Multivariate global sensitivity analysis for discrete-time models.

Matieyendou Lamboni*, INRA, Unité MIA-Jouy, Jouy en Josas, France.

David Makowski, UMR 211 INRA, AgroParisTech, Gignon, France.

Hervé Monod, INRA, Unité MIA-Jouy, Jouy en Josas, France.

Rapport technique 2008-3, 17pp.

Unité Mathématiques et Informatique Appliquées
INRA
Domaine de Vilvert
78352 Jouy-en-Josas Cedex
France

* matieyendou.lamboni@jouy.inra.fr

Abstract

Discrete-time models are frequently used in ecology and agronomy. These models can be used for the management of endangered species, for understanding intraspecific and interspecific competitions, for pest management, or for predicting plant growth. Their outputs can be expressed as time series. It is often impossible to estimate all the parameters of a discrete-time model due to their large number. A common practice consists in selecting a subset of parameters by sensitivity analysis, in estimating the selected parameters from data, and in fixing the others to some nominal values. For a discrete-time model, global sensitivity analysis can be applied separately on each output, but there is a high level of redundancy between close dates and, on the other hand, interesting features of the dynamic may be missed out. In this paper, a method based on principal component analysis and on analysis of variance is presented to compute a generalized sensitivity index for each model parameter. The proposed index synthesizes the influence of the parameter on the whole time series output. It may be used to select a subset of parameters to be calibrated. In addition a quality criterion is proposed for any approximation associated with the ANOVA decomposition on the principal components. The method was applied to a winter wheat dynamic model including seven parameters with few observations for estimating all the parameters. The results showed that two parameters had a strong influence on the wheat biomass simulated by the model at a daily time step: the radiation use efficiency and a parameter of the mathematical function describing the kinetic of the leaf area index. We also showed that the generalized index can be accurately computed by using only the first three principal components. The proposed approach is quite general and can be applied to any dynamic model predicting one or several output variables at a discrete time step.

Keywords: ANOVA decomposition, Discrete-time model, Experimental design, Multivariate sensitivity analysis.

1. Introduction

Discrete-time models are frequently used in ecology and agronomy. These models are useful for management of endangered species (e.g Santangelo et al., 2007), for understanding intraspecific and interspecific competitions (e.g Yakubu et al., 2002; Wu et al., 2007), for pest management (e.g Matsuoka and Seno, in press), and for predicting plant growth (e.g Boote *et al*., 1996; Passioura, 1996; Bechini *et al*, 2005). Their outputs can be expressed as time series.

Discrete-time models can include up to several hundreds parameters whose values must be estimated from past experiments (Makowski et al. 2006a). The estimation of these parameters is an important step because model performances depend for a large part on the accuracy of the parameter estimates (Wallach *et al.*, 2001; Makowski *et al.*, 2006a). Predictions obtained with models are not reliable when inaccurate parameter values are used.

In general, it is impossible to estimate all parameters of complex models (Bechini *et al*, 2005). A strategy consists in selecting a subset of parameters to be calibrated by sensitivity analysis and in fixing the others to some nominal values (Wallach *et al.*, 2001; Makowski *et al*., 2006a; Makowski *et al.*, 2006b; Monod *et al.*, 2006). Several local and global sensitivity analysis methods have been developed and applied for identifying the parameters that deserve an accurate estimation (Homma and Saltelli, 1996; Saltelli *et al.*, 2000b; Saltelli *et al.*, 2004; Saltelli *et al.*, 2006; Cariboni *et al.*, 2007). Methods of global sensitivity analysis are useful and are easy to interpret. They allow modellers to determine which subset of parameters accounts for most of the output variance. Those factors with a small contribution can be set equal to any value within their range. This contributes to a model simplification and a reduction of the number of experiments performed for estimating model parameters.

For a discrete-time model, global sensitivity analysis can be applied separately on each output but there is a high level of redundancy between close dates and, on the other hand, interesting features of the dynamic may be missed out. The application of a sensitivity analysis method to the daily output of a dynamic model can result in a very large number of sensitivity indices (one index per daily output). It is not easy to identify the most important parameters based on such a large number of values (Campolongo *et al*, 2007). As an alternative, Campbell *et al.* (2006) proposed to decompose time series upon a complete orthogonal basis and to compute sensitivity indices on each component of the decomposition. However, no single global index has been proposed for summarizing the sensitivity of a time series to parameter values.

In this paper, we follow on this proposal and present the multivariate sensitivity analysis under a global framework coherent with classical multivariate methods. A generalized index is defined which synthesizes the effect of each model parameter on the whole time series output. It may be used to select a subset of parameters to be calibrated or to simplify a complex model. In addition a quality criterion is proposed for any approximation associated with the decomposition. In section 2, we present a global sensitivity analysis method adapted to discrete-time models. Section 3 illustrates this method on a dynamic crop model, section 4 presents a R-program developed for computing the generalized sensitivity index, and section 5 concludes.

2. Method

2.1. Time series output

We consider a deterministic and dynamic model with discrete time step

$$y(t) = f(\mathbf{z}, t; \theta), \qquad (1)$$

where $y(t)$ is the scalar output on time t for $t = 1, 2, \ldots, T$, $\mathbf{z}$ is a vector of input variables and θ is a vector of parameters. Input variables and parameters will be further referred to as input factors. In the case study presented in Section 3, the input variables will be assumed to be known so that the parameters θ will be the only sources of uncertainty to be considered in the sensitivity analysis.

Suppose that N simulation runs are performed, using equation (1) with different values of the input factors. Then the output can be stored in a $N \times T$ matrix:

$$\mathbf{Y} = \begin{pmatrix} y_1(1) & \ldots & y_1(t) & \ldots & y_1(T) \\ . & . & . & . & . \\ y_i(1) & \ldots & y_i(t) & \ldots & y_i(T) \\ . & . & . & . & . \\ y_N(1) & \ldots & y_N(t) & \ldots & y_N(T) \end{pmatrix}.$$

Each column $\mathbf{y}(t)$ in $\mathbf{Y}$ represents the simulated values of the output variable on a given time t, while each row of $\mathbf{Y}$ is an individual dynamic for a given set of input values. The rows of $\mathbf{Y}$ constitute a sample of output dynamics in $\mathbf{R}^T$ over the uncertainty domain of the input factors. In the sequel, we assume that $N \geq T$.

