Comparison of parameter estimation methods for crop models

Marie Tremblay, Daniel Wallach

Agronomie 24 (2004) 351–365
© INRA, EDP Sciences, 2004
DOI: 10.1051/agro:2004033

Original article

Comparison of parameter estimation methods for crop models

Marie TREMBLAY*, Daniel WALLACH

Unité Mixte de Recherche A.R.C.H.E., INRA Toulouse, BP 27, 31326 Castanet-Tolosan Cedex, France

(Received 3 June 2003; accepted 10 May 2004)

Abstract – Crop models are important tools in agronomic research, a major use being to make predictions. A proper parameter estimation method is necessary to ensure accurate predictions. Until now studies have focused on the application of a particular estimation method and few comparisons of different methods are available. In this paper, we compare several parameter estimation methods, related, on the one hand, to model selection, and on the other, to ridge regression based on an analogy to a Bayesian approach. The different methods are applied to a simplified crop model derived from the STICS model, using simulated data. The criteria for comparison are prediction error and errors in the parameter estimates. Among the methods of model comparison a version of the Schwarz criterion, corrected for small samples and with maximum and minimum bounds for each parameter, is the preferred method. Ridge regression is found to be superior to this best method of model selection.

parameter estimation / crop model / model selection methods / ridge regression

1. INTRODUCTION

Crop models are important tools in agronomic research and are increasingly being considered as potential components of decision support systems for agriculture. The two fundamental components of crop models are the model equations and the values of the model parameters. Both are essential in determining the usefulness of a model, and in particular the quality of model predictions.

The problem of parameter estimation has been very widely studied in statistics. However, the application of the proposed methods to crop models is often not immediate because crop models, like complex dynamic models in other fields such as hydrology, present several characteristics which make parameter estimation difficult. First of all, these models often contain a large number of parameters (for example, approximately 200 parameters in the STICS model [6]), compared with the amount of data available for parameter estimation. It is thus often impossible to estimate all the model parameters from the data, and even in cases where it is numerically possible this may lead to overparameterization and consequent poor predictive quality. In addition, prior information about the parameter values is often available, from experiments on the individual processes or from results that concern the system under conditions not too different than those of interest. There is then the question of how to use this information in the estimation procedure. Finally, the structure of the data is often complex. In a given plot several different variables (for example, grain yield, biomass and leaf area index) may be measured, and certain of these variables may be measured at several different dates during the season. Thus the data available are not independent and there is the question of how to take into account the complex correlation structure of the data.

We will discuss below different aspects of the problem of parameter estimation for complex dynamic models that have been reported in the literature. We will see that while there have been numerous studies on how to calculate parameter values, there are very few studies which systematically compare a range of different statistical approaches to parameter estimation. It seems important to propose such a study, in order to draw conclusions on the advantages and drawbacks of each method. The purpose of this paper then is to propose several methods for parameter estimation that could be applicable to crop models, and to test them in a simulation study on a relatively simple example.

Many of the studies reported in the literature focus on the problem of how many parameters to estimate and how to choose those to be estimated. Olsthoorn [24] suggested that the number of estimated parameters must be substantially fewer than the number of observations to obtain a valid optimum. Other studies have emphasized the importance of limiting the number of estimated parameters lest prediction error be high (for example, Refsgaard [28]). Perrin et al. [26] compared 19 hydrological models of different complexity (between 3 and 9 parameters) in order to see if the increase in the number of parameters improved the performances of a model. They concluded that models of small dimensions can have predictive capacities as good as more complex models.

One can of course reduce the number of estimated parameters by simply deciding that certain parameters will not be fit

* Corresponding author: m-tremblay@chu-montpellier.fr

to the data but will have fixed values. The problem is then to decide which parameters will be estimated and which will be fixed. Sievänen and Burk [31] fixed those parameters of their tree growth model which are easily measured and thus are available without fitting. Parameter values may also be fixed on the basis of values in the literature [34]. A sensitivity analysis of model outputs to different parameters may be the basis for estimating only those parameters to which the outputs are most sensitive. Yan and Haan [38] and Xevi et al. [37] reduced the number of parameters based on the sensitivity for each parameter individually. In their sensitivity analyses, Van der Perk [33] and Brun et al. [7] also took into account parameter interactions. Finally, the data itself may be used to determine how many and which parameters to estimate. This is the approach adopted by Wallach et al. [35], who first classified the parameters using a forward regression analysis and then decided how many parameters to estimate using cross-validation.

As underlined by Brun et al. [7], fixing some parameters may introduce a bias in the estimates of other parameters. Furthermore, a modification of the fixed values may lead to different estimated values because of the interactions between the parameters, as shown by Zhang and Lindström [39].

In linear regression, principal component regression or partial least squares regression may be used to reduce the number of parameters to estimate. Wallach et al. [36] applied continuum regression, which is a generalization of both those methods, to a crop model.

The approaches mentioned above make use of prior information about the parameters in a particular way. Either a parameter is fixed at its prior value, or the prior value is ignored and the parameter is estimated from the data. A Bayesian approach, on the other hand, treats the parameters as random variables. The prior information furnishes a prior distribution for the parameters, and then this is combined with the data to calculate a posterior distribution. Increased computing power and specific numerical methods have made it possible to calculate the posterior distribution even for fairly complex dynamic models. The GLUE method (generalized likelihood uncertainty estimation) [4] assigns a weight proportional to the likelihood to each parameter set sampled from the prior distribution. This method has been applied to distributed hydrological models containing from 4 to 6 parameters [9, 10]. Schulz et al. [29] also applied the GLUE method to a nitrogen budget model including 7 parameters. With the Metropolis Hastings algorithm [15, 23], one can generate a sample from the posterior parameter distribution. This algorithm has been used to estimate the parameters of dynamic models [2, 8, 13]. Kuczera and Parent [19] and Makowski et al. [20] compared GLUE and MCMC on models containing, respectively, 4 and 21 parameters. The Metropolis Hastings algorithm was preferred in both studies. Omlin and Reichert [25] compared frequentist and Bayesian techniques for estimating prediction uncertainty in a simple case (2 parameters). In the case of poor parameter identifiability, the Bayesian method was preferred.

We consider two families of approaches to parameter estimation which seem to be reasonable candidates for crop models. The first family, model selection methods, indicates how to identify a small number of parameters to be estimated. The other parameters are fixed at a fixed value. The second approach is the Bayesian approach, which combines the prior information about the parameters and the information contained in the data to estimate all the parameters. We do not calculate the posterior distribution but rather just the mode of this distribution, calculated using ridge regression. This provides estimates of the parameter values that can be compared with estimates given by the first family of methods.

To compare the different statistical methods we did a simulation study. That is, we invented "true" parameter values and data, and then tested each method to see how well it could retrieve these "true" parameter values given the data. The advantage of working by simulation rather than on real data is that the true parameter values are known. It is thus possible to calculate the quality of the estimated parameters and the predictive quality of the adjusted model for each method. The drawback is that the generality of the results is hard to know. The results may depend on the details of the model, on the way the data are generated and on the specific data that are used.

The different variables and acronyms used throughout the text are shown in Table I.

2. MATERIALS AND METHODS

2.1. Model

The model for which we tested different parameter estimation methods is a part of the STICS model [6], which we shall refer to as Mini-STICS. The use of only part of the STICS model was dictated by the need to limit computing time, since we were led to execute the model several million times.

Mini-STICS simulates sunflower development over a period of 20 days, starting at the stage Maximal Acceleration of Leaf growth (AMF). We suppose that there is no nitrogen stress and that the soil consists of two layers: the first layer extends from 0 to 30 cm and the second from 30 cm to the bottom of the soil.

The five state variables are shown in Table II. The vector of explanatory variables has 75 elements and includes soil characteristics (percentage of clay and bulk density of the first soil layer, and minimum and maximum soil available water of each layer and soil depth), daily climatic variables (minimum and maximum temperature, evapotranspiration and precipitation), management operations (sowing density and depth of ploughing) and initial conditions (leaf area index, root depth and water content of each layer). The parameter vector has 14 elements. The "true" parameter vector, $\theta_{(T)}$, is based on the values suggested by Guiloufi [12] (Tab. III).

The model equations are presented in Appendix. These are the dynamic equations that indicate how each state variable evolves from one day to the next as a function of the current values of the state variables and of the explanatory variables. Integrating these equations over time (i.e. "running" the model), allows one to eliminate the intermediate values of the state variables and relate the state variables at any time to the explanatory variables on each day up to that time. Then for any given response (a response is a state variable or a function of state variables at a given time), the prediction by Mini-STICS can simply be noted $f_j(X;\theta)$ where the index j identifies the

Table I. Variables and acronyms used in the text.

