Sensitivity analysis and uncertainty quantification for environmental models

Thomas Cartailler, Anaïs Guaus, Alexandre Janon, Hervé Monod, Clémentine Prieur, Nathalie Saint-Geours

SENSITIVITY ANALYSIS AND UNCERTAINTY QUANTIFICATION FOR ENVIRONMENTAL MODELS [*], [**], [***]

Thomas Cartailler[1], Anais Guaus[2,1], Alexandre Janon[3,5,7], Hervé Monod[4,7], Clémentine Prieur[5,7] and Nathalie Saint-Geours[6,7]

Abstract. Environmental models often involve complex dynamic and spatial inputs and outputs. This raises specific issues when performing uncertainty and sensitivity analyses (SA). Based on applications in flood risk assessment and agro-ecology, we present current research to adapt the methods of variance-based SA to such models. After recalling the basic principles, we propose a metamodelling approach of dynamic models based on a reduced-basis approximation of PDEs and we show how the error on the subsequent sensitivity indices can be quantified. We then present a mix of pragmatic and methodological solutions to perform the SA of a dynamic agro-climatic model with non standard input factors. SA is then applied to a flood risk model with spatially distributed inputs and outputs. Block sensitivity indices are defined and a precise relationship between these indices and their support size is established. Finally, we show how the whole support landscape and its key features can be incorporated in the SA of a spatial model.

[1] ITK, Montpellier

[2] INRA, UMR Eco&Sols, Montpellier

[3] Université Lyon 1, ISFA

[4] INRA, UR0341 MIA-Jouy, Jouy en Josas

[5] MOISE (INRIA Grenoble Rhône-Alpes / LJK Laboratoire Jean Kuntzmann)

[6] Irstea, UMR TETIS Montpellier

[7] GdR MASCOT-NUM (Méthodes d'Analyse Stochastique des Codes et Traitements Numériques)

Résumé. Les modèles environnementaux contiennent souvent des entrées-sorties complexes de par leur nature dynamique et spatiale, ce qui soulève des problèmes spécifiques pour leurs analyses d'incertitude et de sensibilité (AS). A partir d'applications en évaluation des risques d'inondation et en agro-écologie, nous présentons des recherches en cours pour adapter les méthodes d'AS à de tels modèles. Après un rappel des principes de base, nous proposons une approche de métamodélisation de modèles dynamiques basée sur une approximation par base réduite d'EDPs et nous montrons comment l'erreur sur les indices de sensibilité qui en découle peut être quantifiée. Nous présentons ensuite un cocktail de solutions pragmatiques et méthodologiques pour l'AS d'un modèle agro-climatique dynamique avec des facteurs d'entrée non standards. Puis l'AS est appliquée à un modèle de risque d'inondation avec des entrées-sorties spatialisées. On définit des indices de sensibilité par bloc et une relation précise est établie entre ces indices et leur taille de support. Enfin, nous montrons comment l'ensemble du paysage et de ses caractéristiques clés peut être incorporé à l'AS d'un modèle spatial.

Introduction

Environmental research and engineering require modelling approaches to integrate knowledge about complex systems. The models developed today may involve large numbers of equations, parameters and state variables. Beyond knowledge integration, they are expected to help compare scenarios, predict outcomes, evaluate risks or support decision and policy.

Model complexity and stakes of the associated projects make it compulsory to assess how uncertainty and variability in the inputs propagate to the outputs. These needs, which are shared by all domains of application, have given rise to a rich theoretical and applied literature on sensitivity analysis and uncertainty quantification (see [4, 27, 37]). In particular, it is now widely accepted that these methods must be applied globally, that is, by considering the different inputs simultaneously and through their whole potential ranges of variation. In most cases, global uncertainty and sensitivity analyses are beyond the scope of analytical approaches and so they rely on simulation-based approaches. These approaches raise challenging statistical issues about sampling and estimation in high-dimensional spaces.

In this paper, we give a flavour of some hot topics in this field of research when applied to environmental and agro-ecological models. After a short presentation of the basic notions (Section 1), we consider problems associated with the dynamic (Sections 2 and 3) and spatial (Sections 4 and 5) nature of most environmental models. The more theoretical Section 2 proposes an intrusive approach based on metamodelling to cope with dynamic models. Section 3 illustrates the use of two standard methods of sensitivity analysis (Morris and Sobol') on a model of water stress in grapevine, emphasizing typical issues of dynamic crop models. Based on the context of flood risk assessment, Section 4 presents innovative methods to adapt sensitivity and uncertainty analyses to spatial inputs and outputs. Section 5 considers the spatial case again, when non standard inputs such as geometrical features must be included in the input factors of a sensitivity analysis.

1. Basic principles

1.1. **Global sensitivity analysis, the variance-based approach**

When performing uncertainty or sensitivity analysis, we write the model under study as

$$Y = \mathcal{G}(X_1, \cdots, X_p), \tag{1}$$

where Y denotes the output variable of interest, $\mathcal{G}$ denotes the model equations and $X_1, \ldots, X_p$ denote the input factors. The input factors are considered as random variables, whose probability distribution reflects the uncertainty on their values. Unless mentioned explicitly, they are supposed to be independent and the output

Y is supposed to be scalar. The model $\mathcal{G}$ may be determistic or stochastic, but in both cases Y is a random variable dependent on $X_1, \ldots, X_p$. Uncertainty quantification aims at evaluating its probability distribution, while sensitivity analysis aims at quantifying the contribution of each input factor to that uncertainty.

The conditional expectation $\mathrm{E}(Y|X_i)$ is a random variable which gives the mean of Y over the distributions of the X_j ($j \neq i$), when X_i is fixed. It is the best approximation of Y which depends on X_i only. Its variance quantifies the influence of X_i on the dispersion of Y. The Sobol' sensitivity indices are obtained by normalising this variance. Thus, the Sobol' first order sensitivity index of input factor X_i is defined by

$$\mathrm{S}_{\{i\}} = \frac{\mathrm{Var}\big(\mathrm{E}(Y|X_i)\big)}{\mathrm{Var}(Y)}.$$

It belongs to the interval $[0, 1]$. More generally, one can define sensitivity indices of any order $r \in \{1, \ldots, p\}$, starting from the functional ANOVA decomposition. Indeed, assume that $\mathcal{G}$ is a real square integrable function defined on the unit hypercube $[0, 1]^p$ and that $\mathbf{X} = (X_1, \ldots, X_p)$ is a random vector with independent components uniformly distributed on $[0, 1]$. The choice of uniformly distributed components is not restrictive: using the inverse transformation method allows to consider any marginal distributions. Then for any $\mathfrak{u} \subseteq \{1, \ldots, p\}$, denote by $\mathbf{X}_{\mathfrak{u}}$ the random vector with components X_i, $i \in \mathfrak{u}$. The ANOVA decomposition (see [12] and [5]) then states that $Y = \mathcal{G}(\mathbf{X})$ can be uniquely decomposed into summands of increasing dimensions

$$\mathcal{G}(\mathbf{X}) = \sum_{\mathfrak{u} \subseteq \{1, \ldots, p\}} \mathcal{G}_{\mathfrak{u}}(\mathbf{X}_{\mathfrak{u}}) \tag{2}$$

where $\mathcal{G}_{\varnothing} = \mathbb{E}[Y]$ and the other components have mean zero and are mutually uncorrelated. The Sobol' index [41] of order $r = \mathrm{card}(\mathfrak{u})$ with respect to the combination of all the variables in $\mathfrak{u} \subseteq \{1, \ldots, p\}$ is then defined as

$$\mathrm{S}_{\mathfrak{u}} = \frac{\sigma_{\mathfrak{u}}^2}{\sigma^2} = \frac{\mathrm{Var}\big[\mathcal{G}_{\mathfrak{u}}(\mathbf{X}_{\mathfrak{u}})\big]}{\mathrm{Var}[Y]}. \tag{3}$$

The main effect of the ith factor is measured by $\mathrm{S}_{\{i\}}$, then for $i \neq j$, the interaction effect due to the ith and the jth factors, that cannot be explained by the sum of the individual effects of X_i and X_j is measured by $\mathrm{S}_{\{i,j\}}$, and so on (see [36]). For any $i \in \{1, \ldots, p\}$, we also define a total sensitivity index $\mathrm{S}_{\{i\}}^T$ [13] to express the overall output sensitivity to an input X_i by

$$\mathrm{S}_{\{i\}}^T = \sum_{\mathfrak{v} \subset \{1, \ldots p\} \text{ such that } i \in \mathfrak{v}} \mathrm{S}_{\mathfrak{v}}. \tag{4}$$

Notice that

$$\mathrm{S}_{\{i\}}^T = 1 - S_{\mathbf{X}^{-i}} \tag{5}$$

where $\mathbf{X}^{-i} = (X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_p)$. This formula will be shown of numerical importance later [13].