Conducting separate sensitivity analyses on $\mathbf{y}(1), \ldots, \mathbf{y}(T)$ gives information on how the sensitivity of $y(t)$ evolves over time. However, it leads to a high level of redundancy because of the strong relationship between responses from one date to the next one. It may also miss important features of the $y(t)$ dynamics because many features cannot be efficiently detected through single-time measurements.

Alternatively, sensitivity analyses can be applied to pre-defined functions $h(y(1), \ldots, y(T))$ with an biological interpretation. For example, h may represent the difference in biomass between two stages of plant growth. Many functions of $y(1), \ldots, y(T)$ are potentially interesting to look at and there is a need for methods to identify automatically the most interesting features in the $y(t)$ dynamics.

2.2. Principal Components Analysis

The Principal Components Analysis (PCA) is a method to decompose the whole variability, or total inertia, in $\mathbf{Y}$ (Krzanowski and Marriott, 1990). The total inertia is defined as $I(\mathbf{Y}_c) = \mathrm{trace}(\mathbf{Y}_c{}'\mathbf{Y}_c)$, where $\mathbf{Y}_c$ is the matrix $\mathbf{Y}$ with each column centered around its mean and possibly normalized. Thus $\mathbf{Y}_c{}'\mathbf{Y}_c$ is the empirical variance-covariance matrix of the columns of $\mathbf{Y}$ if $\mathbf{Y}_c$ has not been normalized, or it is the empirical correlation matrix between the columns of $\mathbf{Y}$, with 1 on the diagonal, if the columns of $\mathbf{Y}_c$ have been normalized. It follows that the total inertia is also equal to the sum of the $\mathbf{y}(t)$ variances, or to T if the columns in $\mathbf{Y}_c$ have been normalized. Saporta (2006) provides a detailed treatment of principal component analysis methodology.

The PCA decomposition is based on the eigenvalues and eigenvectors of $\mathbf{Y}_c'\mathbf{Y}_c$. Let $\lambda_1,\ldots,\lambda_T$ denote the eigenvalues of $\mathbf{Y}_c'\mathbf{Y}_c$ in decreasing order (with eigenvalues repeated according to their multiplicity if any) and let $\mathbf{L}$ denote a $T\times T$ matrix of normalized eigenvectors of $\mathbf{Y}_c'\mathbf{Y}_c$, with column $\mathbf{l}_k$ an eigenvector associated with λ_k. The $N\times T$ matrix $\mathbf{H}$ of Principal Components (PC) is obtained by

$$\mathbf{H}=\mathbf{Y}_c\mathbf{L}. \tag{2}$$

Its columns $\mathbf{h}_k$, for $k=1,2,\ldots,T$, are mutually orthogonal linear combinations of the $\mathbf{Y}_c$ columns satisfying $\mathbf{h}_k=\mathbf{Y}_c\mathbf{l}_k$ and $\|\mathbf{h}_k\|^2=\lambda_k$. By construction, $\mathbf{H}$ has the same total inertia as $\mathbf{Y}_c$, but this inertia is concentrated as much as possible in the first principal components.

Example

To follow this approach more easily, let us take a very simple example

$$y(t)=a+b\times t$$

where the uncertain values of parameters a and b are assumed to vary in the interval $[-1,1]$. For a maximum of simplicity, we pay attention to the model output at $t=0$ and $t=T$ only, where T represents the latest time of interest, and we consider four simulations performed at the ±1 combinations of a and b (*a=1* and *b=1*; *a=1* and *b=−1*; *a=−1* and *b=*1; *a=−1* and *b=−1*).

In this example, the $\mathbf{Y}$ matrix of simulation outputs and its normalized version $\mathbf{Y}_c$ are centered. They are given by

$$\mathbf{Y}=\begin{pmatrix}1 & 1+T\\ 1 & 1-T\\ -1 & -1+T\\ -1 & -1-T\end{pmatrix},\qquad \mathbf{Y}_c=\frac{1}{2}\begin{pmatrix}1 & (1+T)/\alpha\\ 1 & (1-T)/\alpha\\ -1 & (-1+T)/\alpha\\ -1 & (-1-T)/\alpha\end{pmatrix},$$

with $\alpha=\sqrt{1+T^2}$.

The eigenvalues of $\mathbf{Y}_c'\mathbf{Y}_c$ are equal to $1+\alpha^{-1}$ and $1-\alpha^{-1}$. Up to normalization constants, the first principal component is equal to the mean of $y(0)$ and $y(T)$ and the second principal component is equal to the difference between $y(0)$ and $y(T)$. More precisely, the $\mathbf{H}$ principal components matrix is

$$\mathbf{H}=\frac{1}{2\sqrt{2}\alpha}\begin{pmatrix}\alpha+1+T & \alpha-1-T\\ \alpha+1-T & \alpha-1+T\\ -\alpha-1+T & -\alpha+1-T\\ -\alpha-1-T & -\alpha+1+T\end{pmatrix}.$$

2.3. Sensitivity principal indices

Sensitivity analyses (SA) can be computed on each principal component (PC). In this paper, we consider the SA methods based on a factorial design for the simulations and analysis of variance (ANOVA) for the calculation of sensitivity indices (Jansen *et al.*, 1994; Saltelli *et al.*, 2000a; Ginot et al, 2006). We denote the s input factors by $F_1,\dots,F_s$ and each factor is supposed to vary at m distinct levels. For example, $F_1,\dots,F_s$ may represent s model parameters taking m possible values. The complete factorial design then consists of $N = m^s$ simulations performed at all combinations of levels of $F_1,\dots,F_s$. From now on, we assume that $\mathbf{Y}$ is the output matrix from those N simulations and that $\mathbf{H}$ is issued from the PCA of $\mathbf{Y}$.

