Reliability Engineering and System Safety 96 (2011) 450–459

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

# Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models

Matieyendou Lamboni [a], Hervé Monod [a,*], David Makowski [b]

[a] INRA, Unité MIA (UR341), F78352 Jouy en Josas Cedex, France
[b] INRA, UMR Agronomie INRA/AgroParisTech (UMR 211), BP 01, F78850 Thiverval-Grignon, France

ARTICLE INFO

*Article history:*
Received 16 October 2009
Received in revised form
29 September 2010
Accepted 8 December 2010
Available online 14 December 2010

*Keywords:*
Dynamic model
Factorial design
Latin hypercube sampling
Principal components analysis
RKHS
Sensitivity analysis
Sobol' decomposition

ABSTRACT

Many dynamic models are used for risk assessment and decision support in ecology and crop science. Such models generate time-dependent model predictions, with time either discretised or continuous. Their global sensitivity analysis is usually applied separately on each time output, but Campbell et al. (2006 [1]) advocated global sensitivity analyses on the expansion of the dynamics in a well-chosen functional basis. This paper focuses on the particular case when principal components analysis is combined with analysis of variance. In addition to the indices associated with the principal components, generalised sensitivity indices are proposed to synthesize the influence of each parameter on the whole time series output. Index definitions are given when the uncertainty on the input factors is either discrete or continuous and when the dynamic model is either discrete or functional. A general estimation algorithm is proposed, based on classical methods of global sensitivity analysis.

The method is applied to a dynamic wheat crop model with 13 uncertain parameters. Three methods of global sensitivity analysis are compared: the Sobol'–Saltelli method, the extended FAST method, and the fractional factorial design of resolution 6.



## 1. Introduction

Global sensitivity analysis is frequently applied to models with multivariate or functional output. For example, many dynamic models used for risk assessment and decision support in ecology and crop science generate time-dependent model predictions, with time being either discretised in a finite number of time steps or considered as continuous. In such situations, as mentioned by Campbell et al. [1], it may be unsufficiently informative to perform sensitivity analyses on each output separately or on a few context-specific scalar functions of the output. Indeed, it may be more interesting to apply sensitivity analysis to the multivariate output as a whole. Consequently, there is a need to define criteria and to develop methods specifically adapted to the sensitivity analysis of multivariate or functional outputs.

In particular, consider a model with dynamic output or multi-outputs $y(1),\ldots,y(T)$. Conducting separate sensitivity analyses on $y(1),\ldots,y(T)$ gives information on how the sensitivity of $y(t)$ evolves over time. This is interesting, but it leads to much redundancy because of the strong relationship between responses from one time step 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.

To improve relevance, sensitivity analysis can be applied to pre-defined scalar functions $h(y(1),\ldots,y(T))$ that have a useful interpretation. However, many functions of $y(1),\ldots,y(T)$ are potentially interesting to look at. A general and more sophisticated approach consists in modelling the output as a joint function of time and of the input variables and uncertain parameters. Several examples are illustrated in Chapter 7 of Fang et al. [2], based on spatio-temporal, functional or semiparametric modelling tools.

However there is also a need to apply data-driven methods that can identify the most interesting features in the $y(t)$ dynamics and perform sensitivity analyses on these features. Campbell et al. [1] proposed a simple and very useful approach to do so. It consists in (i) performing an orthogonal decomposition of the multivariate output, and (ii) applying sensitivity analysis to the most informative components individually. There is a large collection of available methods for the first step: it can be based either on a data driven method such as principal component analysis, or on the projections of output on a polynomial, spline, or Fourier basis defined by the user. The second step can also be performed by several different methods of sensitivity analysis, such as factorial design, FAST, or Sobol' and its most recent versions developed by Saltelli et al. [3,4].

The method proposed by Campbell et al. [1] allows to restrict attention to a few components rather than a whole dynamic. However, there is a need also for a synthetic criterion to summarise the sensitivity over the whole dynamic. This criterion must be adapted to discrete or continuous uncertainty distributions, whereas the examples in [1] are restricted to the first case. In this paper, we first show that there is a full

* Corresponding author. Tel.: +33 1 34 65 28 45.
*E-mail address:* herve.monod@jouy.inra.fr (H. Monod).