1.2. **An estimator for Sobol' indices**

In this subsection, we introduce the classical Monte Carlo estimator of Sobol' indices first introduced in [41]. One of the main advantage of this estimator is that it can be used even for functional or spatial inputs.

We first need some notation. Let $\mathfrak{u}$ be a non-empty subset of $\{1, \ldots p\}$. Let X_i^j, $i = 1, \ldots, 2p - |\mathfrak{u}|$, $j = 1, \ldots n$, with n the number of Monte Carlo replications, be independent random variables uniformly distributed on $[0, 1]$. We also define

$$\begin{aligned}
\mathbf{X}_{\mathfrak{u}}^j &= (X_1^j, \ldots, X_{|\mathfrak{u}|}^j) \\
\mathbf{X}_{\mathfrak{u}^c}^{j,1} &= (X_{|\mathfrak{u}|+1}^j, \ldots, X_p^j) \\
\mathbf{X}_{\mathfrak{u}^c}^{j,2} &= (X_{p+1}^j, \ldots, X_{2p-|\mathfrak{u}|}^j).
\end{aligned}$$

Finally, for $k = 1$ and 2, consider

$$Y_{\mathfrak{u}}^{j,k} = f(\mathbf{X}_{\mathfrak{u}}^j, \mathbf{X}_{\mathfrak{u}^c}^{j,k}) . \tag{6}$$

With this notation, we can define the estimator $\hat{S}_{\{i\},n}$ of $S_{\{i\}}$ (*resp.* $\hat{S}_{\{i\},n}^T$ of $S_{\{i\}}^T$) by

$$\hat{S}_{\mathfrak{u},n} = \frac{\dfrac{1}{n}\sum_{j=1}^{n} Y_{\mathfrak{u}}^{j,1} Y_{\mathfrak{u}}^{j,2} - \left(\dfrac{1}{n}\sum_{j=1}^{n} Y_{\mathfrak{u}}^{j,1}\right)\left(\dfrac{1}{n}\sum_{j=1}^{n} Y_{\mathfrak{u}}^{j,2}\right)}{\dfrac{1}{n}\sum_{j=1}^{n} \left(Y_{\mathfrak{u}}^{j,1}\right)^2 - \left(\dfrac{1}{n}\sum_{j=1}^{n} Y_{\mathfrak{u}}^{j,1}\right)^2} \tag{7}$$

with $\mathfrak{u} = \{i\}$ (*resp.* $\{1,\ldots,p\} \smallsetminus \{i\}$).

The asymptotic properties of $\hat{S}_{\mathfrak{u},n}$ have been recently studied in [14]. The i.i.d. hypothesis on the experimental design can also be relaxed, reducing the computational cost [43]. However, we focus in this paper on the estimator defined by (7), as it is commonly used by practitioners.

1.3. **Screening as a preliminary step**

When the number p of input variables is large (typically $p \gg 10$), estimating Sobol' indices quickly gets too expensive (in time or in storage capacity). In these cases, a preliminary step, which is less time consuming but also less informative, is screening. The aim of screening is to eliminate some input variables whose influence seems to be really negligible. We refer to [36] for a complete overview of tools for screening and we focus in this paper on the method of screening introduced in [28]. Morris [28] proposed an OAT (one-at-a-time) design, in which the impact of changing the values of each factor is evaluated in turn. However, the design introduced by Morris belongs to global sensitivity experiments, as it covers the entire space over which the factors may vary.

Without loss of generality (thanks to a rescaling), we introduce below the approach by Morris [28] for uniformly distributed input factors on the unit hypercube $[0,1]^p$. Let $l \in \mathbb{N}$, $l \geq 2$. The p-dimensional input vector $\mathbf{X}$ is discretized on a l-level grid $\mathbf{G} = \{0, \frac{1}{l-1}, \frac{2}{l-1}, \ldots, 1\}^p$. The choice of l is arbitrary and has an impact on the results (see [3] e.g. for more details on the choice of l). We then choose a predetermined multiple δ of $1/(l-1)$, and define the ith elementary effect for $1 \leq i \leq p$ by

$$d_i(\mathbf{x}) = \frac{\mathcal{G}(\mathbf{x} + \delta_{\mathbf{i}}) - \mathcal{G}(\mathbf{x})}{\delta} = \frac{\mathcal{G}(\mathbf{x_1}, \ldots, \mathbf{x_{i-1}}, \mathbf{x_i} + \delta, \mathbf{x_{i+1}}, \ldots, \mathbf{x_p}) - \mathcal{G}(\mathbf{x})}{\delta} \tag{8}$$

where $\mathbf{x}$ is any value in the grid $\mathbf{G}$ such that $\mathbf{x} + \delta_{\mathbf{i}}$ is still in the grid $\mathbf{G}$ (we say $\mathbf{x} \in \mathbf{G'}$), with $\delta_{\mathbf{i}}$ is a p-dimensional vector whose all components are null except the ith one which is equal to δ. Let $(\mathbf{x}^j)$, $j = 1, \ldots, r$ be an i.i.d. (independent and identically distributed) sample uniformly distributed on $\mathbf{G'}$. Starting from this sample of size r, Morris [28] introduces a strategy to compute rp elementary effects using only $r(p+1)$ evaluations of the model $\mathcal{G}$ (see [36], p. 69-71 for details). For this strategy, l is chosen even and $\delta = \frac{l}{2(l-1)}$, then each of the $l^{p-1}(p - \delta(l-1)) = \frac{l^p}{2}$ elementary effects for the ith input factor has equal probability of being selected. Define $d_i(\mathbf{x}^j)$, $i = 1, \ldots, p$, $j = 1, \ldots, r$ the elementary effects deduced from the initial sample $\mathbf{x}^j$, $j = 1, \ldots, r$. We then compute

$$\mu_i = \frac{1}{r}\sum_{j=1}^{r} |d_i(\mathbf{x}^j)|$$

and

$$\sigma_i = \sqrt{\frac{1}{r}\sum_{j=1}^{r} (d_i(\mathbf{x}^j) - \mu_i)^2} .$$

We know that if the model is linear in X_i and X_i does not interact with the X_k, $k \neq i$, then the elementary effect $d_i(\mathbf{x})$ does not depend on the position of $\mathbf{x}$ in the input space. Then σ_i is equal to zero. Morris' approach thus allows to detect effects which correspond either to non linearities or to interactions (or both), but it does

not allow to separate both effects. In our case we use it to eliminate from influential input variables the ones having

$$|\mu_i| \le c \,, \; |\mu_i| \le 2\frac{\sigma_i}{r}$$

where the threshold c is chosen by the model expert.

1.4. **Toy example**

In this section, we give the intuition of sensitivity indices on a toy example. As a toy example, consider that $p = 2$ and that

$$\mathcal{G}(X_1, X_2) = 2X_1 + X_2,$$

where X_1 and X_2 are independent and uniform on $[0; 1]$.

It is clear that $\mathrm{Var} Y = 5\mathrm{Var} X_1$ and that

$$\mathrm{E}(Y|X_1) = 2X_1 + \frac{1}{2},$$

so that $\mathrm{Var E}(Y|X_1) = 4\mathrm{Var} X_1$, and we have

$$\mathrm{S}_{\{1\}} = \frac{4}{5}.$$

Similarly, we have: $\mathrm{S}_{\{2\}} = \frac{1}{5}$. It is clear, from the formula which defines $\mathcal{G}$, that the uncertainty on X_1 has more impact than the uncertainty on X_2, provided they vary in the same range.