Because of the orthogonality properties of the complete factorial design, there is a unique ANOVA decomposition for the variance of each principal component:

$$\left\|\mathbf{h}_k\right\|^2 = \sum_W \mathrm{SS}_{W,k} \tag{3}$$

where the W's are the factorial terms in the ANOVA (main effects and interactions) and $\mathrm{SS}_{W,k}$ denotes the sum of squares associated with W for the kth principal component. Note that the W's in the summation should include a residual term if the ANOVA model is not complete.

Recall that $\left\|\mathbf{h}_k\right\|^2 = \lambda_k$ with λ_k the inertia associated with the kth principal component. Then $\mathrm{SS}_{W,k}$ is the part of that inertia accounted for by the factorial term W and the sensitivity principal indices in the kth principal component $\mathbf{h}_k$ can be defined as

$$\mathrm{SI}_{W,k} = \frac{\mathrm{SS}_{W,k}}{\lambda_k},$$

so that their values lie between 0 and 1. The first order sensitivity indices correspond to the case when W is a main effect. The total sensitivity indices are calculated by summing the sensitivity indices $\mathrm{SI}_{W,k}$ over all factorial terms W which include a given input factor (Salteli and al, 2000).

From a technical point of view that will be useful in the sequel,

$$\mathrm{SS}_{W,k} = \left\|\mathbf{S}_W \mathbf{h}_k\right\|^2 \tag{4}$$

$$= \mathrm{trace}(\mathbf{Y}_c{}'\mathbf{S}_W \mathbf{Y}_c \mathbf{l_k} \mathbf{l_k}{}') \tag{5}$$

where $\mathbf{S}_W$ is the orthogonal projection matrix on the subspace associated with W in R^N.

Example (continued)

The complete ANOVA model includes the main effects of A and B and the interaction $A.B$. There is no general mean because the data are centered. The projection matrices are

$$S_A = \frac{1}{4}\begin{pmatrix} 1 & 1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & 1 & 1 \end{pmatrix},\quad S_B = \frac{1}{4}\begin{pmatrix} 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ -1 & 1 & -1 & 1 \end{pmatrix},\quad S_{A.B} = \frac{1}{4}\begin{pmatrix} 1 & -1 & -1 & 1 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{pmatrix}.$$

Using equation (4), we get $\mathrm{SS}_{A,1}=\frac{1}{2}(1+\alpha^{-1})^2$, $\mathrm{SS}_{B,1}=\frac{1}{2}(1-\alpha^{-2})$, $\mathrm{SS}_{A,2}=\frac{1}{2}(1-\alpha^{-1})^2$, $\mathrm{SS}_{B,2}=\frac{1}{2}(1-\alpha^{-2})$, $\mathrm{SS}_{A.B,1}=\mathrm{SS}_{A.B,2}=0$.

The sensitivity principal indices are $\mathrm{SI}_{A,1}=\frac{1}{2}(1+\alpha^{-1})$, $\mathrm{SI}_{B,1}=\frac{1}{2}(1-\alpha^{-1})$, $\mathrm{SI}_{A,2}=\frac{1}{2}(1-\alpha^{-1})$, $\mathrm{SI}_{B,2}=\frac{1}{2}(1+\alpha^{-1})$, $\mathrm{SI}_{A.B,1}=\mathrm{SI}_{A.B,2}=0$. $\mathrm{SI}_{A,1}$ and $\mathrm{SI}_{B,1}$ measure the first order sensitivity of the first principal component to the two model parameters A and B. $\mathrm{SI}_{A,2}$ and $\mathrm{SI}_{B,2}$ measure first order sensitivity of the second principal component to the two model parameters.

2.4. Generalized sensitivity indices

In addition to the sensitivity indices on each principal component, it is interesting to quantify the contribution $\mathrm{SS}_{W,\mathrm{total}}$ of each factorial term W to the total inertia. This is precisely the decomposition performed in multivariate analysis of variance (MANOVA), which is the generalisation of ANOVA to multivariate responses (Anderson, 2003). In MANOVA, the decomposition in equation (3) is applied to the total inertia and becomes

$$I(\mathbf{Y}_c)=\sum_W \mathrm{SS}_{W,\mathrm{total}} \tag{6}$$

where

$$\mathrm{SS}_{W,\mathrm{total}}=\left\|\mathbf{S}_W\mathbf{Y}_c\right\|^2 \tag{7}$$

$$=\mathrm{trace}(\mathbf{Y}'_c\,\mathbf{S}_W\mathbf{Y}_c) \tag{8}$$

$$=\sum_{k=1}^{T}\mathrm{trace}(\mathbf{Y}'_c\,\mathbf{S}_W\mathbf{Y}_c\mathbf{l_k}\mathbf{l_k}') \tag{9}$$

$$=\sum_{k=1}^{T}\mathrm{SS}_{W,k}. \tag{10}$$

because $\sum_{k=1}^{T}\mathbf{l_k}\mathbf{l_k}'$ is the $T\times T$ identity matrix.

The various decompositions of the total inertia are summarized in Table 1. The finest decomposition is into the $\mathrm{SS}_{W,k}$ terms. Sums over W give the inertia λ_k associated with the principal components, whereas sums over k give the MANOVA decomposition of the total inertia among the factorial terms. All sensitivity indices presented in this paper arise from the quantities in table 1.

To measure contributions to the total inertia, the *generalized sensitivity index* of any factorial term W is defined as

$$\mathrm{GSI}_W=\frac{\mathrm{SS}_{W,\mathrm{total}}}{I(\mathbf{Y}_c)}. \tag{11}$$

The first order generalized sensitivity indices correspond to the case when W is a main effect. The total generalized sensitivity indices are calculated by summing the generalized sensitivity indices GSI_W over all factorial terms W which include a given input factor.