Acronym	Explanation
AIC	Akaike Information Criterion
AIC_C	Akaike Information Criterion corrected for small samples
BIC	Bayesian Information Criterion
BIC_C	Bayesian Information Criterion for small samples
${BIC_C}^B$	${BIC_C}^B$ with constraints
CV	Cross-validation
MSEP	Mean Squared Error of Prediction
MSE	Mean Squared Error
AMF	Maximal acceleration of leaf growth
LAI	leaf area index
HUR	volumetric water content in full soil, HUR_1+HUR_2

Variable	Explanation
q	parameter vector of Mini-STICS (length 14)
$\hat{\theta}$	Estimator of θ
$\theta_{(T)}$	True value of θ
μ_k	mean of the prior distribution for θ (or initial value of θ)
ω^2_k	variance of the prior distribution for θ
n	number of individuals in the training sample
X	Vector of explanatory variables of the Mini-STICS model
$D_{X_d}(x)$	probability distribution of X, where d represents the distribution number
$i(d)$	index of an individual with explanatory variables drawn from $D_{X_d}(x)$
$X_{i(d)}$	vector of explanatory variables drawn from $D_{X_d}(x)$ for the individual $i(d)$
$y_{i(d),j}$	observed response j, j = 1, …, 4, of the individual $i(d)$
$y_{i(d)}$	vector of observed responses (length 4)
$f_j(X_{i(d)};\theta)$	model prediction of response j for individual $i(d)$
$f(X_{i(d)};\theta)$	vector of model predictions for individual $i(d)$ (length 4)
$\varepsilon_{i(d),j}$	model error for variable j individual $i(d)$
$\varepsilon_{i(d)}$	vector of model errors for individual $i(d)$ (length 4)
Σ	covariance matrix of $\varepsilon_{i(d),}$ of size 4*4.
$\hat{\Sigma}$	estimate of Σ
σ^2_j	variance of $\varepsilon_{i(d),j}$
$\hat{\sigma}^2$	estimate of σ^2_j
$MSEP_j^d(m)$	mean squared error of prediction for response j and individual drawn from $D_{X_d}(x)$, obtained with parameter estimation method m
C_{OLS}	ordinary least squares criterion
C_{GLS}	generalized least squares criterion

response, X is the vector of explanatory variables and θ is the parameter vector. We will consider in particular the four responses shown in Table IV.

2.2. Prior information

We assume the existence of prior information, that is, of information about the parameter values independent of the data

Table II. State variables of the Mini-STICS model.

Name	Meaning	Units
LAI	Leaf area index	m^2 leaves·m^{-2} soil
HUR_1	Volumetric water content of the layer 1	$mm.cm^{-1}$
HUR_2	Volumetric water content of the layer 2	$mm.cm^{-1}$
SUDEVCULT	Cumulative effective temperature for plant development	°C
ZRAC	Root depth	cm

Table III. Parameters of the Mini-STICS model and true parameter values $\theta_{(T)}$.

Parameter	Meaning	Units	True value
ADENS	Parameter of compensation between stem number and plant density	–	–0.8
BDENS	Maximum density above which there is competition between plants	$plants.m^{-2}$	1.25
CROIRAC	Growth rate of the root front	cm. $degree\text{-}day^{-1}$	0.25
DLAIMAX	Maximum rate of the setting up of LAI	m^2 leaves.soil m^{-2}. $degree\text{-}days^{-1}$	0.0078
EXTIN	Extinction coefficient of photosynthetic active radiation in the canopy	–	0.9
KMAX	Maximum crop coefficient for water requirements	–	1.2
LVOPT	Optimum root density	cm root.cm^{-3} soil	0.5
PSISTO	Absolute value of the potential of stomatal closing	bars	10
PSISTURG	Absolute value of the potential of the beginning of decrease in the cellular extension	bars	4
RAYON	Average radius of roots	cm	0.02
TCMIN	Minimum temperature of growth	°C	6
TCOPT	Optimum temperature of growth	°C	32
ZPENTE	Depth where the root density is ½ of the surface root density for the reference profile	cm	120
ZPRLIM	Maximum depth of the root profile for the reference profile	cm	150

Table IV. Response variables $y_{i(d),j}$, j = *1, …, 4,* available for parameter estimation and model evaluation.

Index j	Response variables	Notation
1	LAI, day 20	$y_{i(d),1}$
2	LAI, day 10	$y_{i(d),2}$
3	HUR_1+HUR_2, day 10	$y_{i(d),3}$
4	HUR_1+HUR_2, day 20	$y_{i(d),4}$

Table V. Mean (μ_k) and standard deviation (ω_k), *k=1, ..., 14,* of the two prior distributions for the parameters.

Parameter	Prior distribution 1		Prior distribution 2	
	μ_k	ω_k	μ_k	ω_k
ADENS	–0.6960	0.1206	–0.9048	0.1567
BDENS	1.1029	0.1910	1.4338	0.2483
CROIRAC	0.2913	0.0505	0.3787	0.0656
DLAIMAX	0.0061	0.0011	0.0079	0.0014
EXTIN	0.6396	0.1108	0.8315	0.1440
KMAX	1.4101	0.2442	1.8331	0.3175
LVOPT	0.5672	0.0982	0.7374	0.1277
PSISTO	12.29	2.13	15.98	2.77
PSISTURG	3.79	0.6564	4.9300	0.8539
RAYON	0.0167	0.0029	0.0217	0.0038
TCMIN	7.10	1.23	9.20	1.59
TCOPT	32.10	5.56	41.70	7.22
ZPENTE	113.10	19.59	147.00	25.46
ZPRLIM	154.90	26.83	201.40	34.88

that is used for parameter adjustment. We represent this information as a probability distribution. For each parameter θ_k the prior distribution is assumed to be normal, $\pi_{\theta_k}(\theta_k) = \mathrm{N}(\mu_k, \omega_k^2)$, where μ_k and ω_k^2 are, respectively, the prior mean and the prior variance of the parameter k, *k=1,…,14*. We assume the parameters to be a priori independent. Thus the prior distribution for the full parameter vector is

$$\pi_\theta(\theta) = \prod_{k=1}^{14} \mathrm{N}(\mu_k, \omega_k^2). \quad (1)$$

We considered two different prior distributions. The means and standard deviations are presented in Table V. For each prior distribution, first the prior means were chosen at random within limits proposed by agronomists around the true values. Then the prior variance for each parameter was calculated as $\omega_k^2 = (0.6 \times \mu_k)^2/12$, which is equal to the variance of a uniform distribution with lower limit $0.7 \times \mu_k$ and upper limit $1.3 \times \mu_k$. Note that the prior means are not equal to the "true" parameter values.

2.3. Simulations

We consider that the explanatory variables of Mini-STICS for different individuals (i.e. different site, year and management combinations) are drawn from a probability distribution that represents a range of conditions. We consider 6 different distributions (Tab. VI). The distribution $D_{X_0}(x)$ corresponds approximately to the range of conditions in the area around the INRA experimental farm of the Toulouse research center. For example, (Tab. VI), this distribution has a 0.25 probability of rain each day, and if it rains the amount of precipitation has a uniform distribution with a minimum of 1 mm and a maximum of 15 mm. All the other explanatory variables also have uniform distributions. The explanatory variables are independent of one another and also independent from one day to another. The distributions $D_{X_1}(x)$, $D_{X_2}(x)$, $D_{X_3}(x)$ and $D_{X_4}(x)$ represent, respectively, more shallow soils, greater maximum available soil water, lower sowing densities and less rainfall than $D_{X_0}(x)$. $D_{X_5}(x)$ combines all of these differences.

The "true" response j for an individual *i(d)* drawn from $D_{X_d}(x)$ is generated as

$$y_{i(d),j} = f_j(X_{i(d)}; \theta) + \varepsilon_{i(d),j} \quad (2)$$

where $X_{i(d)}$ is a vector of explanatory variables drawn from $D_{X_d}(x)$ and $\varepsilon_{i(d),j}$ is the model error term. This is a random variable that describes, in terms of probabilities, that part of the response which the model does not explain. We assume that the vector of model errors $\varepsilon_{i(d)} = (\varepsilon_{i(d),1}, \varepsilon_{i(d),2}, \varepsilon_{i(d),3}, \varepsilon_{i(d),4})'$ is distributed according to $N(0,\Sigma)$, where Σ is the (4×4) variance covariance matrix. The hypothesis that the error has expectation 0 and the same variance-covariance matrix regardless of the values of the explanatory variables is a strong hypothesis. We note $y_{i(d)}$ and $f(X_{i(d)}; \theta)$ the vectors of observed and calculated responses, respectively, for an individual *i(d)*.

We generated two sorts of samples using equation (2). First, we generated training samples, which contain the responses that are used for parameter adjustment. All individuals here have explanatory variables drawn from $D_{X_0}(x)$. The responses for different individuals in a training sample are independent.

Secondly, we generated test samples, used for evaluating the predictive quality of a model. The test samples each have 1000 individuals. We generated test samples for each of the 6 distributions of the explanatory variables. The responses in the test samples are independent of one another and of the responses in the training samples.

We did two simulation studies. In the first, we compared the different model selection methods detailed below. The comparison between methods was done for three different sizes of the training sample, namely $n = 7$, 14 or 28 and for two different hypotheses about the prior distribution (Tab. V). The model selection methods make use only of the expectation of the prior distribution, this value being used as the initial value for each parameter. In each case the training sample contained only a single response variable, the leaf area index at day 20, noted $y_{i(d),1}$ for individual *i(d)*, $i = 1,\ldots n$. Parameter estimation was repeated on 80 different training samples. Model predictive quality was evaluated by comparing predicted values with the values in the six test samples. Each test sample had 1000 values of leaf area index at time 20, generated using equation (2) with explanatory variables drawn from one of the six different distributions of explanatory variables described above. Thus the different test samples correspond to testing the model for different ranges of soils or climates.