Now, if we suppose that uncertainty on X_1 has been reduced, and that X_1 is uniform on $[0; \frac{1}{2}]$ (while X_2 is still uniform on $[0; 1]$), we get that $S_1 = \frac{1}{2} = S_2$. As $\frac{1}{2} < \frac{4}{5}$, we see that the Sobol' importance measure is consistent with the intuitive fact that the reduction on the uncertainty on a parameter reduces its impact on the uncertainty of the output (and increases, as a result, the impact of the uncertainty of the other parameters).

The next two sections (Sections 2 and 3) deal with specific issues for sensitivity analysis of dynamic models. Section 2 can be read independently from Sections 3, 4 and 5. It is the most theoretical section. In Sections 4 and 5, we are faced with spatial inputs and outputs. These two sections can be read independently from Sections 2 and 3 .

2. Reduced basis metamodels for sensitivity analysis of a dynamic model

Estimation of Sobol' indices using the estimates $\hat{\mathrm{S}}_{\mathfrak{u},n}$ defined at (7) requires a large number of model runs. More precisely, for a given index subset $\mathfrak{u}$, it requires n realisations of the $Y_\mathfrak{u}^1$ and $Y_\mathfrak{u}^2$ random variables. Using equation (6), these realisations are obtained by performing $2n$ calls to the $\mathcal{G}$ function.

The Monte-Carlo sample size n is generally large (typically, in thousands or tens of thousands), so as to obtain a reliable estimation of $\mathrm{S}_\mathfrak{u}$. This causes no problem if the $\mathcal{G}$ function can be rapidly evaluated. However, in many cases, the evaluation of this function, for a single value of the parameter, can take a long time. For instance, evaluation of $\mathcal{G}$ may involve a numerical solving of a partial differential equation (PDE) on a very fine grid (hence, resolution of large systems of equations), or a numerical optimization of a CPU-intensive function, or a large-sample Monte-Carlo simulation. In these cases, it is impossible, or at least not practical, to perform the sensitivity analysis on the original $\mathcal{G}$ function, as this analysis would require many expensive evaluations of this function.

One solution is to find a function $\widetilde{\mathcal{G}} : [0; 1]^p \to \mathbb{R}$, which approximates the original model function $\mathcal{G}$, but which can be computed faster than $\mathcal{G}$. The $\widetilde{\mathcal{G}}$ function is usually called *metamodel* (or *surrogate model*, or *response surface*). The idea is the following: instead of performing $2n$ calls to $\mathcal{G}$, we begin by building the metamodel $\widetilde{\mathcal{G}}$ (possibly using a small number of calls to $\mathcal{G}$), during an *offline phase* which is performed only once. Then, in an *online phase*, the metamodel is called $2n$ times, in order to produce an approximation of the

required evaluations. This methodology can be used in many contexts (not only in sensitivity analysis), where many evaluations of a costly function are to be performed quickly (this is called a *many-query* context), such as uncertainty quantification, model calibration, optimal design, or real-time simulations.

In the sequel of this section, we give a brief overview of the existing metamodelling methods (Section 2.1), then we focus on a specific method: the reduced basis method, which will be presented, in the linear case, in Section 2.2. The reduced basis (RB) approach is particularly well suited in the context of sensitivity analysis and uncertainty analysis in geophysical models. Such models typically involve a wide range of parameters, such as: source terms (climatic forcings, heat/wind/matter fluxes), boundary conditions (forcings, open boundaries), and the initial state of the system. Their study generally leads to parametrized partial differential equations which involve poorly-known parameters. It is then important to measure the impact of a given parameter on the quality of the solution, and to identify the most influent parameters on the model solution.The RB method has moreover the advantage of giving an explicitly computable bound on the metamodel error (ie., the absolute difference between $\mathcal{G}$ and $\widetilde{\mathcal{G}}$) which, in turn, can be used to provide an error bound on the Sobol' indices estimated with the metamodel. This error quantification will be presented in Section 2.3.

2.1. **Overview of metamodelling methods**

We distinguish two families of metamodelling methods: the *non-intrusive* metamodels and the *intrusive* metamodels. They differ by the type of information considered during the construction of the metamodel.

2.1.1. *Non-intrusive metamodels*

The *non-intrusive* methods use a finite sample of $k \in \mathbb{N}^*$ input-output pairs

$$\mathcal{L} = \left\{ \left(\mathbf{X^1}, \mathcal{G}(\mathbf{X^1}) \right), \ldots, \left(\mathbf{X^k}, \mathcal{G}(\mathbf{X^k}) \right) \right\},$$

where the *design points* $\mathbf{X^1}, \ldots, \mathbf{X^k}$ are generally chosen according to a space-filling design, such as variants of latin hypercube sampling (LHS).

As the computation of the *learning sample* $\mathcal{L}$ requires k calls to the model function, k has to be kept moderate, at most $2n$, for the metamodel approach to be pertinent in terms of computation times.

The metamodel function $\mathcal{G}$ is then an interpolator (in the *RKHS interpolation* method, also known as *Kriging* [38], [30] §3.1, [20]), or a regressor (in the *non-intrusive polynomial chaos* method [42]) of the elements in the $\mathcal{L}$ set.

The $\mathcal{L}$ sample and the interpolation/regression coefficients are computed (and stored) during the offline phase. In the online phase, the fast evaluation of $\widetilde{\mathcal{G}}(\mathbf{X})$ for a given $\mathbf{X}$ is possible using the interpolation or regression formula.

2.1.2. *Intrusive metamodels*

Intrusive metamodelling methods make use not only of a finite learning sample, as non-intrusive methods do, but also of the mathematical (usually implicit) definition of the model function $\mathcal{G}$. Let's consider in the following of this section the particular case where the model output writes as a function of the solution of a $\mathbf{X}$-parametrized PDE (that is, the PDE is driven by uncertain parameters which are taken as the inputs of this model)

$$\mathcal{G}(\mathbf{X}) = \ell(u(\mathbf{X})),$$

where $u(\mathbf{X})$ is the solution of a (discretized) $\mathbf{X}$-dependent PDE

$$\mathcal{D}(\mathbf{X}, u(\mathbf{X})) = f(\mathbf{X}) \tag{9}$$

for a known $f(\mathbf{X})$ (discrete) function, and a (discrete) differential operator $\mathcal{D}(\mathbf{X}, \cdot)$.

For such models, it is possible to construct an intrusive metamodel. In the sequel, we present the reduced basis approach, in the particular case of a linear PDE which depends "affinely" on the parameter vector, as given in [29]. This method has been extended to more general (nonlinear and/or time-dependent) cases [9,21,44]. More

specifically, it has been applied to the Burgers' equation [15], a 1D version of the Shallow-Water equations, which are commonly used to model the behavior of geophysical fluids (ocean, rivers). Thus it has to be considered as a necessary first step to well understand what to do for more (realistic) complex models.

2.2. **The reduced basis method for linear affinely-parametrized equations**

2.2.1. *Hypotheses.*

We suppose that the PDE given in (9) is linear, and can be written in the following discrete weak form

$$a(\mathbf{X}; u(\mathbf{X}), v) = g(v) \quad \forall v \in \mathcal{F}, \tag{10}$$

where $\mathcal{F}$ is a finite-dimensional vector space (usually a finite element space), and $a(\mathbf{X}; \cdot, \cdot)$ and $g(\cdot)$ are, respectively, a bilinear form and a linear form on $\mathcal{F}$.

The most computationally-demanding step during the evaluation of $\mathcal{G}(\mathbf{X})$ is generally the numerical resolution of Equation (10). This resolution is costly because $\mathcal{F}$ has to be of large dimension for the PDE discretization to be accurate. The goal of the reduced basis is to approximate (10) by a system of equations which can be solved faster, replacing $\mathcal{F}$ by a specific space of smaller dimension.

