Case	Error assumption	Error value
OptSTICS-SLS & DREAM-SLS	Variance fixed for all measurements (whatever date or observation value)	-
DREAM-WLS	Nominal variance value computed from replications of observed values	σ_i
DREAM-CV	Global CV value computed from all replications of observed values	0.145

434

435 ## *2.5. Evaluation of the global model output estimates*

436 A crop model is a good representation of reality if it can be used to predict observable
437 phenomena in the range for which it was calibrated (Loague and Green., 1991). This underlines the
438 need to define criteria that will determine whether a model is 'acceptable', in pursuit of set objectives.
439 The first criterion is the Root Mean Square Error (RMSE):

440 $$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\left(X, \theta^{post}_{1,\ldots,d}\right)\right)^2} \quad (9)$$

441 where n is the number of observations, y_i is an available observation of the Y measurement vector, and
442 $\hat{y}_i$ is the corresponding simulated value, which relies on the vector X of inputs. $\theta^{post}{}_{1,\ldots,d}$ represents the
443 vector of d parameter estimated on the *posterior* distribution, using one of the proposed techniques.

444 The model efficiency (EF) criterion presents an upper boundary, which facilitates its
445 interpretation and makes it suitable for comparing different situations:

446 $$EF = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i \left(X, \theta_{1,\ldots,d}^{post} \right) \right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}_i \right)^2}, \quad EF \leq 1 \qquad (10)$$

447 If the model is perfect, then $y_i = \hat{y}_i$ for each i, and $EF = 1$.

448 Ultimately, the normalised deviation (ND) criterion shows the tendency of the model to
449 provide under- or over-estimations, overall, of the real case. This parameter can be positive or
450 negative, but is ideally equal to zero.

451 $$ND = \frac{\sum_{i=1}^{n} \hat{y}_i \left(X, \theta_{1,\ldots,d}^{post} \right) - \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} y_i} \qquad (11)$$

452 RMSE, EF and ND are rarely used alone for evaluating model quality. Brisson et al. (2002)
453 and Beaudoin et al. (2008) used RMSE, EF and ND jointly, on the basis that model calibration or
454 validation is accurate if the *RMSE* is relatively low compared with the mean of the observations, and if

455 $$\begin{aligned} EF &\geq 0{,}5 \\ |ND| &\leq 0{,}1 \end{aligned} \qquad (13)$$

456

457 ## *2.6. Software availability*

458 The software programs (STICS-OptimiSTICS and DREAM) are libraries of Matlab® functions
459 divided into several sub-packages. The STICS interface sub-package is based on the OptimiSTICS
460 codes and is responsible for managing the STICS simulations and their inputs and outputs. The
461 OptimiSTICS codes were obtained upon request by the authors (emmah_web@paca.inra.fr). This
462 sub-package writes inputs and parameter values into the ASCII files read by STICS, called the STICS
463 executable function, and reads the model outputs from the ASCII files written by STICS.

464 The DREAM and DREAM-ZS source codes were obtained from the developer
465 (jasper@uci.edu). Interested users should contact him directly. Other options specific to the DREAM

466 toolbox were discussed by Vrugt et al. (2008a, 2009a).

3. Results and discussions

3.1. Spatial and temporal independence of the biomass coefficients of variation

First, the value of each individual *CV* was calculated for the data obtained for each soil type (2), nitrogen level (7) and date of measurement (±10 per season) (Figure 2) in the original experiment. The linear regression applied to the whole data set took the following form:

$$CV = a.Day + b \tag{14}$$

with '*Day*' being the Julian day of the measurement, and *a* and *b* the parameters. The *a* slope and *b* parameters were respectively equal to -0.0002 (with a 95% confidence interval [-0.0005 ; $+2.698.10^{-5}$]) and 0.2555 (with a 95% confidence interval [0.1187 ; 0.3922]). Considering that (*i*) the block distribution was a complete randomized experiment, (*ii*) the 95% confidence interval of the *a* slope parameter included the zero value and (*iii*) that the coefficient of determination R^2 was low (0.0139), the measurements could be considered as being independent. A mean *CV* value was computed from all measurements (0.145) and introduced in Equation 7.

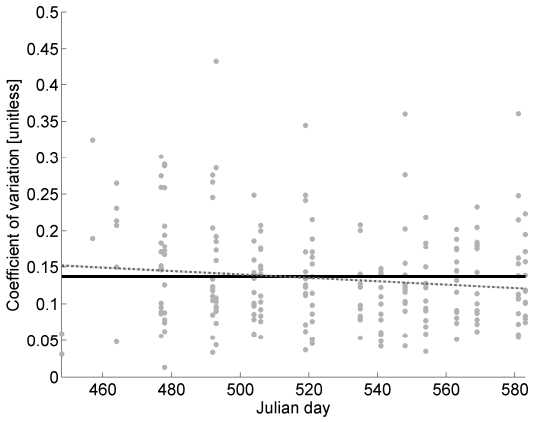


Figure 2: Coefficients of variation (CV) of the total biomass measurements (grey dots). Overall mean value (solid black line) and linear regression (dashed grey line - Eq. 14).

3.2. Parameters identification

As an example, Figure 3 presents the marginal pdf of parameters estimates when the sampling process had achieved a stationary distribution at the end of the WLS process. The results are given for four parameters: *stamflax*, *efcroijuv*, *efcroiveg* and *psisto*. The grey bars represent the histograms drawn using data computation from all the Markov chains.

Figure 3 shows four contrasted cases of marginal pdf. The parameters *efcroijuv*, *efcroiveg* exhibited a marked mode. The *efcroijuv* parameter showed a left dissymmetry in its pdf, which signified that very low values were rejected during the sampling process. The *stamflax* parameter had a relatively irregular shape, indicating some uncertainty about its most likely value. However, the existence of a probable dominant mode around 200 degree-day is clearly noticeable. Finally, the *psisto* parameter showed a pdf clearly close to its prior distribution. This observation may results from two different sources. On the one hand, the STICS model is known to have little sensitivity to the *psisto* parameter (Ruget et al., 2002). On the other hand, as the *psisto* parameter is the critical potential of stomatal closure, one may suppose that the number of observations performed during the water stress events was not high enough to parameterise the model. The plant water potential being seldom reached and/or observed in this rain fed experiment conducted under a temperate climate, the sampling process led to high uncertainty of the posterior distribution of the parameter.

Tables 4 and 5 present the parameter estimates at the end of the various sampling processes. Except for the *psisto* parameter, the optimised parameter set obtained with the OptimiSTICS-SLS algorithm and the sampled parameter set obtained with the DREAM-SLS approach were very close. It would also appear that the close results obtained using OptimiSTICS or the DREAM-SLS case did especially differ from the DREAM-WLS case for the *stamflax* parameter and the three radiation use efficiencies. Finally, apart from the *psisto* and *stamflax* parameters, the DREAM-CV approach tended to converge on the same parameter estimates obtained in the DREAM-WLS case.

With regard to Table 5 which focuses on the DREAM-WLS case, the mean estimators were evaluated in comparison with the absolute optimal estimates that might have been obtained through all the chains. Apart from the *stamflax, durvieF* and *psisto* parameters, the mean estimators were very close to the optimal estimates. The three previous parameters exhibit a marginal shape with high uncertainty (Figure 3). Such differences between the two values could result from an insufficient number of function evaluations, or might appear when the parameter to optimize has a shape without a pronounced mode, which often occurs when at least one of the parameter's *prior* boundaries is taken too close to the final value, when the parameter is physically bounded and exhibits a bimodal pdf (Laloy et al., 2010) or when it shows a tail in the posterior distribution (e.g., *stamflax* parameter).