In the second simulation study, we compared the best model selection method with ridge regression. Here we considered only a single size for the training sample, namely $n = 14$, but again the two different prior distributions for the parameters. Each training sample here had four responses for each individual (Tab. IV). Once again, each method was tested on 80 different training samples, and evaluated using six different test samples of 1000 individuals each, the test samples corresponding to different distributions of explanatory variables. For each individual in each test sample we generated four responses, the same as in the training sample. Thus in this study the training

Table VI. Probability distributions, $D_{X_d}(x)$, *d=0,...,5*, of explanatory variables *X*.

		$D_{X_0}(x)$	$D_{X_1}(x)$	$D_{X_2}(x)$	$D_{X_3}(x)$	$D_{X_4}(x)$	$D_{X5}(x)$
CLIMATE	Précipitations (mm) $P = (P_1 \ldots P_{20})'$ ✱ P_i i.i.d. ✦	$P_i \sim B(0.25)^{\Phi} * U^{\dagger}$ (1,15)	$B(0.25)^{\Phi} * U^{\dagger}$ (1,15)	$B(0.25)^{\Phi} * U^{\dagger}$ (1,15)	$B(0.25)^{\Phi} * U^{\dagger}$ (1,15)	0	0
	Minimum temperature (°C) $Tmin = (Tmin_1 \ldots Tmin_{20})'$ $Tmin_i$ i.i.d.	$Tmin_i \sim U(9.5,17.75)$	U(9.5,17.75)	U(9.5,17.75)	U(9.5,17.75)	U(9.5,17.75)	U(9.5,17.75)
	Maximum temperature (°C) $Tmax = (Tmax_1 \ldots Tmax_{20})'$ $Tmax_i$ i.i.d.	$Tmax_i \sim U(18,31)$	U(18,31)	U(18,31)	U(18,31)	U(18,31)	U(18,31)
	Evapotranspiration (mm) $ETP = (ETP_1 \ldots ETP_{20})'$ ETP_i i.i.d.	$ETP_i \sim U(1.5,6.75)$	U(1.5,6.75)	U(1.5,6.75)	U(1.5,6.75)	U(1.5,6.75)	U(1.5,6.75)
SOIL	Percentage of clay in he surface layer (%) ARGI	ARGI ~ U(30,60)	U(30,60)	U(30,60)	U(30,60)	U(30,60)	U(30,60)
	Bulk density of the first layer ($g \cdot cm^{-3}$) DA	DA ~ U(1,1.3)	U(1,1.3)	U(1,1.3)	U(1,1.3)	U(1,1.3)	U(1,1.3)
	Sowing density ($pl \cdot m^{-2}$) DENSITE	DENSITE ~ U(5,7)	U(5,7)	U(5,7)	U(3,5)	U(5,7)	U(3,5)
	Minimum volumetric water content ($mm \cdot cm^{-1}$) HUMIN‡	HUMIN ~ U(1.5,2)	U(1.5,2)	U(1,1.5)	U(1.5,2)	U(1.5,2)	U(1,1.5)
	Usable reserve ($mm \cdot cm^{-1}$) HUCC-HUMIN	HUCC - HUMIN ~ U(1.2,1.7)	U(1.2,1.7)	U(1.7,2)	U(1.2,1.7)	U(1.2,1.7)	U(1.7,2)
	Soil depth (cm) PROFSOL	PROFSOL ~ U(140,180)	U(40,80)	U(140,180)	U(140,180)	U(140,180)	U(40,80)

✱ denotes the transpose vector;
✦ i.i.d. independent and identically distributed;
$^{\Phi}$ B(0.25) Bernouilli distribution with probability 0.25;
† U uniforme distribution;
‡ Hypothesis: *HUMIN(1)* and *HUMIN(2)* are distributed according to the same distribution;
Hypothesis: *HUCC(1)-HUMIN(1)* and *HUCC(2)-HUMIN(2)* are distributed according to the same distribution;
Hypothesis: depth of the first layer is fixed (30 cm) and depth of the second layer is equal to PROFSOL-30.

samples allow us to test not only how prediction quality varies when soil and climate conditions vary, but also how prediction quality differs for different predicted responses. Table VII summarizes the different simulation studies.

2.4. Criteria for comparing methods

To choose between the various parameter estimation methods, comparison criteria are needed. Since crop models are used in multiple ways, it is natural to use several criteria.

The most important objective of crop models is prediction, which leads us to use the mean squared error of prediction (*MSEP*) as the basic criterion. However, since crop models are used for different types of prediction, we will define a number of different *MSEP* values. First of all, we will consider predictions for the 6 different distributions of the explanatory variables of Table VI. Furthermore, we will consider separately prediction for each of the 4 response variables. Overall then, there are 4 × 6 = 24 prediction quality criteria.

The definition of the mean squared error of prediction for response *j*, *j = 1,..., 4*, distribution $D_{X_d}(x)$, *d = 0,..., 5*, and parameter estimation method *m* is

Table VII. The two simulation studies.

Study	Methods compared	Situations considered
1	AIC, AIC_C, BIC, BIC_C, CV, $BIC_C{}^B$	7, 14 or 28 individuals per training sample × two prior distributions. A single response, 80 different training samples. ($BIC_C{}^B$ tested only for 14 individuals per training sample)
2	$BIC_C{}^B$, Ridge	14 individuals per training sample × two prior distributions. Four response variables, 80 different training samples.

$$MSEP_j^d(m) = \mathrm{E}\left[\left(y^*_{i(d),j} - f_j\left(X^*_{i(d)};\theta\right)\right)^2\right] \qquad (3)$$

where $\hat{\theta}(m)$ is the estimate of the parameter vector obtained with the method *m* and the star indicates that the random variables are independent of those in the training sample on which $\hat{\theta}(m)$ is based. The expectation is over all the random quantities, namely $\hat{\theta}(m)$ and $y^*_{i(d),j}$.

It is easy to see that if the parameter estimate is perfect, that is if $\hat{\theta}(m) = \theta_{(T)}$, then $MSEP_j^d(m) = Var\left[\left(\varepsilon_j\right)^2\right] = \sigma_j^2$. The minimum value of $MSEP_j^d(m)$ is thus σ_j^2, $j = 1, \ldots, 4$.

In the simulations, $MSEP_j^d(m)$ is estimated by

$$\hat{MSEP}_j^d(m) = \frac{1}{80}\sum_{l=1}^{80}\frac{1}{1000}\sum_{i=1}^{1000}\left[y^*_{i(d),j} - f_j\left(X^*_{i(d)};\hat{\theta}_l(m)\right)\right]^2. \quad (4)$$

The first sum, indexed by *l,* is over the 80 different training samples (this estimates the expectation over $\hat{\theta}(m)$) and the second, indexed by *i,* is over the individuals in the test sample.

The quality of the parameter estimates is also important. First of all, the quality of the parameter estimates determines model prediction. Our *MSEP* criteria measure prediction quality directly, but not for all possible situations. Furthermore, certain parameters have a physical or biological significance and so it is of interest to have estimates of their values. For each model parameter, the mean squared error is defined as

$$MSE_k(m) = E\left[\left(\theta_{(T)_k} - \hat{\theta}_k(m)\right)^2\right],\ k = 1\ldots14. \quad (5)$$

In the simulations, *MSE* is estimated by

$$\hat{MSE}_k(m) = \frac{1}{80}\sum_{l=1}^{80}\left(\theta_{(T)_k} - \hat{\theta}_{lk}(m)\right)^2,\ k = 1\ldots14, \quad (6)$$

where the sum, indexed by *l*, estimates the expectation over $\hat{\theta}(m)$.

2.5. Parameter estimation methods

2.5.1. Model selection methods

The model selection methods we consider are the Akaike Information Criterion, *AIC* [1], the Bayesian (or Schwarz) Information Criterion, *BIC* [30], versions of these two approaches corrected for small samples, AIC_C [17] and BIC_C [22] , and an approach where the best candidate model is chosen by cross-validation *CV* [32]. Another method, noted $BIC_C{}^B$, was also used in this study. In this method, the number of parameters to estimate is chosen by BIC_C, but maximum and minimum bounds are fixed for each parameter (see Tab. VIII).

These model selection methods have been widely studied in the linear regression framework. For a linear model, model selection corresponds to the choice of explanatory variables to be included in the model. The parameters associated with variables not included in the final model are set to zero. For crop models as for nonlinear models in general, parameters may occur in complex expressions and setting a parameter to zero does not simply correspond to ignoring one particular explanatory variable. A reasonable analog of model selection for nonlinear models is to select those parameters that will be fixed at their prior means rather than at zero.