We further suppose that $a(\cdot; u, v)$ can be affinely-parametrized, i.e. that

$$a(\mathbf{X}; u, v) = \sum_{q=1}^{Q} \Theta_q(\mathbf{X}) a_q(u, v) \quad \forall \mathbf{X}, \forall u, v \in \mathcal{F},$$

where Θ_q $(q = 1, \ldots, Q)$ are any functions of the parameter vector.

2.2.2. *Offline and online procedures.*

In the reduced basis method, one begins by choosing (during the offline phase) a parameter-independent, *problem-adapted*, linear independent family $\{\zeta_1, \ldots, \zeta_m\}$ of vectors of $\mathcal{F}$. This family is called the *reduced basis* and we define $\widetilde{\mathcal{F}} := \mathrm{Vect}(\zeta_1, \ldots, \zeta_m)$. Then, the following quantities

$$a_q(\zeta_i, \zeta_j) \quad (q = 1, \ldots, Q, \ i, j = 1, \ldots, m)$$

and

$$g(\zeta_i) \quad (i = 1, \ldots, m)$$

can be computed and stored. As they are parameter-independent, this can be done during the offline phase.

Finally, during the online phase, given a parameter vector $\mathbf{X}$, the system of n equations

$$\sum_{i=1}^{m} \left(\sum_{q=1}^{Q} \Theta_q(\mathbf{X}) a_q(\zeta_i, \zeta_j) \right) \widetilde{u}_i(\mathbf{X}) = g(\zeta_j) \quad \forall j = 1, \ldots, m,$$

obtained by the Galerkin projection of (10) onto $\widetilde{\mathcal{F}}$ can be solved in $\widetilde{u}_1, \ldots, \widetilde{u}_m$ so as to find

$$\widetilde{u}(\mathbf{X}) = \sum_{i=1}^{m} \widetilde{u}_i(\mathbf{X}) \zeta_i,$$

which is an approximation of u, solution of (10).

A significant computational gain can be realized if one can find a reduced basis $\{\zeta_1, \ldots, \zeta_m\}$ such that $m \ll \dim \mathcal{F}$ and, at the same time, $u(\mathbf{X}) \approx \widetilde{u}(\mathbf{X})$ for all $\mathbf{X}$. We will see, in Section 2.2.4, two methods to find such a basis.

Once the $\widetilde{u}_i(\mathbf{X})$ coefficients have been found, one may approximate, in the case of a linear output defined by ℓ, the model $\mathcal{G}(\mathbf{X})$ by the *reduced model*

$$\widetilde{\mathcal{G}}(\mathbf{X}) = \ell(\widetilde{u}(\mathbf{X})) = \sum_{i=1}^{m} \widetilde{u}_i(\mathbf{X})\ell(\zeta_i). \tag{11}$$

This quantity can be computed with an online phase of complexity *depending only on m* and Q, and not on $\dim \mathcal{F}$.

2.2.3. *Error quantification.*

The reduced basis method comes with an offline/online procedure for computing an *error bound ϵ_u*, satisfying

$$\|u(\mathbf{X}) - \widetilde{u}(\mathbf{X})\| \leq \epsilon(\mathbf{X}) := \epsilon_u(\mathbf{X}),$$

for a suitably chosen norm $\|\cdot\|$ on $\mathcal{F}$.

The online procedure for the computation of ϵ_u also has a $\dim \mathcal{F}$-independent complexity. One can refer to [29] for details of the computation of ϵ_u.

The ϵ_u can be used so as to give a bound between true and reduced outputs

$$\left|\mathcal{G}(\mathbf{X}) - \widetilde{\mathcal{G}}(\mathbf{X})\right| \leq L\epsilon_u(\mathbf{X}),$$

where L is the Lipschitz constant of ℓ

$$L = \sup_{v \in \mathcal{F}, v \neq 0} \frac{\ell(v)}{\|v\|}.$$

One should note that a much more precise bound for $\left|\mathcal{G}(\mathbf{X}) - \widetilde{\mathcal{G}}(\mathbf{X})\right|$ has been proposed in [17].

2.2.4. *Reduced basis choice.*

We now present two methods for choosing the reduced basis $\{\zeta_1, \ldots, \zeta_m\}$.

POD method.   The first is the POD (proper orthogonal decomposition) method. The idea is to look for an orthonormal family $\{\zeta_1, \ldots, \zeta_m\}$ which minimizes the mean error

$$\int_{\mathbf{X} \in [0;1]^p} \|u(\mathbf{X}) - \widetilde{u}(\mathbf{X})\|^2 \, d\mathbf{X},$$

where $\widetilde{u}(\mathbf{X})$ denotes the orthogonal projection of $u(\mathbf{X})$ on $\mathrm{Vect}\,(\zeta_1, \ldots, \zeta_m)$.

As the integral cannot generally be carried out analytically, one instead minimizes an empirical approximation

$$\sum_{\mathbf{X} \in \Xi} \|u(\mathbf{X}) - \widetilde{u}(\mathbf{X})\|^2,$$

where Ξ is a (usually random) finite subset of the parameter space $[0;1]^p$.

The *snapshot POD* algorithm [39] performs this constrained minimization by computing $u(\mathbf{X})$ for $\mathbf{X} \in \Xi$, and then finding the reduced basis in terms of the m dominant eigenvectors of a matrix of order cardinal of Ξ.

Greedy method.   The *greedy* method [29] iteratively finds a reduced space of the form

$$\widetilde{\mathcal{F}} = \mathrm{Vect}\,(u(\mathbf{X}_k),\ k = 1, \ldots, m).$$

Hence, it is sufficient to define $\mathbf{X}_1, \ldots, \mathbf{X}_m$. To begin with, we take for $\mathbf{X}_1$ a randomly chosen parameter.

At the kth step, suppose that $\mathbf{X}_1, \ldots, \mathbf{X}_{k-1}$ have been found. One takes

$$\mathbf{X}_k = \underset{\mathbf{X} \in \Xi}{\mathrm{argmin}}\, \epsilon_u(\mathbf{X}),$$

where Ξ is a random, finite, large subset of the parameter space, and ϵ_u is the error bound given by the reduced basis method obtained by taking $\{u(\mathbf{X}_1), \ldots, u(\mathbf{X}_{k-1})\}$ as reduced basis (in fact, an orthonormalized family which spans the same subspace, for numerical stability reasons).

The advantage of the greedy method over the POD method is that it requires computation of $u(\mathbf{X})$ for only m (ie., the desired size of the reduced basis) values of $\mathbf{X}$, while the POD requires generally more, which causes a greater offline computational time. However, the POD method does not depend on the existence and the accuracy of an error bound, and have shown to have better mean performance in our experiments.

2.3. **Error quantification for Sobol' indices estimated using a metamodel**

In this Section, we suppose we have a metamodel output $\widetilde{\mathcal{G}}$, which is used as an approximation of the original model output $\mathcal{G}$.

We suppose that the quality of this approximation is asserted by a computable pointwise error bound ϵ,

$$\left| \mathcal{G}(\mathbf{X}) - \widetilde{\mathcal{G}}(\mathbf{X}) \right| \le \epsilon(\mathbf{X}).$$

This error bound can be the one given by the reduced basis method (see above), or any other (possibly heuristic) error estimator.

As said in the introduction, we suppose that the evaluation of $\mathcal{G}$ is costly while evaluation of $\widetilde{\mathcal{G}}$ and ϵ are cheap. Hence, we would like to find two quantities $\hat{S}_{u,n}^{\min}$ and $\hat{S}_{u,n}^{\mathrm{Max}}$ so that

$$\hat{S}_{u,n}^{\min} \le \hat{S}_{u,n} \le \hat{S}_{u,n}^{\mathrm{Max}},$$

where $\hat{S}_{u,n}^{\min}$ and $\hat{S}_{u,n}^{\mathrm{Max}}$ can be computed using evaluations of $\widetilde{\mathcal{G}}$ and ϵ.