0951-8320/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ress.2010.12.002

"factorial by component" decomposition of the output variability or inertia, as illustrated in Lurette et al. [5] and Lamboni et al. [6]. Based on this decomposition, we propose a new synthetic sensitivity criterion for discrete factors first. We then extend this criterion to the cases when the input factors and output are continuous, and estimation methods are proposed and compared through simulations on a crop model.

Section 2 presents the general framework (Section 2.1), and then three special cases: (i) the number of model output variables is finite and a complete or fractional factorial design is used to explore the input domain (Section 2.2); (ii) the number of model output variables is finite but the input domain is continuous (Section 2.3); (iii) the model outputs are defined as a continuous function over time (Section 2.4). In Section 3, the methods are illustrated on a crop model with 13 parameters. The main results are discussed in Section 4.

## 2. Methodology

### 2.1. Framework

To perform sensitivity analysis, some uncertain parameters and input variables are selected for study, while the others are fixed at given nominal values. The selected parameters and input variables yield the $d$ input factors $Z_1$ to $Z_d$ of the sensitivity analysis. Let $\mathbf{z}=(z_1,z_2,\ldots,z_d)'$ denote a scenario, that is, a combination of the levels $z_j$ of the input factors $Z_j$, for $j$ in $1,\ldots,d$. The model of interest in this paper is

$$y(t)=f(\mathbf{z},t),\quad t\in\mathcal{T},\tag{1}$$

where $y(t)$ is the scalar output at time $t$ and, for simplicity, the function $f$ is assumed to be deterministic. The time domain $\mathcal{T}$ may be discrete or continuous, and both cases will be considered. A single model run is determined by its scenario $\mathbf{z}$. Its output is a vector $\mathbf{y}=(y(1),\ldots,y(T))'$ if $\mathcal{T}$ is discrete or more generally a function $\mathbf{y}(t)$, $t\in\mathcal{T}$.

### 2.2. Discrete-time model with discrete input factors

In this subsection, Model (1) is a discrete-time dynamic model, with $\mathcal{T}=\{1,2,\ldots,T\}$. Besides, the uncertainty domain is restricted to a discrete set $D_j$ of $n_j$ values of interest for each input factor $Z_j$. Thus the uncertainty domain of $\mathbf{z}$ is the complete factorial design, that is, the full set of scenarios $\Omega=D_1\times\cdots\times D_d$ of size $N=\prod_j n_j$.

The full set of output dynamics over the complete factorial design $\Omega$, forms the $N\times T$ matrix:

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

Each column $\mathbf{y}(t)$ in $\mathcal{Y}$ contains the values of the output variable at a given time $t$, for the full set of scenarios, while each row $\mathbf{y}_i$ of $\mathcal{Y}$ is an individual dynamic for a given scenario $\mathbf{z}$.

#### 2.2.1. Anova-based decomposition of variance

Consider first global sensitivity analysis for a univariate output. When the input factors are discrete, this is equivalent to analysis of variance (anova), a classical method in statistics [7].

In the full anova decomposition, the output variance is decomposed across $2^d$ factorial terms. Each factorial term $w$ is associated with a subset of factors, and thus it can be identified to the corresponding subset of $\{1,\ldots,d\}$. For example, the factorial effect denoted by $w=\emptyset$ corresponds to the general mean of the output. The subset $w=\{j\}$ denotes the main effect, or first order effect, of factor $Z_j$. The subset $w=\{j_1,j_2\}$ denotes the interaction between factors $Z_{j_1}$ and $Z_{j_2}$, a second order effect.