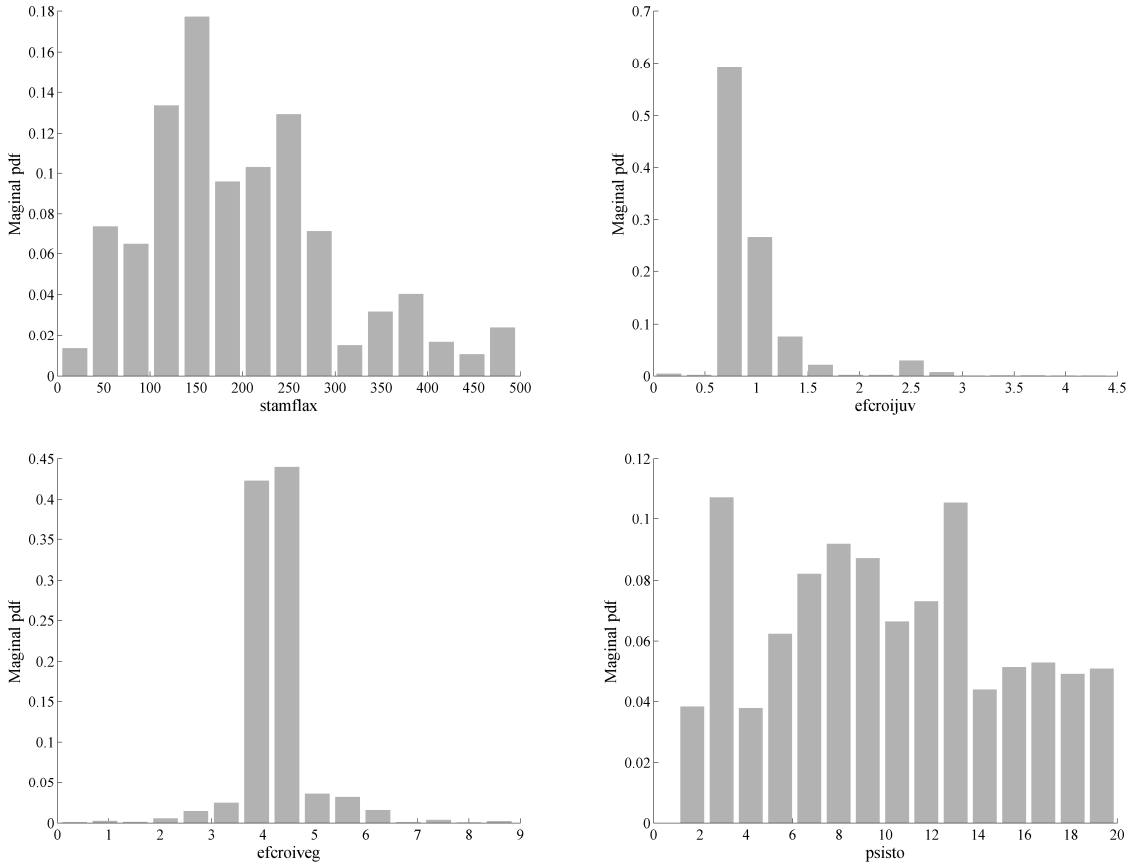


Figure 3: Marginal pdfs for the *stlevamf*, *efcroijuv*, *efcroiveg* and *psisto* parameters. Histogram of the parameter estimates at the end of the DREAM-WLS process.

Table 4: Parameter estimates θ^{mean} at the end of the sampling processes.

Case	OptStics-SLS	DREAM-SLS	DREAM-WLS	DREAM-CV
dlaimaxbrut	1.5E-3	1.4E-3	1.4E-3	1.5E-3
stlevamf	328	324	332	326
stamflax	386	406	198	321
durvieF	370	354	350	347
efcroijuv	0.69	0.41	0.98	1.06
efcroiveg	6.26	6.03	4.26	3.90
efcroirepro	4.49	4.64	5.75	5.86
INNmin	0.29	0.35	0.39	0.45
psisto	6.76	10.55	10.10	6.56

Finally, another interesting aspect of the DREAM and DREAM-ZS sampling algorithm lies in the possibility of studying parameter correlation (Table 5). Moderate to strong correlations were found between model parameters, especially between the radiation use efficiency coefficient (*efcroijuv* and *efcroiveg*) and the *dlaimaxbrut* parameters, which latter controls the overall leaf area index (LAI) development. In particular the correlation between *efcroijuv* and *dlaimaxbrut* was the strongest, with a correlation coefficient of -0.84. It highlighted the important effect of both parameters on LAI and biomass output, during the early growth, *i.e.* before AMF stage. A high value of *dlaimaxbrut* would

lead to an important increase of leaf area, which would have to be compensated by a lower efficiency of radiation use.

The solar radiation use efficiency coefficients are strongly negatively correlated in pairs, *efcroijuv vs. efcroiveg* (-0.59) and *efcroiveg vs. efcroirepro* (-0.26). It clearly meant that an under- or overestimation of one parameter of the pairs was compensated during the next phenological stage to avoid the under- or overestimation of the global simulations in front of the measurements.

Overall, parameters were logically correlated in relation with the preceding stage or the stage during which they are the most expressed (*e.g. efcroijuv* during the *stlevamf* or *stamflax* stages), while poor correlations were observed for parameters referring to different formalisms/physiological aspects (*e.g. durvieF* and *psisto*).

Table 5: Summary of statistics of the marginal posterior parameter distribution in the DREAM-WLS case: optimal parameter set (θ^{opt}), posterior (θ^{mean}), posterior standard deviation (STD), and correlation coefficients over 18,000 generated samples.

Parameter	θ^{opt}	θ^{mean}	*STD*	*dlaimaxb.*	*stlevamf*	*stamflax*	*durvieF*	*efcroijuv*	*efcroiveg*	*efcroirep.*	*INNmin*	*psisto*
dlaimaxb.	1.5E-3	1.4E-3	3.0E-4	1	-0.22	-0.47	-0.03	-0.84	-0.60	0.16	-0.05	-0.08
stlevamf	336	332	10.37		1	0.16	-0.19	0.29	0.49	-0.06	-0.25	0.06
stamflax	328	198	48.41			1	0.12	0.46	0.11	-0.15	-0.04	0.08
durvieF	280	350	33.91				1	0.04	-0.08	0.06	0.24	-0.03
efcroijuv	1.05	0.98	0.09					1	-0.59	0.01	-0.09	0.36
efcroiveg	4.05	4.26	0.21						1	-0.26	-0.46	-0.07
efcroirep.	6.86	5.75	0.69							1	0.15	0.11
INNmin	0.47	0.39	0.02								1	-0.27
psisto	13.2	10.10	2.45									1

Although the parameters were selected to avoid over-parameterization, it appeared that, at the end of the sampling process, some of them were sometimes highly correlated. Remembering that these parameters were not directly linked to the formalism driving the simulated output variable, these results highlighted that the information contained in the measurements were probably not sufficient to identify and accurately estimate all nine parameters. This could never have been shown with a classic Simplex algorithm as it doesn't provide any information on distributions or correlations of parameters.

These observations suggest to adapt the experimental design to the modelling expectations. First of all, the selected parameters should be optimised on the output variable which they impact the most directly the process. In this case study, it would correspond to measure other model outputs (*e.g.* LAI measurements or phenological observations to sample *stlevamf*, *stamflax* and *durvieF* parameters). Another adaptation would be to increase the measurement frequency when needed, *i.a.*

557 during the phases where the growth is the fastest or focusing on identified stress events. Last proposal
558 would be to increase the degree of variation in model driving variables, *i.e.* the weather sequences,
559 and/or the controlling variables, like the assessed experimental nitrogen fertilisation level. These
560 remarks corroborate the researches of Beaudoin et al. (2008) and Basso et al. (2010) who highlighted
561 the importance of numerous measurements and long-term experiments, respectively for the parameter
562 optimisation process and the study of crop yield answer as response to climatic variability.
563

564 ### *3.3. Uncertainty on the predictions for the calibration dataset*

565 As highlighted above, correlation may be strong between parameters. The strength of
566 Bayesian techniques is that one can cope with such correlated parameter sets. After convergence, the
567 posterior distribution of the model parameters derived with DREAM may be used to compute model
568 outcomes ensembles. The predictive uncertainty can then be summarized by model outcomes
569 averaging and confidence interval computation. However, in front of the important computational time
570 needed by such a procedure, it may be interesting to use a unique parameter set. The next two
571 paragraphs will thus focus on the comparison of the set of mean values of parameters and the
572 simulations associated to the posterior distribution of the model parameters.

573 Figures 4, 5 and 6 present respectively the results of the model output simulations after three
574 sampling processes: (*i*) the DREAM-SLS case, (*ii*) the DREAM-WLS case and (*iii*) using the realistic
575 CV value (DREAM-CV).