Factorial term	Principal Component					Inertia
	PC_1	PC_2	PC_3	$\cdots$	PC_T	
A	$SS_{A,1}$	$SS_{A,2}$	$SS_{A,3}$	$\cdots$	$SS_{A,T}$	$SS_{A,\text{total}}$
B	$SS_{B,1}$	$SS_{B,2}$	$SS_{B,3}$	$\cdots$	$SS_{B,T}$	$SS_{B,\text{total}}$
.	.	.	.	.	.	.
$A.B$	$SS_{AB,1}$	$SS_{AB,2}$	$SS_{AB,3}$	$\cdots$	$SS_{AB,T}$	$SS_{AB,\text{total}}$
.	.	.	.	.	.	.
W	$SS_{W,1}$	$SS_{W,2}$	$SS_{W,3}$	$\cdots$	$SS_{W,T}$	$SS_{W,\text{total}}$
.	.	.	.	.	.	.
Inertia	λ_1	λ_2	λ_3	$\cdots$	λ_T	$I(\mathbf{Y}_c)$

Table 1. Sum of squares decomposition of the total inertia $I(\mathbf{Y}_c)$ based on PCA and MANOVA. A and B denote the first two parameters, $A.B$ their interaction, and W a generic factorial term.

Example (continued)

Table 2 presents the decomposition in Table 1 when applied to the example. Thus, $\mathrm{GSI}_A = \frac{1+\alpha^{-2}}{2}$ is the part of inertia explained by factor A and $\mathrm{GSI}_B = \frac{1-\alpha^{-2}}{2}$ the part of inertia explained by factor B. There is no interaction between factors A and B as expected from an additive model. Factor A has more global importance than B, but the difference tends to zero when T increases. Note that the results are strongly influenced by the normalization performed on $\mathbf{Y}$.

Factorial term	Principal Component		Inertia
	PC_1	PC_2	
A	$\frac{1}{2}(1+\alpha^{-1})^2$	$\frac{1}{2}(1-\alpha^{-1})^2$	$1+\alpha^{-2}$
B	$\frac{1}{2}(1-\alpha^{-2})$	$\frac{1}{2}(1-\alpha^{-2})$	$1-\alpha^{-2}$
$A.B$	0	0	0
Inertia	$1+\alpha^{-1}$	$1-\alpha^{-1}$	$I(\mathbf{Y}_c)=2$

Table 2. Sum of squares decomposition for the small example model.

2.5. Approximation quality

In practice, the methodology described in this section is often approximated by considering only the first P principal components and by restricting the ANOVA terms to the main effects and a few interactions. Thus, we define the *approximate generalized sensitivity index* of W as

$$\tilde{GSI}_w = \frac{\sum_{k=1}^{P} \mathrm{SS}_{W,k}}{\sum_{k=1}^{P} \lambda_k}.$$

The quality of the approximation can be quantified by the proportion of inertia preserved by the approximation. We thus define the *approximation global quality* by

$$\mathrm{GQ} = \frac{\sum_{W,k} \mathrm{SS}_{W,k}}{I(\mathbf{Y}_c)}, \tag{12}$$

where the summation is restricted to the pairs (W,k) of principal component and factorial term included in the approximation. For example, GQ in the example equals $\frac{1}{2}(1+\alpha^{-1})$ if all factorial terms but only the first principal component are kept. By analogy to the coefficient of determination, if GQ is close to 1 then the approximation accounts for most the inertia of $\mathbf{Y}$, whereas low GQ suggests that $\tilde{GSI}_w$ indices must be interpreted with much caution.

In addition to the global quality criterion, the *dynamic coefficient of determination* allows to assess the approximation quality directly on the original time series $y(t)$. In fact, going back to the time series from the approximation produces an approximate (normalized or not) output matrix $\tilde{Y}_c$. If the approximation includes the same factorial terms for all of its P principal components, the approximate output matrix $\tilde{Y}_c$ is obtained by the formula

$$\tilde{Y}_c = \sum_{W} S_W . \tilde{H} \tilde{L}', \tag{13}$$

where $\tilde{H}$ (resp. $\tilde{L}$) contains the first P columns of predicted $\hat{H}$ (resp. L) and where the summation on W is restricted to those factorial terms in the approximation. In the general case, the approximation is

$$\tilde{Y}_c = \tilde{H}_{sub} \tilde{L}', \tag{14}$$

where the kth column of $\tilde{H}_{sub}$, for $k = 1,\ldots,P$, equals $\sum_W S_W \hat{h}_k$ with the summation restricted to the W's such that the pair (W,k) is included in the approximation.

The dynamic coefficient of determination is calculated by considering the columns $y(t)$ of the matrix of outputs $\mathbf{Y}_c$ and the columns $\tilde{y}(t)$ of the approximated matrix of output $\tilde{Y}_c$. At time *t*, this coefficient is expressed as

$$R_t^2 = \frac{\sum_{i=1}^{N} \left(\tilde{y}_i(t) - \overline{y}(t) \right)^2}{\sum_{i=1}^{N} \left(y_i(t) - \overline{y}(t) \right)^2}.$$

for $t = 1 \ldots T$.

3. Application

3.1. Model description

We consider a dynamic crop model running at a daily time step, called the Winter Wheat Dry Matter model (WWDM) (Makowski *et al.*, 2004; Monod *et al.*, 2006). It has two state variables, the above-ground winter wheat dry matter $U(t)$ and the leaf area index $LAI(t)$, with t the day number from sowing ($t=1$) to harvest ($t=223$). The state variable $U(t)$ is calculated on a daily basis in function of the cumulative degree-days $T(t)$ (over a basis of $0^o C$) and of the daily photosynthetically active radiation $PAR(t)$. The model equations are defined by

$$U(t+1)=U(t)+E_b E_{i\max}\left[1-e^{-K.LAI(t)}\right]PAR(t)+\varepsilon(t) \tag{15}$$

and

$$LAI(t)=L_{\max}\left(\frac{1}{1+e^{-A(T(t)-T_1)}}-e^{B(T(t)-T_2)}\right) \tag{16}$$

where $\varepsilon(t)$ is a random term with zero expectation representing the model error. In the sequel, we shall consider only the deterministic part of the model and so the latter term will be neglected. The dry matter at sowing ($t=1$) is set to zero : $U(1)=0$. In addition, the constraint $T_2=\frac{1}{B}\log[1+\exp(A\times T_1)]$ is applied, so that $LAI(1)=0$.