2.5.1.1. Construction of candidate models

Since each parameter can either be estimated from the data or fixed at its initial value, there are 2^p different subsets of parameters that we could choose to estimate from the data. In our case that is $2^{14} = 16384$ subsets. To reduce the computational burden, we used a forward regression approach to generate a smaller set of candidate models. One first adjusts in turn each parameter of the model in order to determine the best single parameter to estimate. Then one determines the best second parameter to estimate in combination with the parameter already selected. If one goes all the way to the complete model, where all parameters are estimated, the total number of parameter combinations estimated is $p(p+1)/2$ or 105 in our case with $p = 14$. In general it is not necessary to continue until all parameters are estimated. The model selection methods are based on minimizing a criterion. The criterion is calculated for increasing numbers of estimated parameters, and as soon as the criterion is found to increase, one stops the procedure. Note that other methods than forward regression, such as stepwise regression, could also be used to reduce the number of adjustments.

In the case of a single response measured for each individual, for example, $j = 1$ (leaf area index at day 20), the parameters are adjusted to minimize the ordinary least squares criterion,

$$C_{OLS} = \sum_{i=1}^{n}\left(y_{i(d),1} - f_1\left(X_{i(d)};\theta\right)\right)^2. \quad (7)$$

where the sum is over the n individuals of the training sample.

For a multivariate regression model (several measurements for each individual), the least squares criterion (7) is replaced by the generalized least squares criterion

$$C_{GLS} = \sum_{i=1}^{n}\left(y_{i(d)} - f\left(X_{i(d)};\theta\right)\right)'\Sigma^{-1}\left(y_{i(d)} - f\left(X_{i(d)};\theta\right)\right). \quad (8)$$

The values of θ and Σ are estimated iteratively. From an initial estimation $\hat{\Sigma}_0$ (we use the identity matrix), an estimate $\hat{\theta}_1$ is obtained by minimizing C_{GLS} with regard to θ. Then the estimation of Σ is updated using

$$\hat{\Sigma}_1 = \frac{1}{n}\sum_{i=1}^{n}\left(y_{i(d)} - f\left(X_{i(d)};\theta\right)\right)\left(y_{i(d)} - f\left(X_{i(d)};\theta\right)\right)'. \quad (9)$$

This procedure is repeated until the difference between two successive estimates $\left(\hat{\theta},\hat{\Sigma}\right)$ is negligible. Gallant [11] showed that the estimates obtained with this iterative procedure are equivalent to the maximum likelihood estimates.

The routine DUNLSF of the library IMSL [18] was used to minimize the least squares criterion. This routine uses the Levenberg-Marquardt algorithm.

2.5.1.2. Penalized likelihood methods

The methods *AIC*, *BIC* and their derivatives are penalized likelihood methods. For example, in the case where the residual errors in the training sample are independent and identically distributed as $N(0,\sigma^2)$ (our case if there is only a single response), BIC_C is defined by

$$BIC_C = n\log\left(\hat{\sigma}^2\right) + \frac{k\log(n)n}{n-k-1}, \quad (10)$$

where $\hat{\sigma}^2$ is the maximum likelihood estimate of σ^2, n is the number of individuals in the training sample and k is the number of estimated parameters (including one parameter for the residual variance). The first term is –2 times the logarithm of the likelihood evaluated at the maximum likelihood estimates and the second term is the penalization. The final model chosen by the method BIC_C is the candidate model with the smallest value of

Table VIII. Constraints used in the BIC_C^B method.

Parameter	Lower limit	Upper limit
ADENS	–2.00	0.00
BDENS	1.00	20.0
CROIRAC	0.005	0.30
DLAIMAX	0.0001	0.50
EXTIN	0.10	1.50
KMAX	1.00	4.00
LVOPT	0.05	1.00
PSISTO	1.00	25.00
PSISTURG	1.00	15.0
RAYON	0.005	0.07
TCMIN	–10.0	15.0
TCOPT	10.0	35.0
ZPENTE	10.0	200.0
ZPRLIM	10.0	200.0

BIC_C. Other methods are similar but with different penalization factors.

Note that for a given number of parameters k, the penalization term is constant regardless of which parameters are estimated. Furthermore, $\hat{\sigma}^2$ is equal to $1/n \sum_{i=1}^{n} (y_{i(d),1} - f_1(X_{i(d)}, \hat{\theta}_{OLS}))^2$. The ordinary least squares estimate, $\hat{\theta}_{OLS}$, is equivalent to the maximum likelihood estimate, $\hat{\theta}_{ML}$, since we assume the residual errors in the training sample to be independent and identically distributed as $N(0, \sigma^2)$. Thus, for a given k, the model with the smallest value of C_{OLS} is best according to BIC_C.

The importance of the penalization term is that it serves to decide between models with different numbers of parameters. The penalization term divided by n increases as k increases and decreases as n increases (see Tab. IX).

In the case of multivariate regression, the method AIC_C corresponds to choosing the candidate model that minimizes

$$AIC_C = n \log\left(\left|\hat{\Sigma}\right|\right) + \frac{2n(kr + r(r+1)/2)}{n - (k + r + 1)}, \qquad (11)$$

[3], where $|\hat{\Sigma}|$ is the determinant of the maximum likelihood estimate of Σ and r is the number of responses measured for each individual. The first term of equation (11) is again equal to –2 times the logarithm of the likelihood estimated at the maximum likelihood estimates.

No bibliographical reference to the use of BIC_C in a multivariate regression setting was found. We use the expression for the univariate case but with n, the number of individuals replaced by $n \times r$. Thus

$$BIC_C = n \log\left(\left|\hat{\Sigma}\right|\right) + \frac{(k + r(r+1)/2)\log(nr)nr}{nr - k - r(r+1)/2}. \qquad (12)$$

This is in agreement with the procedure LME of SPLUS [27], where the total number of measurements $n \times r$ is also used, but McQuarrie and Tsaï [21] use simply n.

Table IX. The penalty term of the BIC_C criterion as a function of number of adjusted parameters and training sample size.

Number of adjusted parameters k	Size of the training sample n	k/n	Penalization†/n
2	14	2/14	0.48
4	14	4/14	1.17
8	14	8/14	4.22
2	28	1/14	0.27
4	28	2/14	0.58
8	28	4/14	1.40

† penalization term of the BIC_C method: $k \log(n) n/n - k - 1$.

2.5.1.3. Cross-validation

Cross-validationestimates the prediction error of each candidate model using

$$CV = \frac{1}{n} \sum_{i=1}^{n} \left(y_{i(d),1} - f_1\left(X_{i(d),1}; \hat{\theta}_{-i}\right)\right)^2, \qquad (13)$$

where $\hat{\theta}_{-i}$ arises from adjusting the same parameters as the candidate model, but based on the training data from which the individual i has been removed. CV is computationally expensive since for each candidate model the parameters are estimated n times.

This is "partial" cross-validation in the sense that first the candidate models are chosen using the full set of training data. The cross-validation is then used only to choose between the candidate models. In "true" (or complete) cross-validation, one would also derive the candidate models for each sample of size $(n–1)$ obtained by removing an observation from the original training sample. Breiman and Spector [5] showed that "partial" cross-validation, which we have done, leads to downward biased estimates of $MSEP$ and to more complex models than "true" cross-validation.

2.5.2. Ridge regression or mode of the posterior distribution

Ridge regression [16] can be presented as a method of penalized likelihood. The criterion to minimize is

$$S(\theta, \Sigma) = \sum_{i=1}^{n} \left(y_{i(d)} - f\left(X_{i(d)}; \theta\right)\right)' \Sigma^{-1} \left(y_{i(d)} - f\left(X_{i(d)}; \theta\right)\right) + \sum_{k=1}^{p} \lambda_k (\mu_k - \theta_k)^2, \qquad (14)$$

where λ_k and μ_k are, respectively, weight and initial value for the parameter k, $k = 1, \ldots, 14$.

With this approach, one estimates all the parameters simultaneously, by determining the values that minimize $S(\theta,\Sigma)$. The minimization can be done iteratively, as for the model selection methods.

It can easily be shown that with our assumptions about the prior distribution (normal distribution) and the model errors (normal distribution), then if $\lambda_k = 1/\omega_k^2$, $k = 1, \ldots, 14$, the ridge estimate is equal to the mode of the posterior distribution

Table X. Weights λ_k used in the ridge regression.

Parameter	Prior distribution 1	Prior distribution 2
ADENS	68.76	40.73
BDENS	27.41	16.22
CROIRAC	392.12	232.38
DLAIMAX	826446.3	510204.1
EXTIN	81.46	48.23
KMAX	16.77	9.92
LVOPT	103.70	61.32
PSISTO	0.2204	0.1303
PSISTURG	2.32	1.37
RAYON	118906.1	69252.1
TCMIN	0.6610	0.3956
TCOPT	0.0324	0.0192
ZPENTE	0.0026	0.0015
ZPRLIM	0.0014	0.0008

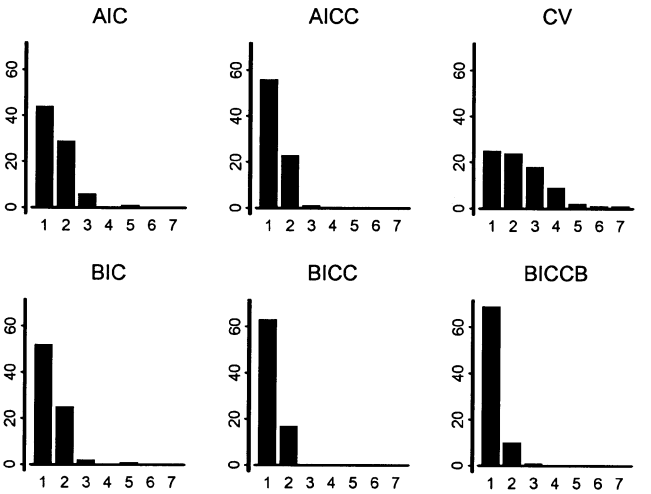

Figure 1. Histogram of number of parameters estimated by the model selection methods *AIC, AIC$_C$, BIC, BIC$_C$, BIC$_C^B$ and CV,* for $n = 14$ and the first vector of initial values. Total number of cases = 80.