576 It seems that the DREAM-SLS approach led to final selected parameter estimators that tend to
577 bias the model output simulations (Figure 4), especially at the early stages, from sowing until Julian
578 day 470. The same phenomenon, and pretty close simulations, were observed with the parameter set
579 obtained at the end of the OptimiSTICS-SLS optimisation process. The corresponding growth of this
580 physiological stage is governed mainly by the *dlaimaxbrut*, *stlevamf* and *efcroijuv* parameters. Since
581 the *dlaimaxbrut* parameter governs the whole LAI growth, and because its value converges
582 approximately on the same value whatever the process, this parameter could be considered as correctly
583 estimated.

584 With regard to Tables 4 and 5, the parameter estimates obtained for *stlevamf* seemed correct

compared with the DREAM-WLS case, although the *efcroijuv* parameter effectively did not converge towards realistic physical values using DREAM-SLS. The lower value during this early stage was then compensated with a higher *efcroiveg* value applied after the *stamflax* stage, until the initiation of flowering.

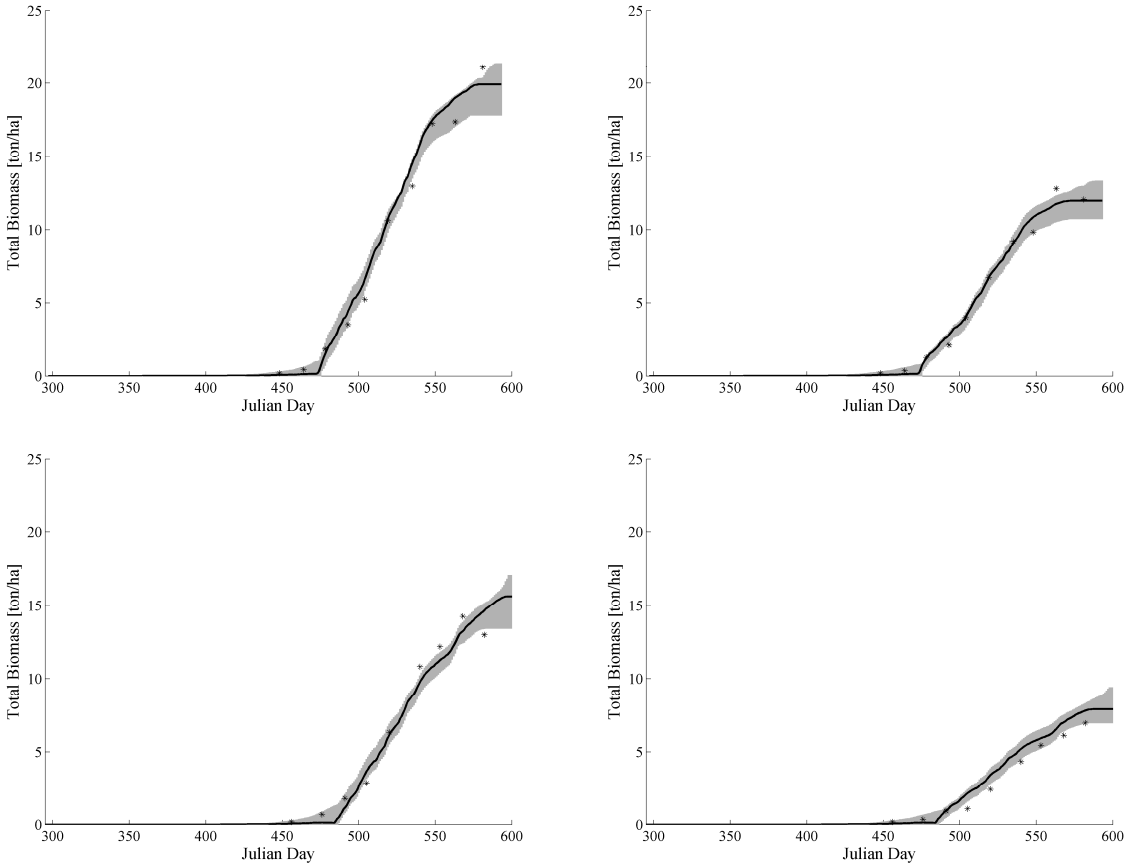


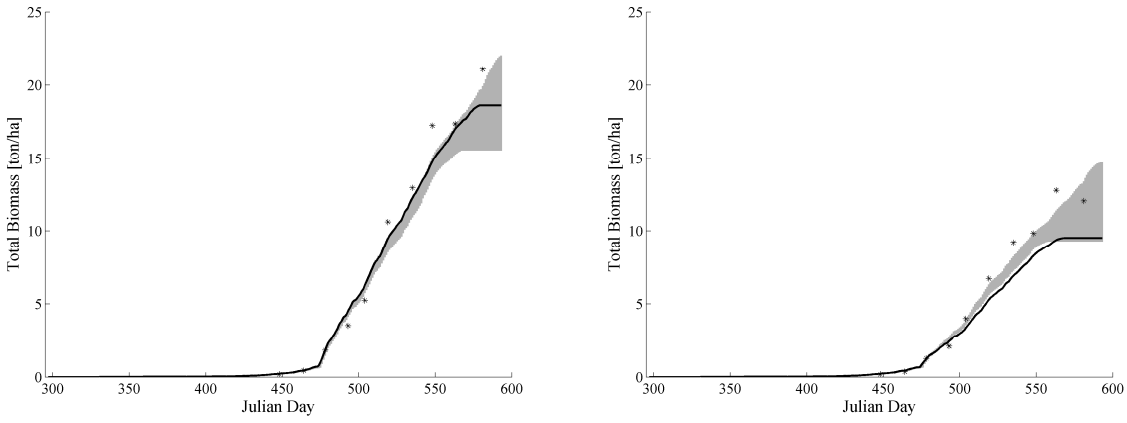
Figure 4: Model output simulations for the DREAM-SLS case. No nitrogen cases (right) and 180kgN.ha^{-1} (left). Winter wheat growing season 2008-09 (upper) and 2009-10 (low). Light grey area represents the 95% uncertainty boundaries. Solid black line represents the simulations obtained with mean estimates for parameters.

In addition, when comparing the DREAM-SLS process with the DREAM-WLS approach, the particular shape of the uncertainty boundaries is noticeable. In the DREAM-SLS case, the 95% uncertainty boundaries exist at the start of the growth, and can be observed from Julian day 450. They seem relatively constant throughout the growing season and for all CS. Analysing the results obtained with the WLS likelihood function, one can immediately notice (i) the precise match of the observation at early stage (before Julian day 500) and for the same period, and (ii) the extremely tight boundaries around the simulated output. These observations are consistent with the assumptions made in the

statistical errors model.

The comparison of the DREAM-WLS case (Figure 5) and the DREAM-CV approach (using a realistic CV value of 0.145 - Figure 6) gives similar results, both in terms of uncertainty interval and simulations based on parameter values selected with the mean estimators. In both cases, the uncertainty intervals are very tight around the simulated output ~~small~~ at the early stages, but widen at the end of the simulation curve. This is due to the noise/standard deviation which is increasing proportionally to the absolute value of measurements (heteroscedasticity), as previously mentioned. The comparison from both the DREAM-WLS and DREAM-CV cases (Figures 5 and 6) showed thus pretty close simulations. This expected result is consistent with theory and the errors model defined within the likelihood function, but it allowed us to conclude that the proposed formula was correctly implemented and computationally as efficient as the WLS likelihood function. Nevertheless, since it takes account of the natural genetic variability of crop species, the proposed formula (Equation 7) opens the door to a new approach in parameter identification. Deeper considerations about the use of such a function are described in the conclusion section.

The other observation concerns case-to-case analysis of the different cultural situations. In general, the sampling/optimization process leads to simulations that fit the measurements properly, taking account of nitrogen and water stresses. Since the model evaluation criteria are of poorer quality compared with the DREAM-SLS case, it seems right that the simulation relying on the WLS or CV likelihood function are slightly further from the measurements.