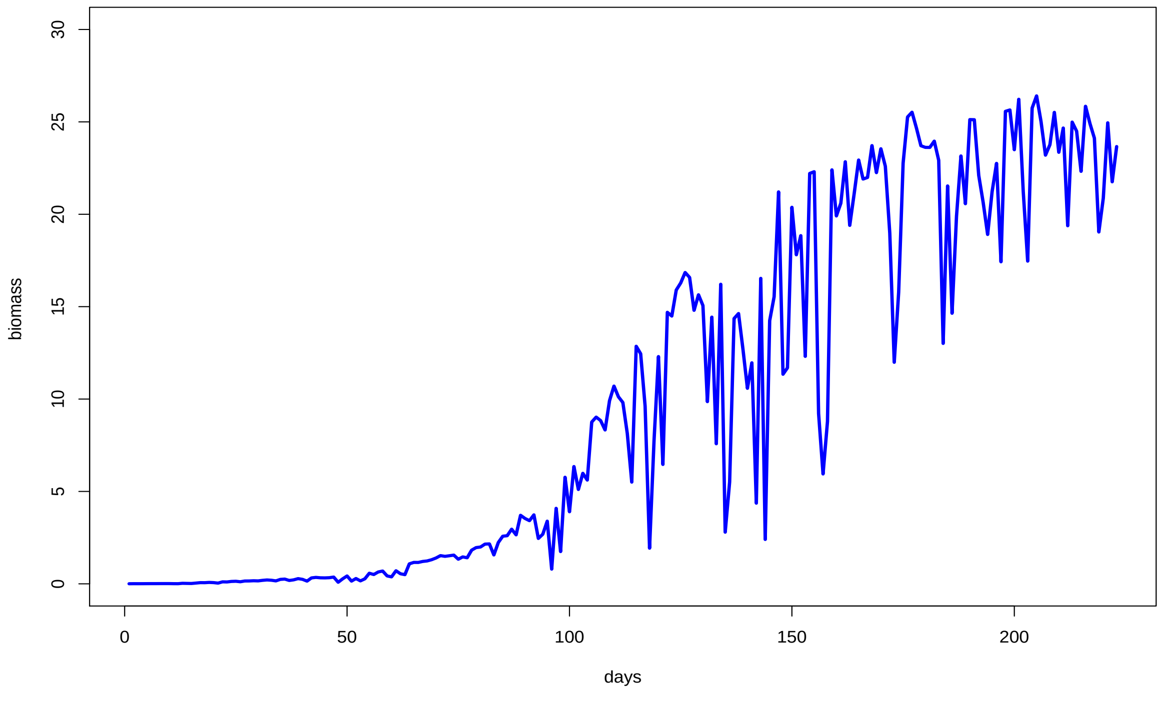

Figure 1. Daily simulated values of the dry matter increase $U(t)-U(t-1)$, using the nominal values of the WWDM model parameters.

Parameter	Interpretation	Nominal value	Uncertainty interval
E_b	radiation use efficiency	185	0.9-2.8
$E_{i\max}$	maximal ratio of intercepted to incident radiation	0.94	0.9-0.99
K	coefficient of extinction	0.7	0.6-0.8
$L_{\max}$	maximal value of *LAI*	7.5	3-12
T_1	temperature threshold	900	700-1100
A	-	0.0065	0.0035-0.01
B	-	0.00205	0.0011-0.0025

Table 3. Uncertainty intervals for the winter wheat dry matter model parameters.

There are seven free parameters, which are considered uncertain for the sensitivity analysis. Most of them have a meaningful interpretation. Uncertainty intervals in Table 3 were given by agronomists (Monod *et al.* 2006). Usually the climate should form one or several input factors for the sensitivity analysis. Here, preliminary investigations on 14 annual climate series showed little differences between years. For simplicity, results with a single series are presented.

The model output to be considered is the dynamic evolution of the dry matter $U(t)$ from sowing ($t=1$) until harvest ($t=223$). It is represented in Figure 1 for the nominal values of the parameters.

3. 2. Simulation design

A complete 3^7 factorial design (seven parameters at three levels) was constructed. With three levels, it is possible to estimate the linear and quadratic effects of quantitative factors. The three levels of each factor were the two bounds of the factor uncertainty interval and their mean. The number of simulations was $N=3^7=2187$.

3.3. Results

3.3.1. Principal components analysis

The PCA showed that 95% of the total inertia between the simulated dry matter dynamics was concentrated in the first three principal components, with $\lambda_1 = 0.54$, $\lambda_2 = 0.36$ and $\lambda_3 = 0.05$.

PCA results and sensitivity principal indices are presented in Figure 2, with one column of graphics per principal component k, for $k = 1,\ldots,3$. The top row displays the correlation of the principal component h_{k}, for $k = 1,\ldots,3$ with the outputs y_t, for $t = 1,\ldots,T$ as a function of day t, and the second row shows the first order and total sensitivity indices of the seven input factors on each PC.