We define, using the notations of Section 1.2,

$$\tilde{Y}_u^{j,k} = \widetilde{\mathcal{G}}(\mathbf{X}_u^j, \mathbf{X}_{u^c}^{j,k}), \quad \epsilon_u^{j,k} = \epsilon(\mathbf{X}_u^j, \mathbf{X}_{u^c}^{j,k}), \quad I_u^{j,k} = \left[\tilde{Y}_u^{j,k} - \epsilon_u^{j,k}; \tilde{Y}_u^{j,k} + \epsilon_u^{j,k} \right], \tag{12}$$

for $k = 1, 2$.

It is clear that since $Y_u^{j,k} \in I_u^{j,k}$, for all $j = 1, \ldots, n$ and $k = 1, 2$, one can take

$$\hat{S}_{u,n}^{\min} = \min_{y_{j,k} \in I_u^{j,k} \, \forall j,k} \frac{\frac{1}{n} \sum_{j=1}^n y_{j,1} y_{j,2} - \left(\frac{1}{n} \sum_{j=1}^n y_{j,1} \right) \left(\frac{1}{n} \sum_{j=1}^n y_{j,2} \right)}{\frac{1}{n} \sum_{j=1}^n y_{j,1}^2 - \left(\frac{1}{n} \sum_{j=1}^n y_{j,1} \right)^2}$$

and

$$\hat{S}_{u,n}^{\mathrm{Max}} = \max_{y_{j,k} \in I_u^{j,k} \, \forall j,k} \frac{\frac{1}{n} \sum_{j=1}^n y_{j,1} y_{j,2} - \left(\frac{1}{n} \sum_{j=1}^n y_{j,1} \right) \left(\frac{1}{n} \sum_{j=1}^n y_{j,2} \right)}{\frac{1}{n} \sum_{j=1}^n y_{j,1}^2 - \left(\frac{1}{n} \sum_{j=1}^n y_{j,1} \right)^2}.$$

In [16], we have proposed a fast, analytical method for computing $\hat{S}_{u,n}^{\min}$ and $\hat{S}_{u,n}^{\mathrm{Max}}$. The spread between $\hat{S}_{u,n}^{\min}$ and $\hat{S}_{u,n}^{\mathrm{Max}}$ accounts for the error induced by the use of the metamodel to estimate the Sobol' sensitivity index S_u.

If one wants to assess the uncertainty (when estimating Sobol' indices) due to the replacement of the model by the metamodel, but also the one due to Monte Carlo estimation, we have to go further.

The Monte-Carlo error can be quantified by using a bootstrap technique [6]. More specifically, notice that $\hat{S}_{u,n}^{\min}$ and $\hat{S}_{u,n}^{\mathrm{Max}}$ are functions of the data

$$\hat{S}_{u,n}^{\min} = \hat{S}_{u,n}^{\min} \left(\left(\tilde{Y}_u^{j,1}, \epsilon_u^{j,1}, \tilde{Y}_u^{j,2}, \epsilon_u^{j,2} \right)_{j=1,\ldots,n} \right), \quad \hat{S}_{u,n}^{\mathrm{Max}} = \hat{S}_{u,n}^{\mathrm{Max}} \left(\left(\tilde{Y}_u^{j,1}, \epsilon_u^{j,1}, \tilde{Y}_u^{j,2}, \epsilon_u^{j,2} \right)_{j=1,\ldots,n} \right).$$

Now, for a list L drawn with replacement from $\{1, \dots, n\}$, we define the bootstrap replications of $\hat{\mathrm{S}}_{\mathfrak{u},n}^{\min}$ and $\hat{\mathrm{S}}_{\mathfrak{u},n}^{\mathrm{Max}}$,

$$\widehat{S}_b^{\min}(L) = \hat{\mathrm{S}}_{\mathfrak{u},n}^{\min}\left(\left(\tilde{Y}_{\mathfrak{u}}^{j,1}, \epsilon_{\mathfrak{u}}^{j,1}, \tilde{Y}_{\mathfrak{u}}^{j,2}, \epsilon_{\mathfrak{u}}^{j,2}\right)_{j \in L}\right), \quad \widehat{S}_b^{\mathrm{Max}}(L) = \hat{\mathrm{S}}_{\mathfrak{u},n}^{\mathrm{Max}}\left(\left(\tilde{Y}_{\mathfrak{u}}^{j,1}, \epsilon_{\mathfrak{u}}^{j,1}, \tilde{Y}_{\mathfrak{u}}^{j,2}, \epsilon_{\mathfrak{u}}^{j,2}\right)_{j \in L}\right).$$

We can compute, a (large) collection of B replications of $\hat{\mathrm{S}}_{\mathfrak{u},n}^{\min}$ and $\hat{\mathrm{S}}_{\mathfrak{u},n}^{\mathrm{Max}}$,

$$\mathcal{R}^{\min} = \{\widehat{S}_b^{\min}(L_1), \dots, \widehat{S}_b^{\min}(L_B)\}, \quad \mathcal{R}^{\mathrm{Max}} = \{\widehat{S}_b^{\mathrm{Max}}(L_1), \dots, \widehat{S}_b^{\mathrm{Max}}(L_B)\}.$$

with $L_1, \dots, L_B$ independent lists drawn with replacement from $\{1, \dots, n\}$. Due to the resampling nature of the bootstrap, this computation does not require any supplementary evaluation of the metamodel or its error bound.

These replications can be used to produce a confidence interval for $\mathrm{S}_{\mathfrak{u}}$ as follows: for a desired risk level $\alpha > 0$, we can consider $q^{\min}$ and q^{Max}, respectively the $\alpha/2$-quantile of $\mathcal{R}^{\min}$ and the $1 - \alpha/2$-quantile of $\mathcal{R}^{\mathrm{Max}}$,

$$\#\{x \in \mathcal{R}^{\min} \text{ s.t. } x \leq q^{\min}\} = \frac{\alpha}{2} B,$$

$$\#\{x \in \mathcal{R}^{\mathrm{Max}} \text{ s.t. } x \leq q^{\mathrm{Max}}\} = \left(1 - \frac{\alpha}{2}\right) B,$$

and take $[q^{\min}; q^{\mathrm{Max}}]$ as confidence interval for $\mathrm{S}_{\mathfrak{u}}$.

This confidence interval combines the two sources (metamodel and estimation) of errors, and it can be shown that it includes the bootstrap confidence interval based on the estimator on the true model output.

In the three next sections, we are interested with spatio and/or temporal environmental models involving non standard inputs (weather scenarios, geometrical features, ...) and/or outputs (spatial outputs). In these sections, contrarily to the RB approach of Section 2, we consider the model as a black box. For a set of input parameters, one can compute the corresponding output, but one ignores the intern structure of the model.

3. Sensitivity and uncertainty analyses of a dynamic model with non standard input factors

This section describes the global sensitivity analysis (SA) of an agro-climatic model embedded in a decision support system (DSS) for the water management of vineyard in the Languedoc-Roussillon region, France. The DSS is used in real time to recommend irrigation amounts in order to maintain optimal vine water stress dynamics, based on the production objective targeted by the winegrower (table wine, aging or laying-down wine, etc.). A major characteristic of agro-climatic models is the difficulty of estimating the numerous input parameters because field measurements are both costly and tedious. This is particularly true when soil-related parameters are involved - which is the case here - because their estimation requires subsoil measurements. The operational use of the model thus requires to find the right balance between data-friendliness and precision [33]: the less input parameters asked to the end-user, the better.

In this context, the practical use of the SA is to identify the most influent parameters in order either to concentrate experimental efforts on their field measurement when possible, or to calibrate them otherwise. After a preliminary screening of the less influential parameters via Morris method, the method by Sobol' is used to achieve the SA. The model includes a functional (temporal) output and correlated temporal inputs, so the study addresses some issues arising from these two aspects when met in a practical context.

3.1. **A model of water stress in grapevine**

The model output is the predawn leaf water potential (PLWP); it measures in MPa the water status in vine leaves. It is the reference indicator of vine water stress [34], which describes the physiological state of the vine under water deficit. It is estimated using a discrete-time mechanistic soil-vegetation-atmosphere-transfer model [23, 31] that is, through the combination of models describing the dynamics of soil water balance, vine canopy growth and solar radiation absorbed by the vegetation. The soil water balance model is again the combination of sub-models describing the various soil water transfer processes: rain and irrigation infiltration, bare soil evaporation, vine root absorption, runoff and drainage. The soil water balance model runs daily starting January 1st, and PLWP is computed daily during the vine vegetative cycle, i.e. when leaves are present, typically from April till October.