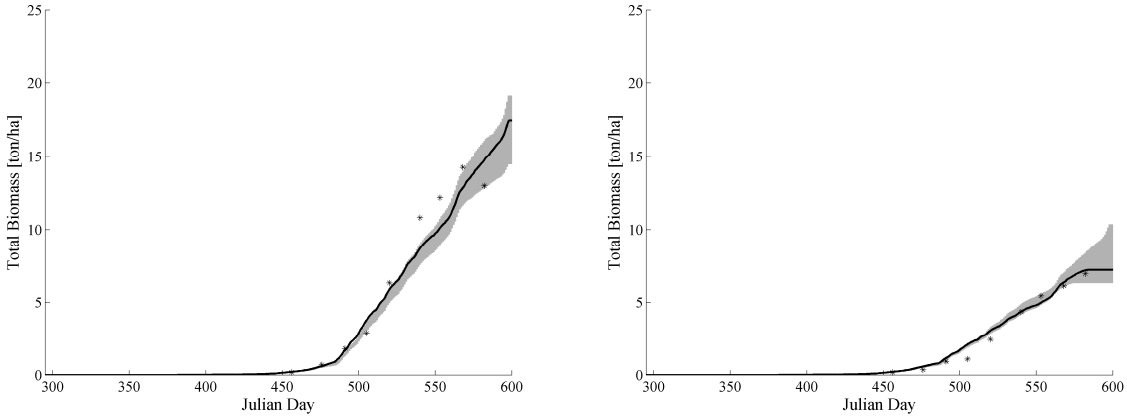


Figure 5: Model output simulations for the DREAM-WLS case. No nitrogen cases (right) and 180kgN.ha^{-1} (left). Winter wheat growing season 2008-09 (upper) and 2009-10 (low). Light grey area represents the 95% uncertainty boundaries. Solid black line represents the simulations obtained with mean estimates for parameters.

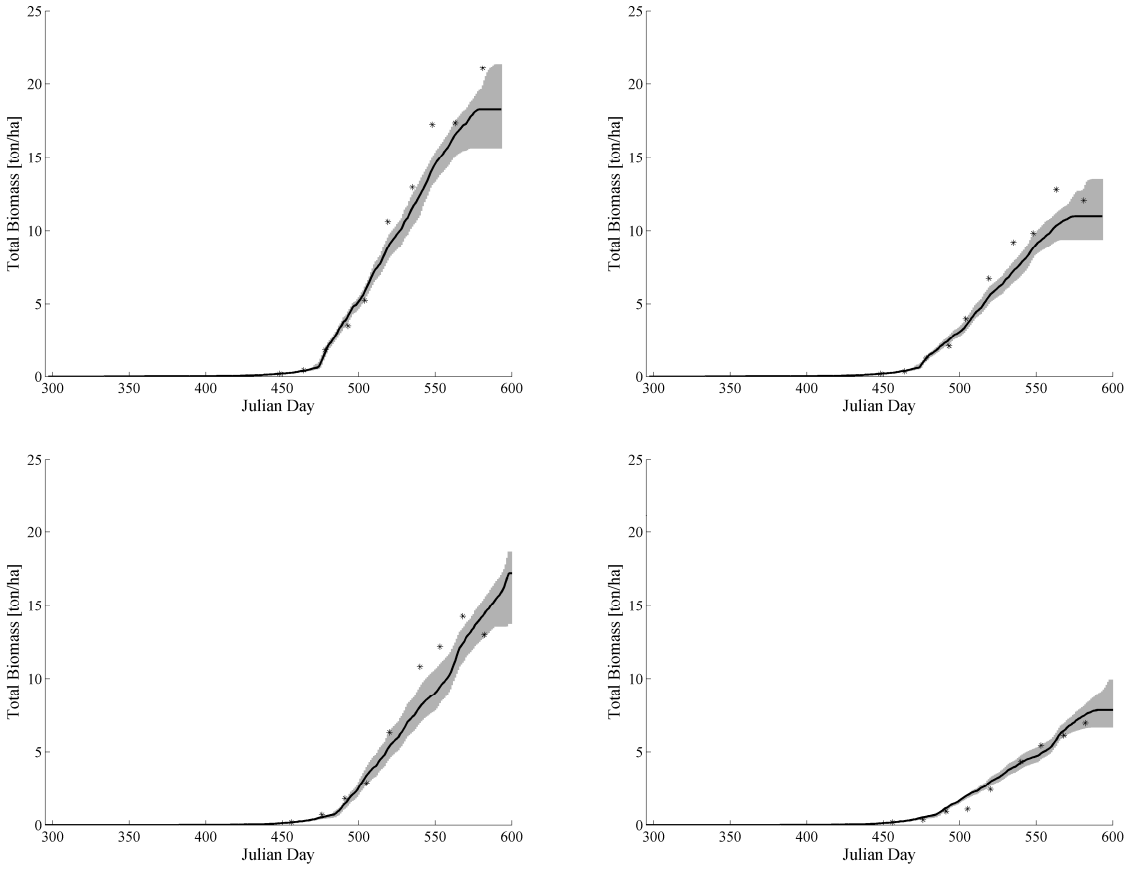


Figure 6: Model output simulations for the DREAM-CV (CV = 0.145) case. No nitrogen cases (right) and 180kgN.ha^{-1} (left). Winter wheat growing season 2008-09 (upper) and 2009-10 (low). Light grey area represents the 95% uncertainty boundaries. Solid black line represents the simulations obtained with mean estimates for parameters.

3.4. Evaluation of the overall model quality

Figure 7 presents the results of the model output evaluation criteria, both for the calibration

637 and validation procedures. The grey-scale histograms correspond to different model output
638 simulations: simulations performed on the basis, respectively, of the mean parameter density estimates
639 (light grey) and the 50% percentile of the last 540 simulations (dark grey). The horizontal black line
640 represents the initial run of the model, based on the initial STICS parameter set.

641 In the sampling processes conducted on the calibration dataset (left graphs on Figure 7), the
642 results of the model evaluation were greatly improved compared with the initial run. RMSE was
643 divided at least by two. The EF criterion, already superior to 0.5, was nevertheless improved and was
644 close to 1. The ND criterion was also enhanced, and was always lower than the expected 0.1 value.
645 With regards to the thresholds generally considered in crop modelling (Brisson et al., 2002 ; Beaudoin
646 et al., 2008), the model was considered has being correctly calibrated.

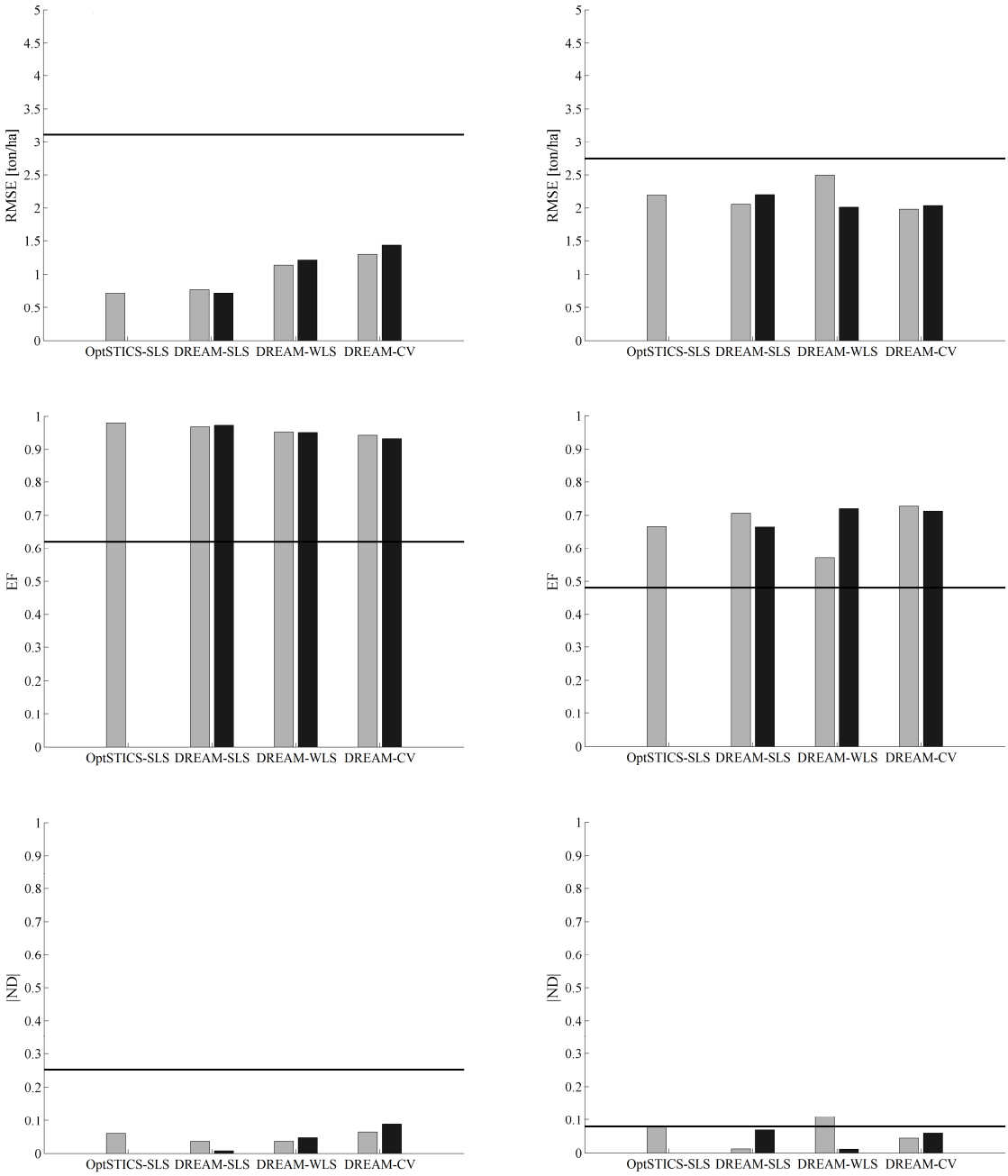
647 Comparing the various optimisation/sampling processes, it appeared that OptimiSTICS-SLS
648 and DREAM-SLS cases always gave the better results. Their similar objective functions, i.e. the
649 minimization of the RMSE between simulations and calibration data, explained why they converged
650 on similar parameter set, and thus gave obviously the best RMSE on the data used for the calibration.
651 The DREAM-WLS and DREAM-CV cases showed also and quite logically similar performances.