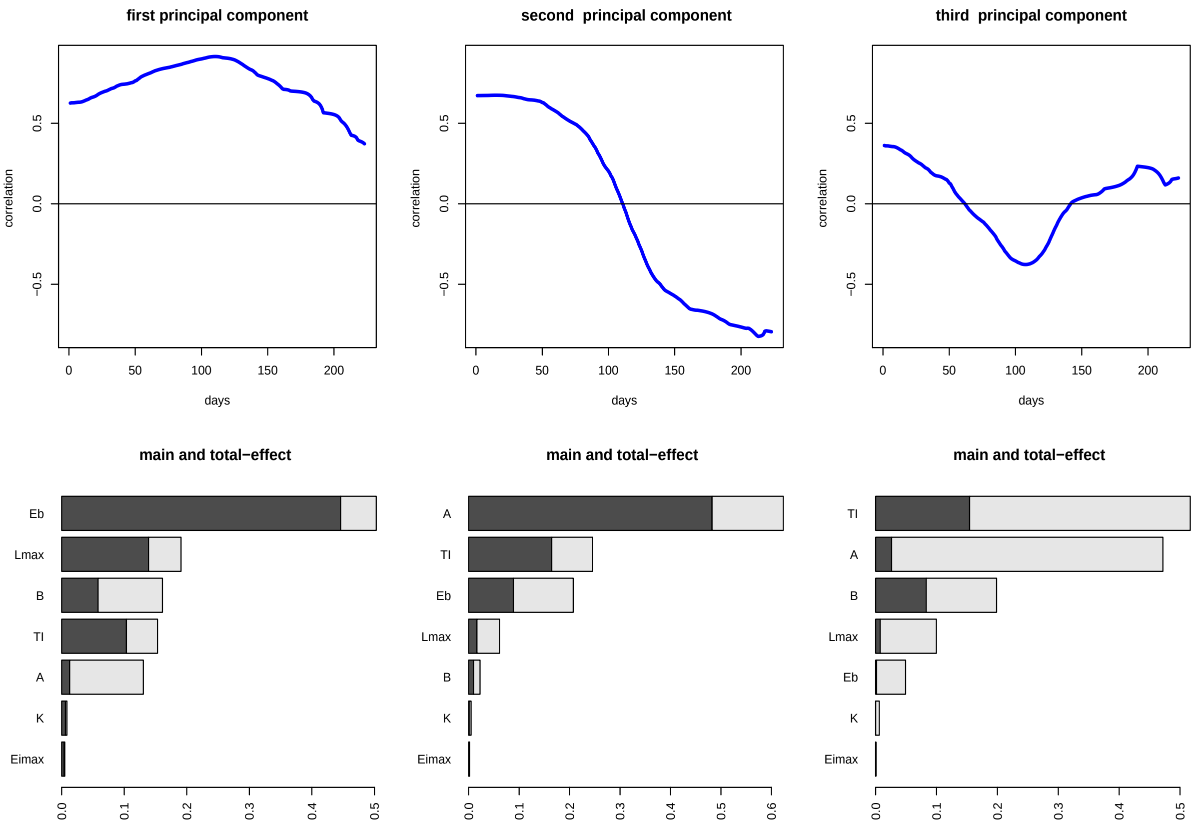

Figure 2. Correlations of the First three principal components with the output variables (top) and sensitivity indices on the first three principal components (bottom). The main sensitivity indices are in dark bars and interaction ones are in pale bars. The total length of any bar represents the total sensitivity index.

The first PC correlation with outputs had positive coordinates, with the largest ones in the middle of the time span. The first PC was thus associated with the global amount of dry matter all along the crop growth. It was mainly sensitive to parameter E_b. The second PC correlation was negative for half the time span, and then positive. Thus it opposed slow to fast growing dynamics relatively to the average behaviour. It was mainly sensitive to parameter A.

The third PC accounted for a much smaller part of inertia, associated with an opposition in the dry matter increase between the middle and both extremes of the dynamics. It was sensitive to the interaction between T_1 but other sensitivity indices were close. Clearly, the most three important parameters appeared to be A, E_b and T_i, but sensitivity principal indices were not sufficient to identify the most important among these three parameters.

3.3.2. *Generalized Sensitivity Index*

The Generalized sensitivity indices (GSI) are shown in Figure 3. GSI and the approximated $\tilde{GSI}$ with $P = 3$ lead to the same conclusions. Indeed, the rankings of factorial terms are the same. The three important parameters are E_b (the most important parameter), A (the second most important parameter) and T_1 (the third most important parameter). These results confirm the previous sensitivity analyses on PCs and bring new information by ranking the parameters. For instance, if one has to choose two parameters to calibrate, these parameters should be E_b and A.

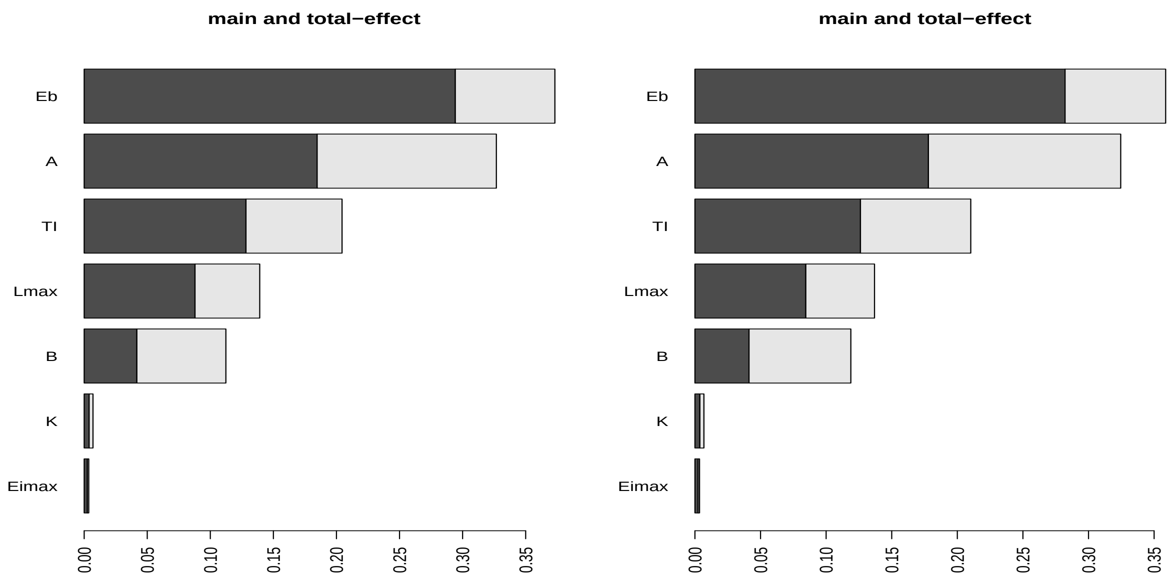

Figure 3. Generalized Sensitivity Indices GSI (right) and approximated indices $\tilde{GSI}$ (left) for the WWDM model. The main sensitivity indices are in dark bars and interaction ones are in pale bars. The total length of any bar represents the total sensitivity index.