$\pi(\theta|Y)$. The values of λ_k corresponding to each of our prior distributions are presented in Table X.

The parameter estimation was carried out with the routine DBCONF of the library IMSL of Fortran [18]. This routine uses a Quasi-Newton algorithm.

3. RESULTS AND DISCUSSION

3.1. Comparison of model selection methods

In the first simulation study we compared the different model selection methods presented in Section 3.5.1 using just the response variable 1 (*LAI* on day 20). We present the results for $n = 14$ and for the first initial value vector. The results for the other training sample sizes and the other vector of initial values were similar.

Figure 1 presents the number of parameters estimated using each method. The methods AIC_C, BIC_C and BIC_C^B are the most parsimonious. BIC_C selects between 1 and 2 parameters, depending on the training sample, and the other two methods between 1 and 3 parameters. *CV*, on the other hand, selects between 1 and 7 parameters.

Since we are using simulation studies it is possible to determine, for each training sample, the optimal number of parameters to estimate in order to minimize $MSEP_1^{d=0}$. For each candidate model, we use the test sample based on $D_{X_0}(x)$ to estimate the true value of $MSEP_1^{d=0}$. The optimal candidate model, and correspondingly the optimal number of parameters to estimate, is that which gives the smallest mean squared error of prediction. This number varies between 1 and 3 depending on the training sample. This is coherent with the recommendation of Harrel [14] that one should normally not estimate more than *n/20* or *n/10* parameters (in our case with $n = 14$ this translates into between 1 and 2 parameters). It can be seen that the penalized likelihood methods indeed very rarely estimate more than 3 parameters, whereas cross-validation often estimates too many parameters.

Table XI presents the values of mean squared error for each parameter and each method. Each value is an average over the 80 training samples. For those training samples where the parameter is not estimated but keeps its initial value, the mean squared error is that of the initial value. The first column of the table corresponds to the mean squared error for the initial value. According to this table, BIC_C^B has the smallest mean squared error among the different methods for every parameter. However, even this method has mean squared errors that are systematically larger than for the initial parameter values.

Table XII presents the values of $MSEP_1^d(m)$ for the various model selection methods. The first line of the table gives the *MSEP* values calculated using the initial parameters values. The last line gives the minimum value of *MSEP*, namely $\sigma_1^2 = 0.36$. The columns correspond to different distributions of the explanatory variables.

BIC_C^B gives the lowest mean squared error of prediction for every distribution of the explanatory variables. For the distribution $D_{X_0}(x)$, the reduction in prediction error compared with the error using the initial parameter values is significant for all methods. The initial value is 0.63, reduced to 0.44 for *AIC* and *CV*, and to 0.39 for BIC_C^B. The *MSEP* value for BIC_C^B is close to the minimum value.

The extrapolation populations 3 and 5 are of particular interest. For these populations, only BIC_C^B gives prediction errors which are lower or essentially the same as those obtained with the initial values. All the other methods give worse predictions. For example, the value of $MSEP_1^{d=3}$ for the initial values is 0.65, but is 0.87 for BIC_C and 1.31 for *AIC*.

The fact that BIC_C^B has larger mean squared errors than the initial parameter values, but predicts better, is noteworthy. This is due to the fact that the estimated values of the 1 or 2 parameters estimated by BIC_C^B compensate for the errors in the initial values in the other parameters. This compensation leads to

Table XI. Mean Squared Error (MSE) of parameter estimate $\hat{\theta}$ obtained with each model selection method, *n = 14* and first vector of initial parameter values.

Parameter	Initial	AIC	AIC_C	BIC	BIC_C	CV	BIC_C^B
ADENS	0.0108	0.6290	0.4558	0.5229	0.3782	0.5482	0.0934
BDENS	0.0216	1.09	1.32	1.32	1.32	1.13	0.7712
CROIRAC	0.0017	3.83	0.0017	0.0017	0.0017	2574.3	0.0017
DLAIMAX	2.9×10^{-6}	0.1818	1.87×10^{-5}	1.60×10^{-4}	1.72×10^{-5}	5.60×10^{-4}	9.56×10^{-5}
EXTIN	0.0678	0.0731	0.0731	0.0731	0.0678	0.1362	0.0724
KMAX	0.0441	4.93	4.17	4.32	4.17	11.38	0.1678
LVOPT	0.0045	0.0045	0.0045	0.0045	0.0045	0.0931	0.0070
PSISTO	5.24	336588.9	5.24	336588.9	5.24	233.80	5.24
PSISTURG	0.0441	6.12	0.6767	6.01	0.4124	1149.7	0.2680
RAYON	1.1×10^{-5}	1.1×10^{-5}	1.1×10^{-5}	1.1×10^{-5}	1.1×10^{-5}	0.0071	1.1×10^{-5}
TCMIN	1.21	8782.4	8675.5	8779.8	8675.5	4890.3	7.29
TCOPT	0.01	0.01	0.01	0.01	0.01	0.01	0.01
ZPENTE	47.61	47.61	47.61	47.61	47.61	2.84×10^{11}	47.61
ZPRLIM	24.01	511.5	473.6	473.6	191.7	343957	24.01

Table XII. Mean squared error of prediction for predicting leaf area index on day 20 using different parameter estimation methods. These are averages over 80 training samples, each with n=14 individuals. The first vector of initial parameter values is used. The different columns correspond to predictions for different distributions of explanatory variables. The first row uses the initial parameter values. The last row is the minimal theoretical mean squared error of prediction, equal to the residual variance.

	$MSEP_1^{d=0}$	$MSEP_1^{d=1}$	$MSEP_1^{d=2}$	$MSEP_1^{d=3}$	$MSEP_1^{d=4}$	$MSEP_1^{d=5}$
Initial	0.6297	0.4573	0.6931	0.6513	0.5011	0.3978
AIC	0.4362	0.3969	0.4647	1.3112	0.4084	0.6298
AIC_C	0.4189	0.3843	0.4370	0.9009	0.3958	0.5039
BIC	0.4225	0.3904	0.4446	0.9724	0.3994	0.5297
BIC_C	0.4141	0.3823	0.4290	0.8657	0.3917	0.4946
CV	0.4360	0.4132	0.4741	0.9621	0.4208	0.5610
BIC_C^B	0.3976	0.3766	0.4098	0.4869	0.3811	0.4012
Minimum	0.36	0.36	0.36	0.36	0.36	0.36

larger errors for the parameter values but better predictions. Of course, there is no assurance that the predictions will be better for all possible distributions of the explanatory variables and indeed, for distribution $D_{X_5}(x)$, BIC_C^B does not predict better than the initial values.

We have just argued that smaller errors in the parameters does not necessarily imply smaller prediction errors. Thus the fact that BIC_C^B has smaller errors for the parameters than the other methods does not suffice to explain why this method predicts better. To get a better understanding of these results, it is helpful to examine the different training samples individually. Figure 2 shows the difference between $MSEP_1^{d=3}(m = BIC_C)$ and $MSEP_1^{d=0}(m = BIC_C)$ for each training sample. Remember that $D_{X_0}(x)$ corresponds to interpolation and $D_{X_3}(x)$ to extrapolation. The figure shows that while for most of the training samples the difference is very small, for a small number of samples the difference is quite large. The largest differences correspond to training samples where both the parameter *ADENS* and one of the parameters *BDENS*, *DLAIMAX* or *TCMIN* are estimated. In these cases there is a strong correlation between the estimators of the two parameters, and as a consequence, the estimates are far from the true values. For example, for the sample 55, the correlation between the estimators of *ADENS* and *TCMIN* is equal to 0.99. The estimated values are, respectively, equal to –3.03 and –779.0 °C. These bad estimates more or less cancel out in interpolation, but can lead to very bad predictions in extrapolation. This is typical behavior for parameter estimation in the face of high correlations.

The constraints imposed on the parameter values in the method BIC_C^B make it impossible to have parameter estimates that are totally unreasonable, like the values quoted above. The estimated values of *ADENS* and *TCMIN* in the case cited above are now –1.51 and –10.0 °C, respectively. That is the basic reason that BIC_C^B is the best model selection method. It avoids the worst prediction errors of the other methods.