Let $\mathbf{h}$ in $\mathbb{R}^N$ denote a zero mean output vector across all $N$ scenarios. Because of the orthogonality properties of the complete factorial design, there is a unique decomposition of the Sum of Squares $SS(\mathbf{h})=\|\mathbf{h}\|^2$, where $\|\cdot\|^2$ denotes the quadratic norm:

$$SS(\mathbf{h})=\sum_{w,w\neq\emptyset} SS_w,\tag{2}$$

where $SS_w=\|\mathbf{S}_w\mathbf{h}\|^2$ denotes the anova sum of squares associated with the factorial term $w$ and $\mathbf{S}_w$ denotes the orthogonal projection matrix on the subspace $V_w$ associated to $w$. For more details and for the uniqueness of the anova decomposition, see Appendix A and Scheffé [8].

#### 2.2.2. Principal components analysis

Consider now the dynamic output $\mathbf{y}$. Principal components analysis (PCA) allows its expansion in a new basis, so that most information is concentrated in the first few components [9–12].

Let $\Sigma$ denote either the variance–covariance matrix or the correlation matrix of the columns of $\mathcal{Y}$. Thus

$$\Sigma=\frac{1}{N}\mathcal{Y}'_c\mathcal{Y}_c,$$

with $\mathcal{Y}_c$ the matrix obtained by centering and possibly normalising each column of $\mathcal{Y}$. The PCA decomposition is based on the $\Sigma$ expansion

$$\Sigma=\sum_{k=1}^{T}\lambda_k\mathbf{v}_k\mathbf{v}'_k,$$

with $\lambda_1\geq\cdots\geq\lambda_T$ the eigenvalues of $\Sigma$ and $\mathbf{v}_1,\ldots,\mathbf{v}_T$ a set of normalised and mutually orthogonal eigenvectors associated to these eigenvalues. For simplicity, we assume that $N\geq T$. Then the principal components (PCs) $\mathbf{h}_k$, for $k=1,2,\ldots,T$, are the mutually orthogonal linear combinations of the $\mathcal{Y}_c$ columns defined by $\mathbf{h}_k=\mathcal{Y}_c\mathbf{v}_k$ or, in matrix form, the columns of the $N\times T$ matrix $\mathcal{H}=\mathcal{Y}_c\mathcal{V}$, where $\mathcal{V}$ denotes the $T\times T$ matrix with $\mathbf{v}_k$ in column $k$. The PCs appear in the expansion of the output vectors in the basis defined by the eigenvectors $\mathbf{v}_k$, which reads

$$y_i(t)=\overline{y}(t)+\sum_{k=1}^{T}(\mathbf{h}_k)_i\mathbf{v}_k(t),$$

where $\overline{y}(t)$ is the mean of $\mathbf{y}(t)$.

The inertia of $\mathcal{Y}$ is defined by $\mathbb{I}=\text{trace}(\Sigma)$. If $\Sigma$ is a correlation matrix, then $\mathbb{I}=T$. Otherwise, the inertia measures the total dispersion or variability among the rows of $\mathcal{Y}$. The principal components satisfy $\|\mathbf{h}_k\|^2=\lambda_k$, and the eigenvalues satisfy $\sum_k\lambda_k=\text{trace}(\Sigma)=\mathbb{I}$. Thus, by construction, the principal component matrix $\mathcal{H}$ has the same total inertia as $\mathcal{Y}_c$, but it is mostly concentrated in its first columns.

#### 2.2.3. Sensitivity indices on the principal components

Sensitivity analysis (SA) can be applied to each principal component. By combining the anova and PCA decompositions, the following definitions generalise the univariate global sensitivity indices defined, e.g., in Chan et al. [13].

**Definition 1.** For Model (1) with discrete input factors and discrete-time output:

(i) the sensitivity index of factorial term $w$ ($w\neq\emptyset$) for the $k$th principal component is defined by

$$\text{SI}_{w,k}=\frac{SS_{w,k}}{\lambda_k}\quad\text{where } SS_{w,k}=\|\mathbf{S}_w\mathbf{h}_k\|^2;$$

(ii) the first order sensitivity index of $Z_j$ for the $k$th principal component corresponds to the main effect of $Z_j$ ($w=\{j\}$), and so is defined by $\text{FSI}_{Z_j,k}=\text{SI}_{\{j\},k}$;