The dynamic coefficients of determination are shown in Figure 4, for an ANOVA model with main effects and two-factor interactions and for various numbers of PCs. It is logical that the dynamic coefficients of determination increase with the number of PCs included in the approximation, but for this application, considering the first three PCs or all the PCs lead to the same results. The approximations are good ($\mathrm{GQ} \geq 0.80$) except for the end of the crop growth.

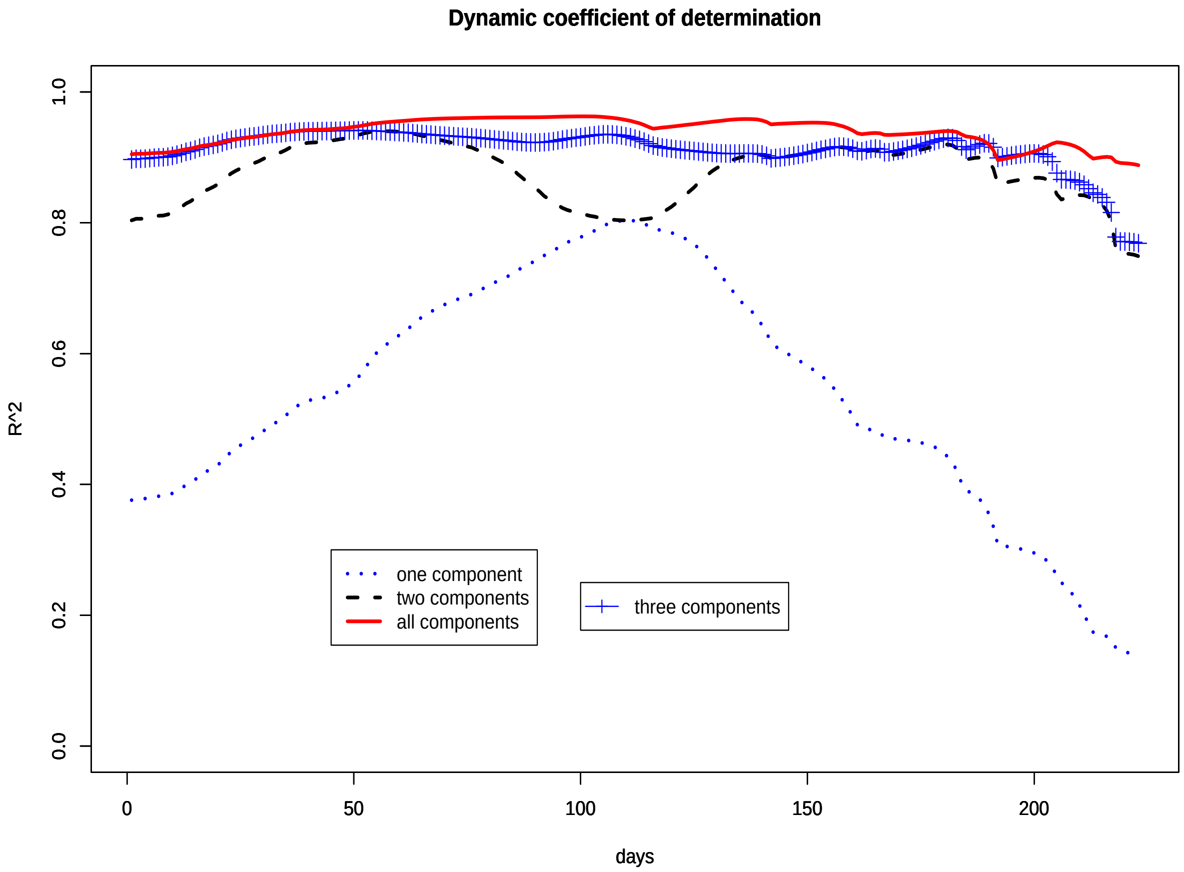

Figure 4. Dynamic coefficient of determination with the first one, the first two, the first three or with all principal components, and with main effects and two-factor interactions in the ANOVA models.

4. GSI R-Program

A R function was developed to compute the global sensitivity index defined in section 2. This function is named GSI and is available upon request. It is based on the free language R (www.cran-r.org) and can be easily used to implement our method with other dynamic models. The inputs of the function are: the experimental design including the levels of the model input factors, the output matrix including the model simulations corresponding to the experimental design, the maximal order of the interactions among factors, the number of principal component to be considered. The function proceeds in four steps:

(a) principal components analysis on the output matrix;
(b) ANOVA analyses on the principal components;
(c) computation of the sensitivity indices for each principal component;
(d) computation of the generalized sensitivity indices and goodness of fit assessment.

The outputs of the GSI function are: the values of the classical sensitivity index for each principal component (main and total effects), the generalized sensitivity index (main and total effects), the values of the dynamic coefficient of determination, and the global criterion. Results are also presented graphically as shown in figures 2 and 3.

5. Conclusion

In this paper, we proposed a generalized index to study the sensitivity of time series output to parameter values. This index is based on a principal component analysis and can be easily computed by using a set of model simulations derived from a factorial design. It can be used to identify the parameters of a discrete-time model that must be estimated from experimental data. The method considered in this study allows biologists to summarize the sensitivity of a time series output to several parameters and so, to determine which subset of parameters accounts for most of the output variance. Our approach could contribute to model simplification and a reduction of the number of experiments performed for estimating the parameters of discrete-time models.