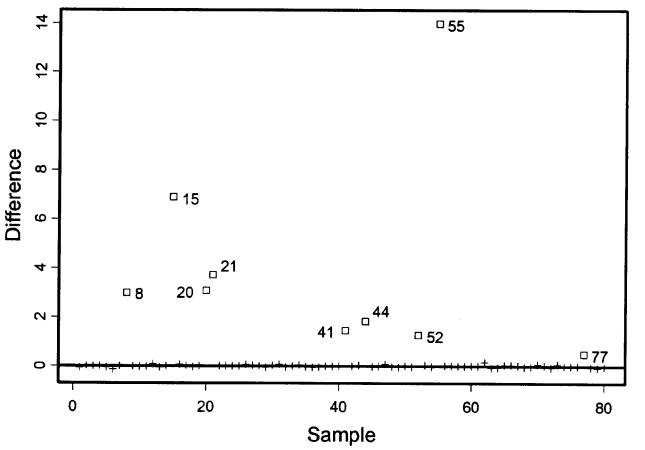

Figure 2. Difference between $MSEP_1^{d=3}(m = BIC_C)$ and $MSEP_1^{d=0}(m = BIC_C)$ for each of the training samples, for $n = 14$ and the first vector of initial values.

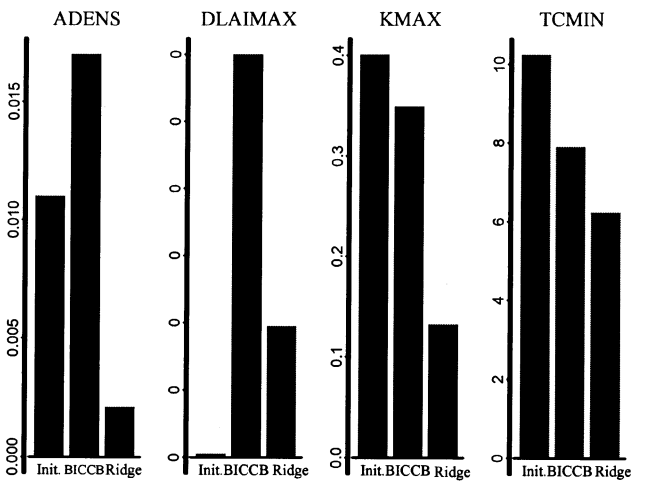

Figure 3. Mean squared error (*MSE*) of the parameters *ADENS*, *DLAIMAX*, *KMAX* and *TCMIN* estimated by BIC_C^B and by ridge regression for $n = 14$ and the second vector of initial values.

3.2. Comparison between the best model selection method and ridge regression

In the second simulation study, the best model selection method, BIC_C^B, was compared with ridge regression. Here the training sample size was $n = 14$. Four responses were considered, both for parameter estimation and for predictive quality. The following results correspond to the second initial parameter distribution. The results for the first vector of initial values were similar.

Consider first the effect of having 4 responses instead of only 1. The method BIC_C^B estimates 1 or 2 parameters. That is, the increase in the amount of data did not lead to an increase in the number of estimated parameters. There is, however, a difference in the choice of parameters to estimate. The most frequently selected parameters are here *TCMIN*, *ADENS*, *DLAIMAX* and *KMAX*. When only leaf area index at time 20 was used to estimate the parameters, the parameters most frequently selected by BIC_C^B were *ADENS, TCMIN*, *DLAIMAX* and *BDENS*, parameters which are all directly linked to leaf growth. The *KMAX* parameter, selected for only 1 training sample in the case of one response, is now selected 9 times. A sensitivity analysis (in which each parameter was varied by 30%) showed that *KMAX* is the parameter to which *HUR* is the most sensitive. It is thus not surprising that adding measurements of *HUR* to the training data leads to estimating this parameter.

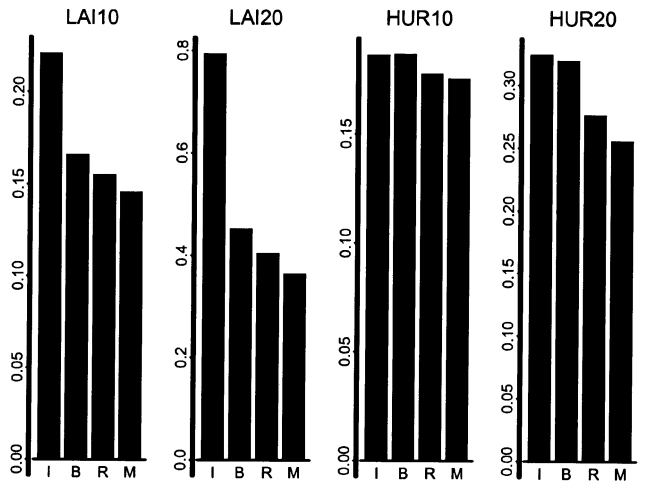

Figure 4. Mean squared errors of prediction. The explanatory variables are drawn from the distribution $D_{X_0}(x)$, $n = 14$ and the second vector of initial values is used. The index $j = 1, \ldots, 4$, indicates which response is considered. The columns labeled "I", "B", "R" and "M" correspond, respectively, to using the initial parameter values, to using the method BIC_C^B for parameter estimation, to using ridge regression for parameter estimation and to the minimal value of $MSEP_j^{d=0}$.

Figure 3 presents the mean squared errors for *ADENS*, *DLAIMAX*, *KMAX* and *TCMIN*. In this case, BIC_C^B has smaller errors than the initial values for 2 of the 4 parameters. For all four parameters, ridge regression has smaller errors than BIC_C^B. The difference is particularly flagrant for *DLAIMAX*, whose initial value is close to the true value. Here both methods are worse than the initial value, but ridge regression is much better than BIC_C^B.

Figure 4 shows the values of $MSEP_j^{d=0}(m)$, $j = 1, \ldots, 4$, for the initial values of the parameters and the two estimation methods, as well as the minimal value of *MSEP*. Both estimation methods reduce prediction error compared with the initial parameter values. The improvement is substantial for the prediction of *LAI*, but less significant for *HUR*.

Figure 5 presents the values of $MSEP_j^{d=5}(m)$. Now the prediction errors for *HUR* on day 10 and on day 20 are greater with BIC_C^B than for the initial parameter values, while the values of these prediction errors for ridge regression are comparable with the values for the initial values, themselves close to the minimal values. Similar behavior is observed for the other distributions of the explanatory variables.

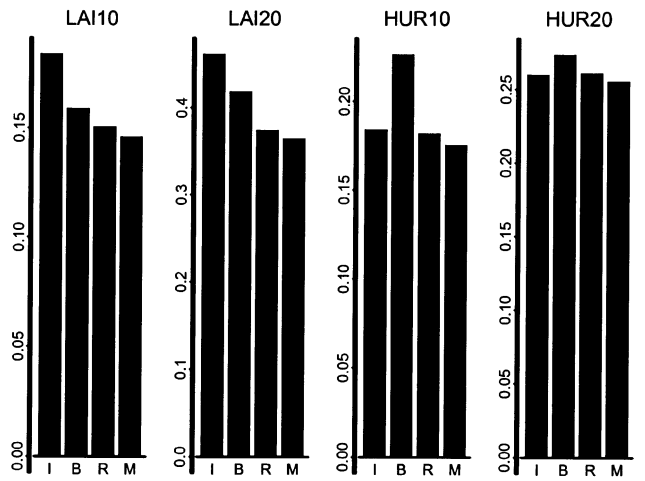

Figure 5. Mean squared errors of prediction. The explanatory variables are drawn from the extrapolation distribution $D_{X_5}(x)$, n = 14 and the second vector of initial values is used. The index j = *1,..., 4,* indicates which response is considered. The columns labeled "I", "B", "R" and "M" correspond, respectively, to using the initial parameter values, to using the method $BIC_C{}^B$ for parameter estimation, to using ridge regression for parameter estimation and to the minimal value of $MSEP_j^{d=5}$.

4. CONCLUSIONS

Several model selection methods were compared. Among these methods, $BIC_C{}^B$ is the best both in terms of mean squared error of the parameter estimates and in terms of prediction error for 6 different distributions of the explanatory variables. Of course, the results could be different for different minimum and maximum bounds on the parameter values. In particular, if the constraints are very weak, they will be less useful in forcing the parameter values to be reasonable. If the constraints are strong but exclude all good parameter values, then prediction error will be increased. The initial values used for the nonselected parameters are also important and the results might also be different depending on these values. However, we might have a bit more confidence here, since we did in fact test two different sets of initial values.

The optimal number of parameters to estimate for training samples of size n = 14 was found to be 1–3. The number of parameters selected by $BIC_C{}^B$ is 1–2. This underlines the importance of estimating relatively few parameters with the model selection methods.

The comparison between $BIC_C{}^B$ and ridge regression is to the advantage of ridge. Ridge regression leads to smaller prediction errors, even though the differences we found are not very large, and can lead to much better estimates of the parameter values. This is no doubt due to the fact that with $BIC_C{}^B$ only 1 or 2 parameters are estimated. In order to compensate for errors in the other parameters, the estimated parameters may take values that are rather far from the true values. On the other hand, this situation does not occur with ridge regression, where all parameters are estimated simultaneously.

One can make some remarks about the comparison between the results here and certain common practices concerning parameter estimation for crop models. A common practice is to estimate only certain of the model parameters, and to impose constraints on the values of those parameters. If the choice of methods is limited to model selection methods, then both these practices are certainly supported by the results here. One should severely limit the number of parameters to estimate (a number in the range of 1/10 to 1/20 seems reasonable), and it is advantageous to constrain the parameters. It is probably important that the upper and lower limits really bracket the true value without defining an excessively large interval.