652 It is worth mentioning that, in all the calibration cases, the mean parameter set obtained at the
653 end of the sampling process led to similar results than the 50% percentile computed out of the last 540
654 simulations.

655 Considering the validation dataset (right graphs on Figure 7), the three criteria were enhanced
656 in comparison of the initial run, even if the performances were slightly lower than in the calibration
657 run. The RMSE was approximately 1 t/ha lower than the initial run. The model efficiency (EF) which
658 was below the 0.5 threshold under the initial parameter set, was improved till more or less 0.65. The
659 ND criterion, was improved under all the considered error assumptions and remained always under the
660 validation threshold of 0.1. In presence of these results, the model was considered as validated
661 whatever the error assumption made.

662 Concerning the intercomparison on the validation dataset, the four error assumptions led to
663 quite similar results. One could however notice the lower quality of the simulations obtained with the
664 mean parameter set computed at the end of the DREAM-WLS sampling process. It was shown that the

665 posterior distribution of parameters could have multimodal or dissymmetrical distributions. While the
666 correlation between parameters was obviously maintained during the sampling process, due to the
667 shape of posterior distribution, a set of parameters calculated as the mean of the last given element of
668 all chains may thus not necessarily represent a combination that will provide a good model evaluation,
669 especially when assessed on an independent/validation data set.
670

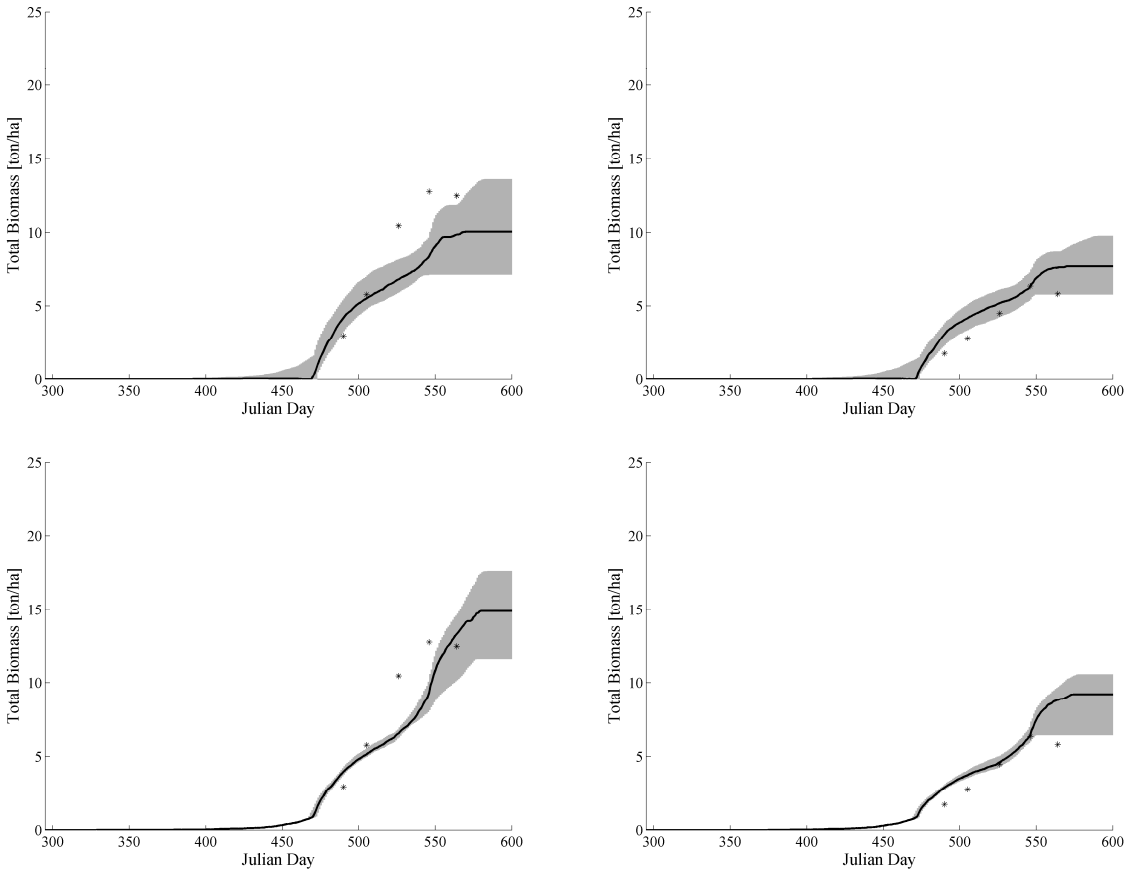


671
672
673

674 **Figure 7 : Model evaluation criteria based on the calibration dataset (left) and the validation dataset**
675 **(right). Initial model run (horizontal black line). Model evaluation using the parameters estimated with**
676 **mean density (light grey), and the percentile 50% of the 540 model output simulations (black).**

3.5. Uncertainty on the predictions for the validation dataset

To validate previous statements, the temporal evolution of the model's outputs were computed for the validation dataset and for the different error assumptions (Figure 8). As previously observed, in the DREAM-SLS case (and this was also true with the parameter set obtained with OptimiSTICS-SLS), the simulations did not respect the biophysical behaviour of plants at early stages, compensating later the biomass growth with higher radiation use efficiencies. The results obtained under the 'no nitrogen' case and for the DREAM-WLS and DREAM-CV error assumptions were satisfactory. With 180 kgN/ha, the simulations were of poor quality when compared to the measurements. In a global way, the crop season 2010-11 is known to be particularly challenging in terms of modelling, since water deficits occurred, deeper than the ones observed in 2009-10. Although the criteria used to evaluate the model quality encountered the validation thresholds, the temporal evolution indicates the need for model improvement by selecting other or more parameters for identification. It may also suggest the need to use other formalisms better adapted to take into account water deficits.