The model requires the definition of 4 temporal and 21 scalar inputs (see Table 1 for a brief definition). The 4 temporal inputs are to some extent correlated. They represent the weather data station measurements necessary for driving the model: daily precipitations, solar radiation, mean air temperature and potential evapotranspiration. For most of the scalar inputs, like vegetation dimensions or vine rows orientation, the winegrower can easily give at least a rough estimate. However, this is not the case for 4 scalar inputs that are the parameters of the runoff (CN) and evaporation (b_1 and U) models and the total transpirable soil water (TTSW) that controls the maximum amount of soil-water available to the vine. The runoff parameter CN controls the amount of water that actually infiltrates the soil during a rain event. Parameters U and b_1 characterize the influence of the top soil layer on the evaporative processes. All 3 are known to be related to other more easily measured soil properties, mainly soil texture in the top soil layer and soil surface feature (presence of grass, chemical weeding, etc.), but no consensus exists in literature on the precise relations. Here they are estimated from literature, based on relations with the mean soil texture only, because it is the parameter that the end-user can the most easily estimate. Similarly, the TTSW is estimated based on soil texture and an estimate of the rooting depth. One aim of the SA is to evaluate the choice made for the estimation of these parameters: it might be suitable to spend more effort to reduce the uncertainty on the estimation of the most influent parameters. In the following, the 21 scalar input parameters will be assumed to be independent, which was considered as reasonable by practitioners.

3.2. **Definition of the input factors**

The SA was achieved at the Languedoc-Roussillon scale in order to gain a general insight into the model behaviour. The distributions of all scalar parameters were set rather easily from literature or field expertise in order to scan their whole variation range at the regional scale. The case of the weather data was more an issue, since to the authors' knowledge, the introduction of correlated functional inputs in a SA is still under research. We chose to use an equivalent of the map-labeling method developed for spatial inputs and described in Section 4.2.2. It consists in grouping the 4 temporal inputs into a single input factor, defining a weather scenario. The weather scenarios are then equiprobably drawn among 22 sets of data (i.e. scenarios) collected by weather stations in various spots of the Languedoc-Roussillon in the past 40 years, and representative of dry, medium-dry and humid years. One drawback of the method is that it does not allow to quantify the individual influence of the weather components, and especially of precipitations that are empirically known to be the most influent. The whole set of input factors is described in Table 1.

3.3. **Results**

A preliminary screening via Morris method was achieved (results not shown). Morris method was developed for scalar continuous inputs, so several SA were achieved with fixed weather scenarios, since they are defined as discrete inputs. Three independent SA were performed with 3 years of weather data, representative of dry, medium-dry and humid years and in all cases, the same 6 parameters were identified as being negligible. These are the soil albedo, which depicts the radiation reflecting power of the soil surface, and the 5 cumulative thermal times defining the transition between phenological stages (leaf appearance, flowering, leaf senescence, etc.). As

Parameters	Definition	Distribution	Range (b_{inf}, b_{sup})
weather	climatic data (see main text)	discrete	-
TTSW $[mm]$	total amount of transpirable soil water	uniform	$50 - 350$
CN $[-]$	parameter of the runoff model	uniform	$60 - 99$
U $[m]$	parameter of the evaporation model	uniform	$2 - 6$
b_1 $[mm]$	parameter of the evaporation model	truncated $N(14, 5)$	$5 - 21$
b_2 $[-]$	parameter of the evaporation model	truncated $N(0.12, 0.03)$	$0.05 - 0.18$
H_{max} $[m]$	vegetation maximum height	uniform	$0.5 - 1.5$
L_{max} $[m]$	vegetation maximum width	uniform	$0.2 - 1$
Po_{min} $[-]$	vegetation minimum porosity	uniform	$0.1 - 0.5$
$FTSW_{initial}$ $[-]$	water content at January 1st	uniform	$0.3 - 1$
$FTSW_{threshold}$ $[-]$	parameter of the root absorption model	uniform	$0.3 - 0.9$
D $[m]$	inter-row distance	uniform	$1.5 - 3.5$
orientation $[rad]$	vine row orientation	uniform	$0 - \pi$
THT_{Hmax} $[°C.day]$	parameter of the vegetation growth model	uniform	$350 - 750$
THT_{Lmax} $[°C.day]$	parameter of the vegetation growth model	uniform	$350 - 750$
THT_{Pomin} $[°C.day]$	parameter of the vegetation growth model	uniform	$350 - 750$

TABLE 1. Definition of the input factors for Sobol' SA of the model for water stress in grapevine in Section 3 with the probability distribution chosen, i.e. either a discrete distribution, a uniform distribution in the range (b_{inf}, b_{sup}) or a normal distribution $N(\mu, \sigma)$ of mean μ and standard deviation σ truncated in the range (b_{inf}, b_{sup}). The input factors eliminated with Morris screening (see Section 3.3) are not included.

the results are coherent with field expertise, they were supposed extensible to every weather data, so these parameters were fixed to nominal values in the following Sobol' SA.

The results of Sobol' SA, obtained with $N = 15\,000$, are shown in Figure 1. The first order and total sensitivity indices were computed sequentially at each simulation step, which enables to follow the variation of parameters influence over time. Top Figure 1(a) shows the variation range of PLWP dynamics represented by its median, interquartile and interdecile range. The PLWP evolution is characteristic of the Languedoc-Roussillon pedoclimatic context. At the beginning of the vegetation growing season, the PWLP remains high and within a small variation range. Early June, i.e. at the beginning of the drying season (marked in the figure by a vertical dotted line), it begins to drop with an increasingly large variation range. Then PLWP slightly increases again in September, which marks the beginning of soil-water refilling by rain events.

Bottom Figure 1(a) shows the corresponding daily stacked area chart of Sobol' sensitivity indices (individual parameters and parameter interaction). The contribution of individual parameters reduces to that of 4 soil-related parameters plus the weather data, the total individual contribution of all other parameters (labeled 'others' in the figure) remains small. All along the season, interactions are high. They mainly occur between the input factors whose individual contributions are the highest. Comparison between first order and total sensitivity indices is shown in Figure 1(b). From an operational point of view, the parameters of interest are those whose influence is high when PLWP variability is high, i.e. from June to October. In this regard, a specificity of SA for temporal outputs is that the interpretation of the sensitivity indices on their own, i.e. without reference to the output uncertainty, can be misleading. This is the case for parameter $FTSW_{initial}$ (soil-water content at January 1st); its total contribution explains up to 40% of the output variability in the first phase of the season but its influence is rapidly damped when the PWLP variability increases. Identically, the total contribution of the runoff coefficient CN, which goes up to 40% at the beginning of the season, remains greatly under 15% in the phase of interest. The parameter $FTSW_{threshold}$, which is a soil-water content threshold controlling root absorption, participates in the model when the soil-water content is low enough, i.e. when PLWP is low enough. This explains why its influence is zero at the beginning of the growing cycle. It