The choice of parameters to estimate is no doubt quite important. All the methods here choose these parameters on the basis of their joint ability to increase the likelihood. This seems logical, but we did not explicitly test other approaches.

An important result here is that ridge regression in a form corresponding to a Bayesian approach was found to be better than the best model selection method. This should orient future research into the problem of parameter estimation.

The importance of evaluating model prediction quality for different conditions (different distributions of the explanatory variables) should also be stressed. Notice that this multiplicity of prediction criteria, a consequence of the multiple possible uses of crop models, is a specificity of this type of model. In most regression studies, the model is just tested on data generated in the same way as the data in the training sample. All of the methods give good results for interpolation (prediction for conditions drawn from the same distribution as the training sample), but can differ substantially for extrapolation (prediction for conditions drawn from a different distribution than the training sample.

The ridge regression approach supplies only the mode of the posterior parameter distribution. It would be very advantageous to have the full posterior parameter distribution or a sample therefrom, which would make it possible to generate a distribution of model results. This could be one answer to the problem of defining "validity domain" as far as parameter errors are concerned. Given the posterior distribution and a vector of explanatory variables, one could calculate the distribution of model results. If this distribution is very dispersed, the model is of little use (not "valid") for these conditions. If the distribution is very concentrated, the model gives relatively precise results ("valid").

The MCMC (Markov Chain Monte Carlo) algorithms should make it possible to generate a sample from the posterior parameter distribution, even for very complex models. We are currently working on applying the Metropolis Hastings algorithm [15, 23] to the Mini-STICS model.

Another extension of this work is to consider cases other than those where the same variables are measured at the same dates for all the individuals of the training sample. Generally, with true data the measured variables are not of the same type and/or the measurements are not made on the same dates. The structure of the variance-covariance matrix of the measured responses can then be much more complex than here.

APPENDIX: MODEL EQUATIONS

The state variables, parameter and explanatory variables of the Mini-STICS model are presented (name and meaning) in Tables I and II, respectively.

Appendix Table I. Parameters assumed to be constant.

Name	Signification	Value
ACLIM	climatic component of A	20 mm
$NBDJ_{AMF}$ $_{LAX}$	sum of development units between the stages AMF and LAX	420 degree-day
Q_0		
TDMAX	maximum threshold for the development	32 °C
TDMIN	minimum threshold for the development	6 °C
ZESX		

The parameters *ACLIM, Q_0, ZESX*, $NBDJ_{AMF\text{-}LAX}$, *TDMAX* and *TDMIN* are assumed to be fixed. Their values are presented in Appendix Table I. Other variables are intermediate variables and are presented in Appendix Table II.

Efficient temperature for development

$$UDEVCULT(J)=\begin{cases}0\\ TCULT(J-1)-TDMIN\\ TDMAX-TDMIN\end{cases}$$

$$\begin{cases}\text{if } TCULT(J-1)>TDMIN\\ \text{if } TDMIN\le TCULT(J-1)<TDMAX\\ \text{else}\end{cases}$$

Root Depth

$$DELTAZ(J)=\min\left(DELTAZEff(J),PROFSOL-ZRAC(J-1)\right),$$

with

$$DELTAZEff(J)=CROIRAC\times\left(\min(TCULT(J-1),TCOPT)-TCMIN\right)^{2}$$

$$\times\ PFZ(ZRAC(J-1))(J)$$

and

$$PFZ(Z)(J)=\begin{cases}1 & \begin{cases}\text{if } HUR(Z)(J-1)>HUMIN(Z)\\ \text{else}\end{cases}\\ 0\end{cases}.$$

Leaf Area Index

$$DELTAI(J)=\frac{DLAIMAX}{1+\exp\left[5.5\times(2.2-ULAI(J))\right]}$$

$$\times\left(TCULT(J-1)-TCMIN\right)\times TURFAC(J)$$

$$\times\ EFDENSITE\times DENSITE$$

Appendix Table II. Intermediary variables.

Name	Signification	Units
A	parameter of the soil evaporation during the reduction stage	mm
AMF	maximal acceleration of leaf growth, end in the juvenile phase	
CUMULRACZ	sum of the effective root lengths	cm root.cm^{-2} soil
DELTAI	daily increase in the leaf index	m^2 leaves.m^{-2} soil.day^{-1}
DELTAZ	deepening of the root front	cm
DRAIN	water flux drained out of the soil	mm.day^{-1}
EFDENSITE	density effect acting on the setting up of the LAI	–
EO	intermediary variable for the computation of evapotranspiration	mm.day^{-1}
EOP	maximum transpiration flux	mm.day^{-1}
EOS	maximum evaporation flux	mm.day^{-1}
EP	actual transpiration flux	mm.day^{-1}
EPAIS(Z)	thickness of the layer Z	cm
ES	actual soil evaporation flux	mm.day^{-1}
HA	residual moisture of the soil	cm^3.cm^{-3}
HI	volumetric water content at the field capacity of the surface layer	mm.cm^{-1}
ILAX	day of the stage LAX	–
ILEV	day of the stage LEV: emergence	–
J	running day	–
LAX	maximal leaf area index	
LRAC(Z)	effective root density in the layer Z	cm root.cm^{-3} soil
PFZ(Z)	water status of the layer Z	–
PROF(Z)	depth of the layer Z	
S	parameter for the calculation of root density	cm^{-1}
SWFAC	stomatal stress index	–
TCULT	surface temperature in daily average	°C
TETA	water content of the soil, available to the plant	cm^3.cm^{-3}
TETSTOMATE	threshold of TETA below which SWFAC decreases	cm^3.cm^{-3}
TETURG	threshold of TETA below which TURFAC decreases	cm^3.cm^{-3}
TURFAC	turgescence stress index	–
UDEVCULT	effective temperature for the development, computed with TCULT	°C
ULAI	physiological time units for the calculation of the leaf area index between ILEV and ILAX	–
Z	layer index	–
ZDEMI	depth where the root density is ½ of the surface root density	cm

with

$$ULAI(J) = 2.2 + (3 - 2.2) \times \frac{SUDEVCULT(J)}{NBDJ_{AMF-LAX}}$$

and

$$EFDENSITE = \left(\frac{DENSITE}{BDENS} \right)^{ADENS}.$$

Soil Water Evaporation

$$EOS(J) = ETP(J) \times \exp[-(EXTIN - 0.2) \times LAI(J)]$$

$$HA = \frac{ARGI/100}{15} \times DA(1)$$

$$HI = \frac{HUCC(1)}{10}$$

$$A = 0.5 \times ACLIM \times (0.63 - HA)^{5/3} \times (HI - HA)$$

$$\sum ES(J) = \sqrt{2 \times A \times \sum EOS(J) + A^2} - A$$

$$\sum EOS(J) = \begin{cases} \sum EOS(J-1) + EOS(J) \\ EOS(J) \end{cases}$$

$$\text{if } \begin{array}{l} \sum ES(J-1) > PLUIE(J) \\ \sum ES(J-1) \le PLUIE(J) \end{array}$$

$$ES(J) = \begin{cases} \sum ES(J) - \sum ES(J-1) \\ \sum ES(J) \end{cases}$$

$$\text{if } \begin{array}{l} \sum ES(J-1) > PLUIE(J) \\ \sum ES(J-1) \le PLUIE(J) \end{array}$$

$$ES(1)(J) = \min(HUR(1)(J-1) - 10HA \times EPAIS(1), ES(J))$$

$$HUR(1)(J) = HUR(1)(J-1) - \frac{ES(1)(J)}{EPAIS(1)}.$$

Transpiration

$$EO(J) = ETP(J)\left[1 + \frac{KMAX - 1}{1 + \exp[-1.5 \times (LAI(J) - 3)]}\right]$$

$$EOP(J) = (EO(J) - EOS(J))\left(1.4 - 0.4\frac{ES(J)}{EOS(J)}\right)$$

$$S = \frac{-4.6}{ZLABOUR - ZPENTE}$$

$$ZDEMI(J) = \max\left(ZRAC(J) - ZPRLIM + ZPENTE, \frac{1.4}{S}\right)$$

$$LRAC(Z)(J) = \frac{LVOPT}{1 + \exp[-S \times (PROF(Z) - ZDEMI(J))]}$$

$$CUMULRACZ(J) = \sum LRAC(Z)(J) \times EPAIS(Z),\ Z = 1 \text{ or } 2$$

$$TETSTOMATE(J) = \frac{1}{40} \ln\left[\frac{EOP(J)}{2\pi \times CUMULRACZ(J) \times PSISTO \times 10^{-3}}\right] \times \ln\left[\frac{1}{RAYON}\sqrt{\frac{ZRAC(J)}{\pi \times CUMULRACZ(J)}}\right]$$

$$TETA(J) = \begin{cases} \max\left(\frac{HUR(1)(J) - HUMIN(1)}{10}, 0\right) & \text{if } ZRAC(J) \le EPAIS(1) \\ \frac{TETA(1)(J) + TETA(2)(J)}{ZRAC(J) \times 10} & \text{else} \end{cases},$$