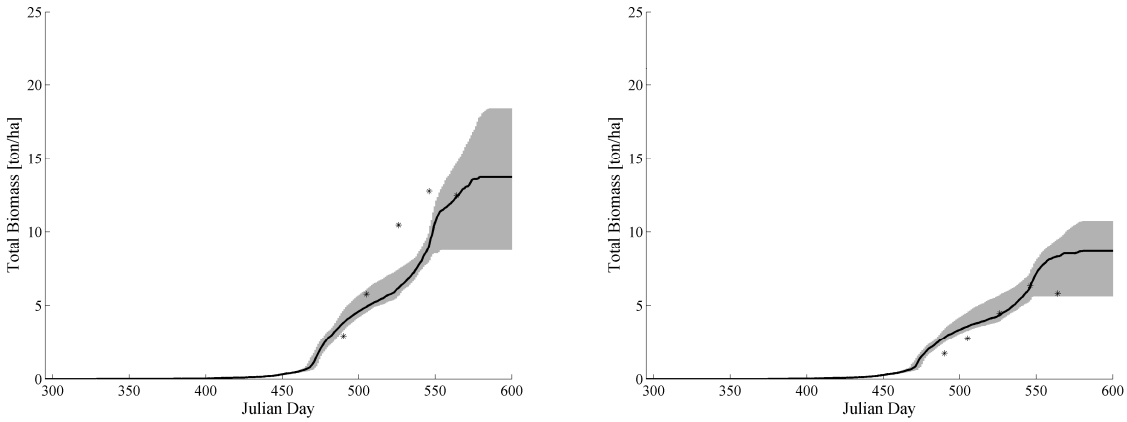


Figure 8: Model output simulations for season 20010-11, under 180kgN/ha (left graphs) and 0 kgN/ha (right graphs), and for the different error assumptions, DREAM-SLS case (upper graphs), DREAM-WLS case (middle graphs) and DREAM-CV case (Lower graphs). Light grey area represents the 95% uncertainty boundaries. Solid black line represents the simulations obtained with mean estimates for parameters.

3.6. Residual analysis

The Figure 9 shows the error analysis results for the residuals between measurements and the fitted models using the SLS, WLS and CV likelihood functions for the four cultural cycles (CS 1 to 4). Three aspects are considered: (*i*) the residuals versus the simulated biomass, (*ii*) the comparison of assumed and observed pdf and (*iii*) the partial autocorrelation coefficients of residuals.

In the DREAM-SLS case, the distribution of residual errors against simulated biomass appeared quite stationary, suggesting homoscedasticity. Although the number of measurements was low, the error histogram did not seem inconsistent with the assumed Gaussian pdf. Finally, errors were not correlated whatever the lag, highlighting the independence of the measurements. The assumptions for the statistical error model were thus *a posteriori* validated.

The error analysis graphs drawn using the DREAM-WLS or DREAM-CV approaches were relatively close. In both cases, the residuals seemed to increase with simulated biomass, up to 10 t/ha and were quite constant thereafter (between 10 and 20 t/ha). The residual histograms did not really match the expected Gaussian distribution (e.g., if the histograms peaked around the zero value, there was a dissymmetry and a tail in the positive values of residuals). Finally, the residuals were at least significantly correlated at a lag of one measure, but in both cases this correlation was very close to the 95% boundaries.

In line with the graphical analysis depicted in the simulations obtained under the DREAM-WLS or DREAM-CV cases (in Figures 5 and 6), it appeared clear that the model underestimated the measurements step by step, meaning that if at a particular moment a measurement was underestimated, there would be a tendency for this to happen with the next measure. Moreover, the underestimation seemed to appear only after Julian day 520, when half the biomass had yet to be produced (for biomass superior to 10t/ha).

As mentioned by Wallach et al. (2006), since only a part of the between site-year variability can be predicted by crop models, correlations between model residuals often arise when several measurements are performed at different dates in a given site-year. However, this issue was acknowledged by the application of three mechanisms, respectively (*i*) the implementation of the experiment as a *complete randomised block distribution*, (*ii*) the consideration of the coefficient of variation within the error model, and finally (*iii*) by the implementation of a single-step sampling procedure, involving all the parameters to optimize and the variables output of the 4 cultural situation at the same time (Guillaume et al., 2011).

For these reasons, and assuming that observational climatic inputs and in-field output measurement errors were negligible, meaning, in our case, that they are due solely to genetic variability, the systematic under-estimation of the in-field biomass samples for the higher values of biomass could be fully attributed to modelling inadequacies. Rather than enabling the conclusion that could be drawn about the non-stationarity or the non-normality of the error model, it suggested the need for model improvement, whether by selecting other (or more) parameters or other formalisms (stresses effects).

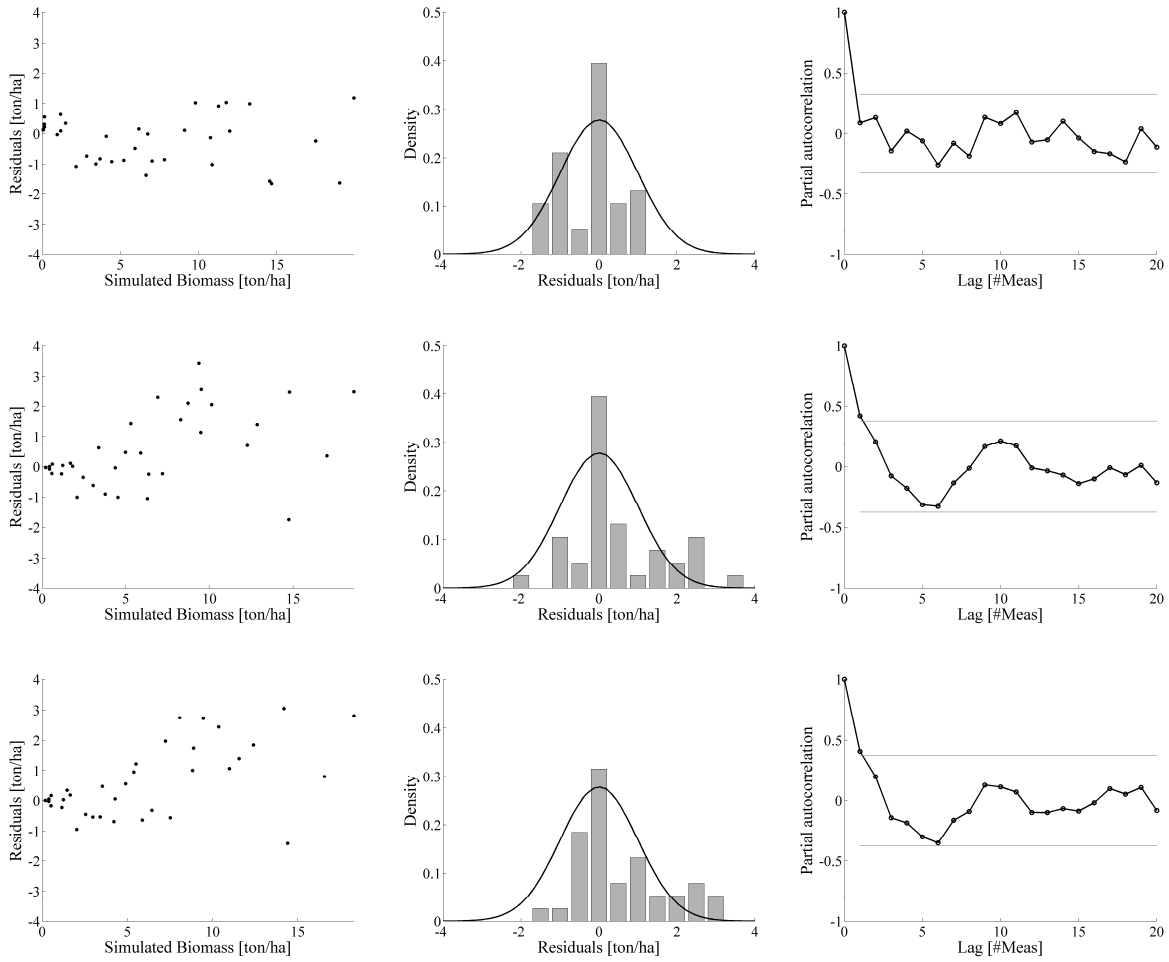


Figure 9: Model residual analysis, for three different likelihood functions: SLS (upper graphs), WLS (middle graphs) and likelihood function for which CV =0.145 (lower graphs). Analysis of the residuals against biomass simulation (left graphs), assumed (solid black line) and observed (grey bar) pdf of residuals (centred graphs) and partial autocorrelation coefficients of residuals (solid circle black line) with 95% significance boundaries (horizontal grey line) (right graphs).