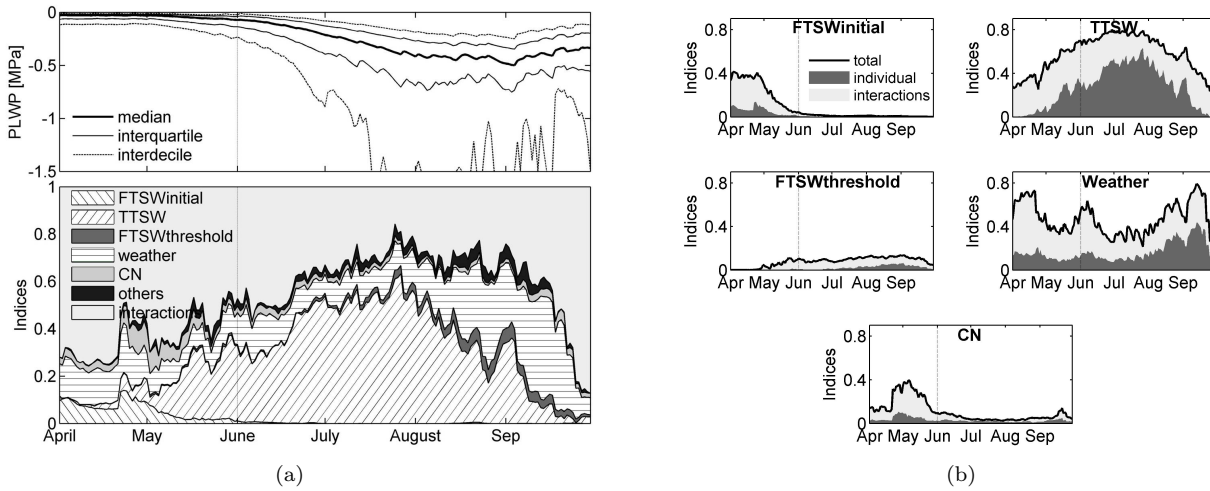(a)   (b)

FIGURE 1. Results of Sobol' SA of the model for water stress in grapevine in section 3. (a) The top figure shows the variation range of the model output dynamics represented by its median, interquartile range (range between the 1st and 3rd quartiles) and interdecile range (range between the 1st and 9th deciles). The bottom figure shows the corresponding daily stacked area chart of sensitivity indices (individual parameters and interactions). (b) Comparison of the temporal evolution of 1st order (individual) and total sensitivity indices for the 5 most influential parameters. In all figures, the vertical dotted lines mark the beginning of the phase with a high output variability.

starts playing a part around June and explains 10 to 15% of the variability at the end of the season. The 2 prevailing input factors when the PLWP variability is high are by far the weather station measurements and the TTSW. This confirms the empirical knowledge of the modelers [24]. As stated above, the methodology chosen for including the weather data as an input factor does not allow to discriminate which of the weather components have the most influence. However, additional SA (not shown) achieved for fixed weather scenarios showed the predominant role of precipitations since both PLWP and sensitivity indices dynamics strongly vary with rain events.

3.4. **Discussion**

The SA allowed to assess the general behavior of the model at the Languedoc-Roussillon regional scale. For a practical use of the model for decision support, a complementary step is now to quantify the sensitivity of PLWP at the vine plot scale, and most importantly of the irrigation amount and triggering date recommended by the DSS, and this for the whole range of pedo-climatic variability of the region. The present results, by identifying parameter interaction, and thus the parameters whose influence vary from one plot to another, will allow to define a set of region-representative vine plots.

This additional study is essential to assess the practical use of the model in the DSS, and to determine which parameters require calibration. Considering the influence of the TTSW at the regional scale, and from empirical knowledge, the best option seems to calibrate it against field data for each vine plot. Among the 3 parameters related to soil texture, only the runoff coefficient CN stands out. This result has to be confirmed at the vine plot scale, but the suitability of its calibration is currently under discussion.

4. Sensitivity analysis of a spatial model for flood risk quantification

Flood risk research makes intensive use of numerical modelling to forecast flooding events, assess the impacts of potential flooding events in terms of human well-being, economic and social development, and design efficient flood management policies. These models simulate hydrological, hydraulic and economic processes over a given study area. They are usually spatially distributed and often make use of Geographic Information Systems (GIS tools). We present in this section how sensitivity analysis has been performed on one of these flood risk models.

4.1. **NOE model for the economic assessment of flood risk**

4.1.1. *Model description*

The NOE model is a spatially distributed code which is used to assess the economic impact of flood risk [7]. It calculates the so-called EAD risk indicator (Expected Annual Damage [€/year]), which is defined as the expected economic losses due to flooding events over one year on a given study area. The EAD indicator is spatially distributed and mapped over the floodplain. Its estimation requires to compute flood damages and flood return intervals for a range of flood scenarios of various magnitudes (e.g. a 10-year flood, a 30-year flood, etc.). The return intervals of these scenarios are computed from a series of annual maximum flows at a gauging station. Estimation of economic damages caused by a given flood scenario (Figure 2) is based on the combination of various input data: i) a *hazard map* which gives maximum water depths, water velocities and flood durations over the study area for this flood scenario; ii) a *landuse map* which identifies and locates the various assets at stake over the study area (e.g., buildings, agricultural land, shops, factories, etc.); and iii) a set of so-called *damage functions* that describe the vulnerability of flood-exposed assets.

As a study area, we selected the Lower Orb River fluvial plain, located in the south of France (Hérault). This 63 sq. km^2 area has a Mediterranean subhumid regime and suffers from frequent flooding events, that may result in large economic losses. For example, the flood of Dec. 1995 - Jan. 1996, with a peak discharge of 1 700 m^3/s, affected approximately 15 000 people and caused a total amount of damage of 53 M€ [40].
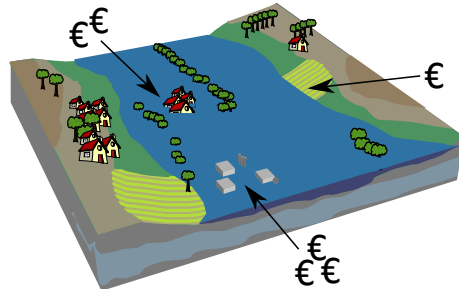

Figure 2. NOE model: estimation of economic losses for a given flood scenario.

4.1.2. *Notations*

The NOE model has both spatially distributed inputs and a spatial output. We thus adapt the general notations given in section 1 as follows:

$$\mathcal{Y} = \mathcal{G}(X_1, \ldots, X_p, \mathcal{Z}_{p+1}, \ldots, \mathcal{Z}_{p+p'}) \tag{13}$$

where X_i are scalar inputs (X_1: flood return intervals, X_2: damage functions), $\mathcal{Z}_i = \left\{ Z_i(s) \mid s \in \mathbb{R}^2 \right\}$ are spatially distributed inputs ($\mathcal{Z}_3$: hazard map, $\mathcal{Z}_4$: landuse map) and $\mathcal{Y} = \left\{ Y(s) \mid s \in \mathbb{R}^2 \right\}$ is a spatially distributed output (the EAD map). The NOE model has two important features that need to be stressed out. First, the model end-user (e.g., a water manager) is interested in the whole output map $\mathcal{Y} = \left\{ Y(s) \mid s \in \mathbb{R}^2 \right\}$ of the EAD

risk indicator, but also in knowing the aggregated value of the EAD risk indicator over one district, one city or the entire floodplain. We thus say that the NOE model is *spatially additive*, which means that the output of interest over a given spatial support $\nu \subset \mathbb{R}^2$ is the integral $Y_\nu = \int_{s \in \nu} Y(s)ds$.[1] Second, the NOE model is also a *point-based* model, meaning that the value of model output $Y(s)$ at a given location $s \in \mathbb{R}^2$ only depends on the set of scalar inputs $X_1, \ldots, X_p$ and on the values $Z_i(s)$ of spatially distributed inputs *at that same location s only* [11]. Such point-based models are encountered in various fields of environmental and earth sciences, whenever spatial interactions in the physical processes under study can be neglected in a first approximation.

4.2. **Variance-based sensitivity analysis for a spatial model**

4.2.1. *Modelling uncertainty on spatial inputs*

The hazard map ($\mathcal{Z}_3$) is a GIS raster map (regular grid) that gives water depths across the study area. Uncertainty in $\mathcal{Z}_3$ derives from the combination of various input uncertainties in the inundation mapping process; however, for the sake of simplicity, we considered the error on the high-resolution Digital Terrain Model (DTM) as the single uncertainty source in the hazard map $\mathcal{Z}_3$. The DTM – a raster surface of 5 m cell size – was initially built by stereo-photogrammetry. Both measurement errors and interpolation errors affect the quality of this input data [46]. These errors are spatially auto-correlated and are modelled by a zero-mean Gaussian Random Field with an exponential variogram model $\gamma(\cdot)$, whose characteristics were determined from a set of 500 control field points (sill = 17 cm, range = 500 m, nugget = 0.02).