with

$$TETA(1)(J) = \max((HUR(1)(J) - HUMIN(1)) \times EPAIS(1), 0)$$

and

$$TETA(2)(J) = \max((HUR(2)(J) - HUMIN(2)) \times EPAIS(2), 0).$$

$$EP(J) = \begin{cases} EOP(J) & \text{if } TETA(J) > TETSTOMATE(J) \\ \frac{EOP(J) \times TETA(J)}{TETSTOMATE(J)} & \text{else} \end{cases}$$

$$EP(Z)(J) = \frac{EP(J) \times LRAC(Z(J)) \times EPAIS(Z)}{CUMULRACZ(J)}$$

$$EP(1)(J) = \min[(HUR(1)(J) - HUMIN(1)) \times EPAIS(1), EP(1)(J)]$$

$$EP(2)(J) = \min[(HUR(2)(J-1) - HUMIN(2)) \times EPAIS(2), EP(J) - EP(1)(J)]$$

$$HUR(1)(J) = HUR(1)(J) - \frac{EP(1)(J)}{EPAIS(1)}$$

$$HUR(2)(J) = HUR(2)(J-1) - \frac{EP(2)(J)}{EPAIS(2)}.$$

Water Budget

$$HUR(1)(J) = \frac{HUR(1)(J) \times EPAIS(1) + PLUIE - DRAIN(1)(J)}{EPAIS(1)},$$

with

$$DRAIN(1)(J) = \begin{cases} HUR(1)(J) \times EPAIS(1) + PLUIE(J) - HUCC(1) \times EPAIS(1) \\ 0 \end{cases}$$

$$\begin{cases} \text{if } HUR(1)(J) \times EPAIS(1) + PLUIE > HUCC(1) \times EPAIS(1) \\ \text{else} \end{cases}$$

$$HUR(2)(J) = \frac{HUR(2)(J) \times EPAIS(2) + DRAIN(1)(J) - DRAIN(2)(J)}{EPAIS(2)},$$

with

$$DRAIN(2)(J) = \begin{cases} HUR(2)(J) \times EPAIS(2) + DRAIN(1)(J) - HUCC(2) \times EPAIS(2) \\ 0 \end{cases}$$

$$\begin{cases} \text{if } HUR(2)(J) \times EPAIS(2) + DRAIN(1)(J) > HUCC(2) \times EPAIS(2) \\ \text{else} \end{cases}$$

Stress Index

$$TETURG(J) = \frac{1}{40} \ln\left[\frac{EOP(J)}{2\pi \times CUMULRACZ(J) \times PSISTURG \times 10^{-3}}\right] \times \ln\left[\frac{1}{RAYON}\sqrt{\frac{ZRAC(J)}{\pi \times CUMULRACZ(J)}}\right]$$

$$TURFAC(J) = \begin{cases} 1 & \begin{cases} \text{if } TETA(J) > TETURG(J) \\ \text{else} \end{cases} \\ \dfrac{TETA(J)}{TETURG(J)} \end{cases}$$

REFERENCES

[1] Akaike H., Information theory and an extension of the maximum likelihood principle, in Proc. 2nd Int. Symp. Information Theory, Supp. to Problems of Control and Information Theory (1972), pp. 267–281.

[2] Bates B.C., Campbell E.P., A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modelling, Water Resour. Res. 37 (2001) 937–947.

[3] Bedrick E.J., Tsaï C.L., Model selection for multivariate regression in small samples, Biometrics 50 (1994) 226–231.

[4] Beven K., Binley A., The future of distributed models: model calibration and uncertainty prediction, Hydrol. Process. 6 (1992) 279–298.

[5] Breiman L., Spector P., Submodel selection and evaluation in regression. The X-random case, Int. Stat. Rev. 60 (1992) 291–319.

[6] 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., Delecolle R., STICS: a generic model for the simulation of crops and their water and nitrogen balances. I Theory and parametrization applied to wheat and corn, Agronomie 18 (1998) 1–346.

[7] Brun R., Reichert P., Künsch H.R., Practical identifiability analysis of large environmental simulation models, Water Resour. Res. 37 (2001) 1015–1030.

[8] Campbell P.C., Fox D.R., Bates B.C., A bayesian approach to parameter estimation and pooling in nonlinear flood event models, Water Resour. Res. 35 (1999) 211–220.

[9] Durand P., Gascuel-Odoux C., Cordier M.O., Parametrisation of hydrological models: a review and lessons learned from studies of an agricultural catchment (Naizin, France), Agronomie 22 (2002) 217–228.

[10] Franks S.W., Gineste P., Beven K.J., Merot P., On constraining the predictions of a distributed model: The incorporation of fuzzy estimates of saturated areas into the calibration process, Water Resour. Res. 34 (1998) 787–797.

[11] Gallant R., Nonlinear statistical models, John Wiley & Sons, 1987.

[12] Guiloufi M., Méthodologie d'adaptation de STICS à de nouvelles cultures : application au tournesol et à la betterave à sucre, Rapport de DAA, INA-PG (1999) 48 p. + annexes.

[13] Harmon R., Challenor P., A Markov Chain Monte Carlo method for estimation and assimilation into models, Ecol. Model. 101 (1997) 41–59.

[14] Harrel F.E., Regression modelling strategies: with applications to linear models, logistic regression, and survival analysis, Springer-Verlag, New York, 2001.

[15] Hastings W.K., Monte Carlo sampling methods using Markov chains and their applications, Biometrika 57 (1970) 97–109.

[16] Hoerl A.E., Kennard R.W., Ridge regression: biased estimation for non-orthogonal problems, Technometrics 12 (1970) 55–68.

[17] Hurvich C.M., Tsaï C.L., Regression and time series model selection in small samples, Biometrika 76 (1989) 297–307.

[18] IMSL Fortran 90 MP Library, Visual Numerics, Houston, Texas, USA, 1996.

[19] Kuczera G., Parent E., Monte Carlo assessment of parameter uncertainty in conceptual catchement models: the Metropolis algorithm, J. Hydrol. 211 (1998) 69–85.

[20] Makowski D., Wallach D., Tremblay M., Using a Bayesian approach to parameter estimation; comparison of the GLUE and MCMC methods, Agronomie 22 (2002) 191–203.

[21] McQuarrie A.D., Tsaï C.L., Regression and time series model selection, World Scientific Publishing Co. Pte. Ltd, 1998.

[22] McQuarrie A.D., A small-sample correction for the Schwarz SIC model selection criterion, Stat. Prob. Lett. 44 (1999) 79–86.

[23] Metropolis N., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E., Equation of state calculations by fast computing machines, J. Chem. Phys. 21 (1953) 1087–1091.

[24] Olsthoorn T.N., Effective parameter optimization for ground water model calibration, Groundwater 33 (1995) 42–48.

[25] Omlin M., Reichert P., A comparison of techniques for the estimation of model prediction uncertainty, Ecol. Model. 115 (1999) 45–59.

[26] Perrin C., Michel C., Andréassian V., Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments, J. Hydrol. 242 (2001) 275–301.

[27] Pinheiro J.C., Bates D.M., Mixed-effects models in S and S-PLUS, Springer-Verlag, New York, 2000.

[28] Refsgaard J.S., Parametrisation, calibration and validation of distributed hydrological models, J. Hydrol. 198 (1997) 69–97.

[29] Schulz K., Beven K., Huwe B., Equifinality and the problem of Robust Calibration in Nitrogen Budget simulations, Soil Sci. Soc. Am. J. 63 (1999) 1934–1941.

[30] Schwarz G., Estimating the dimension of a model, Ann. Stat. 6 (1978) 461–464.

[31] Sievänen R., Burk T.E., Adjusting a process based growth model for varying site conditions through parameter estimation, Can. J. For. Res. 23 (1993) 1837–1851.

[32] Stone M., Cross-validatory Choice and Assessment of Statistical Predictions, J. Roy. Stat. Soc. Ser. B 36 (1974) 111–147.

[33] Van der Perk M., Calibration and identifiabilty analysis of a water quality model to evaluate the contribution of different processes to the short-term dynamics of suspended sediment and dissolved nutrients in the surface water of a rural catchment, Hydrol. Process. 12 (1997) 683–699.

[34] Vold A., Breland T.A., Soreng J.S., Multiresponse estimation of parameter values in model of soil carbon and nitrogen dynamics, J. Agric. Biol. Environ. Stat. 4 (1999) 290–309.

[35] Wallach D., Goffinet B., Bergez J.E., Debaeke P., Leenhardt D., Aubertot J.N., Parameter estimation for crop models: a new approach and application to a corn model, Agron. J. 93 (2001) 757–766.

[36] Wallach D., Goffinet B., Tremblay M., Parameter estimation in crop models: exploring the possibility of estimating linear combinations of parameters, Agronomie 22 (2002) 171–178.

[37] Xevi E., Christiaens K., Espino A., Sewnandan W., Mallants D., Sorensen H., Feyen J., Calibration, validation and sensitivity analysis of the MIKE-SHE model using the Neuenkirchen catchment as case study, Water Res. Manage. 11 (1997) 219–241.

[38] Yan J., Haan C.T., Multiobjective parameter estimation for hydrologic models. Weighting of errors, Trans. ASAE 34 (1991) 135–141.

[39] Zhang X., Lindström G., Development of an automatic calibration scheme for the HBV hydrological model, Hydrol. Process. 11 (1997) 1671–1682.