4. Conclusion

This study assessed the potential of the DREAM algorithm for optimizing the STICS crop model parameters with the aim of improving the simulations of the biomass growth of a winter wheat culture (*Triticum aestivum* L.).

Nine parameters involved in leaf area development, radiation use efficiency and stress effects were chosen for optimization of the biomass growth output. Different likelihood functions and error assumptions were evaluated: a standard least square (SLS), a weighted least square (WLS) and a transformed likelihood function that makes explicit use of the coefficient of variation (CV). The performances of the DREAM algorithm were compared with a Nelder-Mead Simplex algorithm adapted to the STICS model under the OptimiSTICS package.

This study showed that it was possible to successfully use the DREAM algorithm combined to a complex crop model such as STICS. The DREAM algorithm offers the advantage of Bayesian techniques and MCMC simulations, i.e. the approximation of the parameters' posterior distribution, the evaluation of correlations, and the uncertainties estimation in the output predictions.

The parameters' sampling using the SLS likelihood function within the DREAM algorithm showed close results to those obtained using OptimiSTICS. The model evaluation criteria, RMSE, EF, and ND were substantially improved compared with the initial set of parameter values. The residual analysis also showed the validity of the SLS approach of DREAM. These results were very satisfactory and encouraging. However, when using the SLS likelihood function, considering the temporal evolution of the simulated biomass during a crop cycle characterised by significant water deficit that occurred at the early season, it appeared that the simulations did not fully respect the biophysical behaviour of plant growth which compensated later the biomass growth thanks to higher efficiency of radiation use. Insignificant or unrealistic values occurred thus for some parameters. This was probably due to the fact that too limited information was included in the dataset to efficiently sample the posterior distribution of the selected nine parameters. Although parameters with low interaction were chosen, correlation appeared in the parameters' posterior distribution. This observation highlighted the importance of adapting the experimental design to the plant-soil system

777 dynamics with modeling purposes.

778 Finally, it is worth mentioning that the proposed likelihood function based on an explicit
779 formulation of the CV presented several advantages. The results were very close to those obtained
780 with the standard WLS likelihood function and were satisfactory in terms of evaluation criteria,
781 RMSE, EF and ND. From a biophysical point of view, relevant values for all parameters were
782 obtained. Furthermore, the proposed CV likelihood function allows taking into account not only the
783 noise on measurements but also the heteroscedasticity regularly encountered in crop modeling.

Acknowledgements

The authors would like to thank the SPW (DGARNE) for its financial support for the project entitled '*Suivi en temps réel de l'environnement d'une parcelle agricole par un réseau de microcapteurs en vue d'optimiser l'apport en engrais azotés*'.

They would also like to thank the OptimiSTICS team that allowed them to use the Matlab running code of the STICS model.

And finally they wish to thank Jasper Vrugt for his precious advices that greatly improved the quality of the present paper and for the interesting discussions about the DREAM and DREAM-ZS algorithm options.

We also wish to thank the two anonymous reviewers for their helpful comments.

5. References

Basso, B., Ritchie, J.T., 2005. Impact of animal manure, compost and inorganic fertilizer on nitrate leaching and yield in a six-year maize alfalfa rotation. Agri. Ecosys. Environ. 108, 329-341.

Basso B., Cammarano D., Troccoli A, Chen D., Ritchie J.T., 2010. Long-term wheat response to nitrogen in a rainfed Mediterranean environment: Field data and simulation analysis. Eur. J. Agron., 33, 132-138

Beaudoin N., Launay M., Sauboua E., Ponsardin G., Mary B., 2008. Evaluation of the soil crop model STICS over 8 years against the 'on farm' database of Bruyères catchment. Eur. J. Agron., 29, 46-57.

Bechini, L., Bocchi, S., Maggiore, T., Confalonieri, R., 2006. Parameterization of a crop growth and development simulation model at sub model component level. An example for winter wheat (*Triticum aestivum* L.). Environ. Model. Softw. 21, 1042-1054.

Beven, K., 1989. Changing ideas in hydrology. The case of physically based models. J. Hydrol. 105, 157-172.

Beven, K., 2008. Comment on ''Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling?'' by Vrugt JA, ter Braak CJF, Gupta HV, Robinson BA (2008). Stoch. Environ. Res. Risk Assess.

Beven, K., Binley, A., 1992. The future of distributed models: Model calibration and uncertainty prediction. Hydrol. Process. 6, 279-298.

Beven, K., Smith, P., Freer, J., 2008. So just why would a modeller choose to be incoherent? J. Hydrol., 354, 15-32.

Box, G., Cox, D., 1964. An analysis of transformations. Journal of the Royal Statistical Society, Series B 26(2), 211-252.

Box, G., Tiao, G., 1973. Bayesian inference in statistical analysis. Addison-Wesley, Reading, PA.

Brisson, N., Mary, B., Ripoche, D., Jeuffroy, M.-H., Ruget, F., Nicoullaud, B., Gate, P., Devienne-Barret, F., Antonioletti, R., Durr, C., Richard, G., Beaudoin, N., Recous, S., Tayot, X., Plenet, D., Cellier, P., Machet, J.-M., 1998. STICS: a generic model for the simulation of crops and their water and nitrogen balances. Theory and parameterization applied to wheat and corn. Agron. 18, 311-346.

Brisson N., Ruget F., Gate P., Lorgeau J., Nicoulaud B., Tayo X., Plenet D., Jeuffroy M.H., Bouthier A., Ripoche D., Mary B., Justes E., 2002. STICS: a generic model for simulating crops and their water and nitrogen balances. II. Model validation for wheat and maize. Agron., 22, 69-82.

Brisson, N., Gara, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussière, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillère, J.P., Hénault, C., Maraux, F., Seguin, B., Sinoquet, H., 2003. An overview of the crop model STICS. Eur. J. Agron. 18, 309-332.

Brisson, N., Launay, M., Mary, B., Beaudoin, N., 2009. Conceptual basis, formalisations and parameterization of the STICS crop model. Editions Quae, Collection Update Sciences and Technologies, p. 297.

832 Brooks, S.P., 1998. Markov chain Monte Carlo and its application, The Statistician. J. Royal Statist.
833 Soc. Series D.

834 Campolongo, F., Cariboni, J., Saltelli, A., 2007. An effective screening design for sensitivity analysis
835 of large models. Environ. Model. Softw. 22, 1509-1518.

836 Confalonieri, R., Bechini, L., 2004. A preliminary evaluation of the simulation model CropSyst for
837 alfalfa. Eur. J. Agron. 21, 223-237.

838 Dietzel, A., Reichert, P., 2012. Calibration of computationally demanding and structurally uncertain
839 models with an application to a lake water quality model. Environ. Model. Softw. 38(0) 129-146.

840 Flenet, F., Villon, P., Ruget, F., 2004. Methodology of adaptation of the STICS model to a new crop:
841 spring linseed (*Linum usitatissimum*, L.). Agron. 24, 367-381.

842 Gelman, A., Carlin, J.B., Stren, H.S., Rubin, D.B., 1997. Bayesian Data Analysis. Chapman and Hall,
843 London, p. 526.

844 Gilks, W.R., Richardson, S., Spiegehalter, D.J., 1996. Introducing Markov Chain Monte Carlo, in
845 Markov Chain Monte Carlo in Practice: Chapman & Hall, London, pp. 1-19.

846 Guillaume, S., Bergez, J.E., Wallach, D., Justes, E, 2011. Methodological comparison of calibration
847 procedures for durum wheat parameters in the STICS model. Eur. J. Agron. 35, 115-126

848 Hansen, S., Jensen, H., Nielsen, N., Swenden, H., 1990. DAISY – Soil Plant Atmosphere System
849 Model S NP0 Research, in the NAEP report, The Royal Veterinary and Agricultural University, Nr
850 A10.

851 Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications.
852 Biometrika 57, 97-109.

853 Jansen, M., Hagenaars, T., 2004. Calibration in a Bayesian modelling framework, In: van Boekel,
854 M.A.J.S., Stein, A., van Bruggen, A.H.C. (Ed.), Bayesian Statistics and Quality Modelling in the Agro-
855 food Production Chain: Kluwer Academic, Dordrecht.

856 Jeremiah, E., Sisson, S.A., Sharma, A., Marshall, L., 2012. Efficient hydrological model parameter
857 optimization with Sequential Monte Carlo sampling. Environ. Model. Softw. 38(0) 283-295.