In this paper, our method was applied to a model including seven parameters. The results showed that two parameters had a strong influence on the wheat biomass simulated by the model at a daily time step; the radiation use efficiency and a parameter of the mathematical function describing the kinetic of the leaf area index. We also showed that the index could be accurately computed by using only the first three principal components.

The method presented in this paper is quite general and can be applied to any dynamic model predicting one or several output variables at a discrete time step (e.g KLM model used in Campolongo *et al*. (2007), SWEB model described in Rizo *et al*. (2005) etc.). We provide a R-function for running any sensitivity analysis on discrete-time models. In this paper, the decomposition of the total inertia was based on principal component analysis and the simulations were computed from a factorial design. In the future, it will be interesting to extend the method. For example, principal components could be made more flexible by considering functional principal components (Ramsey and Silverman, 1997). Alternatively they could be replaced by Legendre polynomials (Campbell et al, 2006) and factorial designs by Monte Carlo simulations. It will be interesting to study the consequences of these alternatives on the results of the sensitivity analysis.

Acknowledgements: We are grateful to colleagues of the Mexico ("Méthodes pour l'EXploration Informatique des modèles COmplexes") network for their useful advices.

References

Anderson T.W., 2003. *An Introduction to Multivariate Statistical Analysis*, (3rd eds). Wiley, New York. , 721 pp.

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.). *Environmental modelling & software* **21**, 1042-1054.

Boote K.J., Jones J.W., Pickering N.B., 1996. Potential uses and limitations of crop models. *Agronomy Journal* **88**, 704-716.

Campbell K., McKay, M.D., Williams B.J., 2006. Sensitivity analysis when model outputs are functions. *Reliability Engineering and System Safety* **9**1, 1468-1472.

Campolongo F., Cariboni J., Saltelli A., 2007. An effective screening design for sensitivity analysis of large models. *Environmental modelling & software* **22**, 1509-1518.

Cariboni J., Galtelli D., Liska R., Saltelli A., 2007. The role of sensitivity analysis in ecological modelling. *Ecological Modelling* **203**, 167-182.

Ginot V., Gaba S., Beaudouin R., Aries F., Monod H., 2006. Combined use of local and ANOVA-based global sensitivity analyses for investigation of a stochastic dynamic model: Application to the case study of an individual–based model of a fish population. *Ecological Modelling* **193**, 479-491.

Homma T., Saltelli A., 1996. Importance measure in global sensitivity analysis of nonlinear models. *Reliability Engineering and System Safety* **52**, 1-17.

Jansen M.J.W, Rossing W.A.H, Daamen R.A, 1994. Monte carlo estimation of uncertainty contribution from several independent multivariate sources. *In: Predictability and Nonlinear Modelling in Natural Sciences and Economics* (Grasman J. and van Straten G. eds),. Kluwer, Dordrecht, pp.334-343.

Krzanowski W.J., Marriott F.H.C., 1990. *Multivariate Analysis. Part 1. Distribution, ordination, inference*. E. Arnold, London.

Makowski, D., Jeuffroy M.-H., Guérif M., 2004. Bayesian methods for updating crop model predictions, applications for predicting biomass and grain protein content. *In : Bayesian Statistics and quality modelling in the agro-food production chain* (Van Boekel *et al.* eds). Kluwer, Dordrecht. pp.57-68

Makowski, D., Hillier J., Wallach D., Andrieu B., Jeuffroy M.-H., 2006a. Parameter estimation for crop models. *In: Working with Dynamic Crop Models*. (Wallach D., Makowski D. and Jones J. eds). Elsevier, Amsterdam, pp. 101-150.

Makowski D., Naud C., Jeuffroy M.-H., Barbottin A., Monod H., 2006b. Global sensitivity analysis for calculating the contribution of genetic parameters to the variance of crop model predictions. *Reliability Engineering and System Safety* **91**, 1142-1147.

Matsuoka T., Seno H. Ecological balance in the native population dynamics may cause the paradox of pest control with harvesting. Journal of Theoretical Biology, in press.

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

Passioura J.R., 1996. Simulation models: science, snake oil, education or engineering. *Agronomy Journal* **88**, 690-694.

Ramsey I., Silverman B., 1997. *Functional Data Analysis* . Springer, New York, 311 pp.

Rizzo M. D., Mouser J. P., Whitney H. D., Mark D C., Magarey D. R., Voinoy A A. The comparison of four dynamic systems-based software packages: translation and sensitivity analysis. *Environmental modelling & software* **21**, 1491-1502.

Saltelli A., Chan K., Scott E.M. (eds.), 2000. *Sensitivity Analysis*. Wiley , New York, 475 pp.

Saltelli A., Tarantola A., Campolongo F., 2000. Sensitivity analysis as an ingredient of modelling. *Statistical Science* **15**, 377-395.

Saltelli A., Tarantola A., Campolongo F., Ratto M., 2004. *Sensitivity Analysis in Practice*. Wiley, New York.

Saltelli A., Ratto M., Tarantola A., Campolongo F., 2006. Sensitivity analysis practices: strategies for model based inference. *Reliability Engineering and System Safety* **91**,1109-1125.

Santangelo G., Bramanti L., Iannelli M. , 2007. Population dynamics and conservation biology of over-exploited Mediterranean red coral. Journal of Theoretical Biology 244, 416-423.

Saporta G., 2006. *Probabilité, Analyse des Données et Statistique*. Technip, Paris, 622 pp.

Wallach D., Goffinet B., Bergez J.E., Debaeke P., Leenhardt D., Aubertot J.N., 2001**.** Parameter estimation for crop models: a new approach and application to a corn model. *Agronomy Journal* **93**, 757-766.

Wu H-H., Lee H-J., Horng S-B., Berec L. , 2007. Modeling population dynamics of two cockroach species: effects of the circadian clock, interspecific competition and pest control. Journal of Theoretical Biology 249, 473-486.

Yakubu A-A., Castillo-Chavez C. 2002. Interplay between local dynamics and dispersal in discrete-time metapopulation models. Journal of Theoretical Biology 218, 273-288.