858 Joseph, J.F., Guillaume, J.H.A., 2013. Using a parallelized MCMC algorithm in R to identify
859 appropriate likelihood functions for SWAT. Environ. Model. Softw. 46(0) 292-298.

860 Laloy, E., Fasbender, D., Bielders, C.L., 2010. Parameter optimization and uncertainty analysis for
861 plot-scale continuous modeling of runoff using a formal Bayesian approach. J. Hydrol. 380, 82-93.

862 Laloy, E., Vrugt, J.A., High-dimensional posterior exploration of hydrologic models using multiple-try
863 DREAM_(ZS) and high-performance computing, Water Resources Research, 48, W01526,
864 doi:10.1029/2011WR010608, 2012.

865 Lamboni, M., Makowski, D., Lehuger, S., Gabrielle, B., Monod, H., 2009. Multivariate global
866 sensitivity analysis for dynamic crop models. Field Crop. Res. 113, 312–320.

867 Launay, M., Flenet, F., Ruget, F., Garcia de Cortazar Atauri, I., 2005. Généricité et méthodologie
868 d'adaptation de STICS à de nouvelles cultures. Séminaire STICS, Carry-le-Rouet, 55-57 p.

869 Loague, K., Green, R.E., 1991. Statistical and graphical method for evaluating solute transport models:

overview and application. J. Contam. Hydrol. 7, 51-73.

Makowski, D., Wallach, D., Tremblay, M., 2002. Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods. Agron. 22, 191-203.

Makowski, D., Naud, C., Jeuffroy, M.H., Barbottin, A., Monod, H., 2006. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model prediction. Reliab. Eng. Sys. Saf. 91(10), 1142-1147.

Mansouri, M., Dumont, B., Destain, M.-F., 2013. Modeling and prediction of nonlinear environmental system using Bayesian methods. Computers and Electronics in Agriculture 92(0) 16-31.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Physi. 21, 1087-1092.

Minunno, F., van Oijen, M., Cameron, D.R., Cerasoli, S., Pereira, J.S., Tomé, M., 2013. Using a Bayesian framework and global sensitivity analysis to identify strengths and weaknesses of two process-based models differing in representation of autotrophic respiration. Environ. Model. Softw. 42(0) 99-115.

Monod, H., Naud, C., Makowski, D., 2006. Uncertainty and sensitivity analysis for crop models. *In:* Wallach, D., Makowski, D., Jones, J. W. (eds.) *Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications*, Chapter 4. Elsevier, pp. 55-100.

Monteith, J., 1996. The quest for balance in crop modeling. Agron. J. 88, 695-697.

Preece, D., 1990. R.A. Fisher and experimental design : a review. Biometrics, 46(4), 925-935.

Rasmussen, R., Hamilton, G., 2012. An approximate Bayesian computation approach for estimating parameters of complex environmental processes in a cellular automata Environ. Model. Softw. 29(1) 1-10.

Ritchie, J.T., Otter, S., 1984. Description and performance of CERES-Wheat, a user-oriented wheat yield model. USDA-ARS-SR Grassland Soil and Water Research Laboratory Temple RX, 159-175.

Ruget F., Brisson N., Delécolle R., Faivre R., 2002. Sensitivity analysis of a crop simulation (STICS) in order to determine accuracy needed for parameters. Agron., 22:133-158.

Schoups, G. and Vrugt, J., 2010. A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. Water Ressour. Res., 46, W10531, doi:10.1029/2009WR008933

Singh, A., Tripathy, R., Chopra, U., 2008. Evaluation of CERES-Wheat and CropSyst models for water−nitrogen interactions in wheat crop. Agric. Water Manag. 95, 776-786.

Ter Braak, C., 2006. A Markov chain Monte Carlo version of the genetic algorithm differential evolution: Easy Bayesian computing for real parameter spaces. Stat. Comput. 16, 239-249.

Tremblay, M., Wallach, D., 2004. Comparison of parameter estimation methods for crop models. Agron. 24, 351-365.

Van Diepen, C., Wolf, J., Van Keulen, H., Rappoldt, C., 1989. WOFOST: a simulation model of crop production. Soil Use Manag. 5, 16-24.

Varella, H., 2011. Inversion d’un modèle de culture pour estimer spatialement les propriétés des sols et

908 améliorer la prédiction de variables agro-environnementales. PhD thesis. 207 pp, Univ of Avignon.

909 Varella, H., Guérif, M., Buis, S., 2010a. Global sensitivity analysis measures the quality of parameter
910 estimation: the case of soil parameters and a crop model. Environ. Model. Softw. 25, 310–319.

911 Varella, H., Guérif, M., Buis, S., Beaudoin, N., 2010b. Soil properties estimation by inversion of a
912 crop model and observations on crops improves the prediction of agro-environmental variables. Eur. J.
913 Agron. 33(2), 139-147.

914 Vrugt, J., Gupta, V., Bouten, W., Sorooshian, S., 2003. A shuffled Complex Evolution Metropolis
915 algorithm for optimisation and uncertainty assessment of hydrologic model parameters. Water
916 Ressour. Res. 39(8), 1201.

917 Vrugt, J.A., Gupta, H.V., Nualláin, B., Bouten, W., 2006. Real-Time Data Assimilation for Operational
918 Ensemble Streamflow Forecasting. J. Hydrometeor. 7(3) 548-565.

919 Vrugt, J., ter Braak, C., Clark, M., Hyman, J., Robinson, B., 2008a. Treatment of input uncertainty in
920 hydrologic modeling using adaptive Markov Chain Monte Carlo sampling. Water Ressour. Res. 44.

921 Vrugt, J., ter Braak, C., Gupta, H., Robinson, B., 2008b. Response to comment by Keith Beven on
922 'Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic
923 modeling?' Stoch. Environ. Res. Risk Assess., 23, 7p.

924 Vrugt, J., ter Braak, C., Diks, C., Robinson, B., Hyman, J., Higdon, D., 2009a. Accelerating Markov
925 chain Monte Carlo simulation by self adaptive differential evolution with randomized subspace
926 sampling. Int. J. Nonlinear Sci. Numer. Simul. 10, 271–288.

927 Vrugt, J., ter Braak, C., Gupta, H., Robinson, B., 2009b. Equifinality of formal (DREAM) and
928 informal (GLUE) Bayesian approaches in hydrologic modeling. Stoch. Environ. Res. Risk Assess., 23,
929 7 p.

930 Vrugt, J.A., Laloy, E., and ter Braak, C.J.F., DiffeRential Evolution Adaptive Metropolis with
931 Sampling from past states, SIAM journal on Optimization, In Review, 2011

932 Wallach, D., Goffinet, B., Bergez, J.-E., Debaeke, P., Leenhardt, D., Aubertot, J.-N., 2001. Parameter
933 estimation for crop models: A new approach and application to a corn model. Agron. J. 93(4), 757-
934 766.

935 Wallach, D., Buis, S., Lechapentier, P., Bourges, J., Clastre, P., Launay, M., Bergez, J.E., Gueriff, M.,
936 Soudais, J., Justes, E., 2011. A package of parameter estimation methods and implementation for the
937 STICS crop-soil model. Environ. Model. Softw. 26, 386-394.

938 Wallach, D., Makowski, D., Jones, J., 2006. Working with Dynamic Crop Models: Evaluation,
939 Analysis, Parameterization and Applications. Wallach D., Makowski D., Jones J., Elsevier, Amsterdam
940 (Ed).

941 Wallach, D., Van Evert, F., Adam, M., 2009. Parameter estimation software for crop models.
942 Proceedings of the AgSAP Conference 2009: Egmond aan Zee, The Netherlands, pp. 342-343.

943 Williams, J.R., Jones, C.A., Kiniry, J.R., Spanel, D.A., 1989. The EPIC crop growth model. Trans.
944 ASAE 32, 497-511.

945 Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of

946 hydraulic properties of European soils. Geoderma 90, 169–185.

947 Wu, Y., Liu, S., 2012. Automating calibration, sensitivity and uncertainty analysis of complex models
948 using the R package Flexible Modeling Environment (FME): SWAT as an example. Environ. Model.
949 Softw. 31(0) 99-109.

950