The landuse map ($\mathcal{Z}_4$) is a GIS vector map, which exhibits at least two types of uncertainty: i) semantic errors (misclassification of polygonal features representing assets); and ii) geometric errors [2]. Semantic errors are described by a confusion matrix that gives a confusion probability $p(a, b)$ for each pair of possible land use types (a, b) [8]. Geometric errors are accounted for by randomizing the surface area of each polygonal feature, using a corrective random coefficient $\alpha \sim U[0.75; 0.85]$ — which corresponds to a digitalising error of 0.3 mm at the map scale [10].

From this description of uncertainty, two stochastic processes $\mathcal{P}_3$ and $\mathcal{P}_4$ were designed to generate any number of random realisations of $\mathcal{Z}_3$ and $\mathcal{Z}_4$. Stochastic processes $\mathcal{P}_3$ and $\mathcal{P}_4$ are *fixed*, and the sensitivity analysis focuses on the overall uncertainty on $\mathcal{Z}_3$ and $\mathcal{Z}_4$ that is generated by $\mathcal{P}_3$ and $\mathcal{P}_4$.[2]

4.2.2. *Propagating uncertainty: the map labelling method*

To estimate variance-based sensitivity indices $S_{\mathcal{Z}_3}$ and $S_{\mathcal{Z}_4}$ associated with spatial inputs $\mathcal{Z}_3$ and $\mathcal{Z}_4$, we use the *map labelling* method [25]. The uncertainty of $\mathcal{Z}_i$ is represented using a set of n_i randomly generated realisations (n_i possibly large), generated by the stochastic process $\mathcal{P}_i$. Each realisation is then *labelled* by a single integer. The steps are as follows:

Step 1: generate n_i random realisations of the spatial input $\mathcal{Z}_i = \{Z_i(s)|s \in \mathbb{R}^2\}$ using a stochastic process $\mathcal{P}_i$. Store them in some permanent disk space. These realisations are considered as equiprobable;

Step 2: label each realisation of $\mathcal{Z}_i$ with a unique integer in $[\![1; n_i]\!]$; the random realisation associated with label l is denoted by $\mathcal{Z}_i^{(l)}$;

Step 3: consider a random label $\mathbf{L}_i$, which is assumed to follow a discrete uniform pdf in $[\![1; n_i]\!]$;

Step 4: sample random label $\mathbf{L}_i$ along with other model inputs to generate an input matrix, following the sampling procedure described in Section 1.2;

---

[1]Many other environmental models are also spatially additive (e.g. models that describe the rainfall over a study area), but some are not (e.g., a model that predicts the maximum pollutant concentration $\max\{Y(s)|s \in \nu\}$ over a study area).

[2]Other choices could have been made, such as: i) considering that $\mathcal{P}_i$ is itself uncertain—so-called *second-level uncertainty*, e.g., modelling uncertainty on the variogram model $\gamma(\cdot)$ or on the probabilities confusion $p_{i,j}$; or ii) not considering the overall uncertainty on $\mathcal{Z}_i$, but focusing on the impact of some important scalar features of $\mathcal{Z}_i$—which is the perspective adopted in Section 5.3.

Step 5: evaluate the model $\mathcal{G}$ for each line of the input matrix; spatially distributed input $\mathcal{Z}_i$ is replaced by a reference to the random label l_i: on line j, the sampled value $l_i^{(j)}$ of the label indicates that random realisation $\mathcal{Z}_i^{(l_i^{(j)})}$ must be considered to evaluate model $\mathcal{G}$ for this sample line;

Step 6: from the output vector, calculate the estimators of first and total-order sensitivity indices $\hat{\mathrm{S}}_{\mathbf{L}_i}$ and $\hat{\mathrm{S}}_{\mathbf{L}_i}^T$ of random label $\mathbf{L}_i$ with Equation (7). These indices are taken as a measure of the influence of $\mathcal{Z}_i$ on the variance of model output.

Sample sizes $n_3 = 1000$ and $n_4 = 100$ were used for the two spatial inputs of the NOE model (hazard map $\mathcal{Z}_3$ and landuse map $\mathcal{Z}_4$).

4.2.3. *Which output of interest? Site and block sensitivity indices*

Variance-based sensitivity indices must be computed with respect to a scalar quantity of interest derived from the spatially distributed model output $\mathcal{Y} = \{Y(s) | s \in \mathbb{R}^2\}$. In the case of a spatially additive model such as NOE, we suggest to use two different types of sensitivity indices: *site* and *block* sensitivity indices.

We first call *site sensitivity indices*, denoted by $\mathrm{S}_{\{i\}}^T(s)$, the total-order indices computed on a point support, i.e., with respect to the output of interest $Y(s)$ at a specific location $s \in \mathbb{R}$; $\mathscr{S}_{\{i\}} = \{\mathrm{S}_{\{i\}}^T(s) \mid s \in \mathbb{R}^2\}$ denotes the 2D-field of site sensitivity indices.

Next, we call *block sensitivity indices*, denoted by $\mathrm{S}_{\{i\}}^T(\nu)$, the total-order indices computed on a block support $\nu \subset \mathbb{R}^2$, i.e., with respect to the block average $Y_\nu = \int_{s \in \nu} Y(s) ds$. On the NOE model, we considered three individual spatial supports ν_1, ν_2 and ν_3 of increasing surface area. These three supports were selected at random, and their study is only meant to be illustrative: ν_1 is a single house located on the western bank of the Orb Delta ($|\nu_1| = 1$ ha), second support ν_2 is the administrative district of Sauvian ($|\nu_2| = 13$ sq. km^2), and ν_3 is the entire floodplain ($|\nu_3| = 63$ sq. km^2).

4.3. **Results**

4.3.1. *Maps of site sensitivity indices*

Figure 3 displays the maps $\mathscr{S}_{\{i\}}$ of site sensitivity indices for uncertain model inputs X_1, X_2, $\mathcal{Z}_3$ and $\mathcal{Z}_4$. Spatial distribution of site sensitivity indices proves to be heterogeneous. For example, by comparing these sensitivity maps with a map of the study area, we can find that the hazard map $\mathcal{Z}_3$ and the landuse map $\mathcal{Z}_4$ display smaller sensitivity indices on the urban areas than on the areas covered with agricultural land. The interpretation of this finding will not be discussed here, but it brought a better understanding of the NOE model behaviour.

4.3.2. *Block sensitivity indices*

Figure 4 displays the block sensitivity indices $\mathrm{S}_{\{i\}}^T(\nu)$ computed for each uncertain model input with respect to the aggregated model output $Y_\nu = \int_{s \in \nu} Y(s) ds$ over spatial supports ν_1 (house), ν_2 (district) and ν_3 (floodplain). It clearly suggests that the ranking of uncertainty sources depends on the surface area of the spatial support ν (*change of support effect*). The variance of the aggregated model output on ν_1 (smallest support) appears to be mainly explained by the uncertainty on the two spatially distributed inputs, that is, the hazard map $\mathcal{Z}_3$ and the landuse map $\mathcal{Z}_4$ (sensitivity indices: 0.65 and 0.8, respectively). On the contrary, the non spatially distributed inputs (flood return intervals X_1 and damage functions X_2) prove to be the most important sources of uncertainty when aggregating the EAD risk indicator over the total floodplain ν_3 (largest support).

4.4. **Change of support effect**

The *change of support* effect observed in Section 4.3.2 can be explained theoretically. Let consider from now on a model $\mathcal{Y} = \mathcal{G}(\mathbf{X}, \mathcal{Z})$ with a *single* spatial input $\mathcal{Z}$ which is assumed to be a second-order stationary random field (SRF) with an isotropic covariance function $C(\cdot)$. Let's also assume that model $\mathcal{G}$ is point-based, that is, there exists a mapping $\mathcal{G}_{\mathrm{loc.}} : \mathbb{R}^p \times \mathbb{R} \to \mathbb{R}$ such that $\forall s \in \mathbb{R}^2, \quad Y(s) = \mathcal{G}_{\mathrm{loc.}} [\mathbf{X}, Z(s)]$. Under the stationary
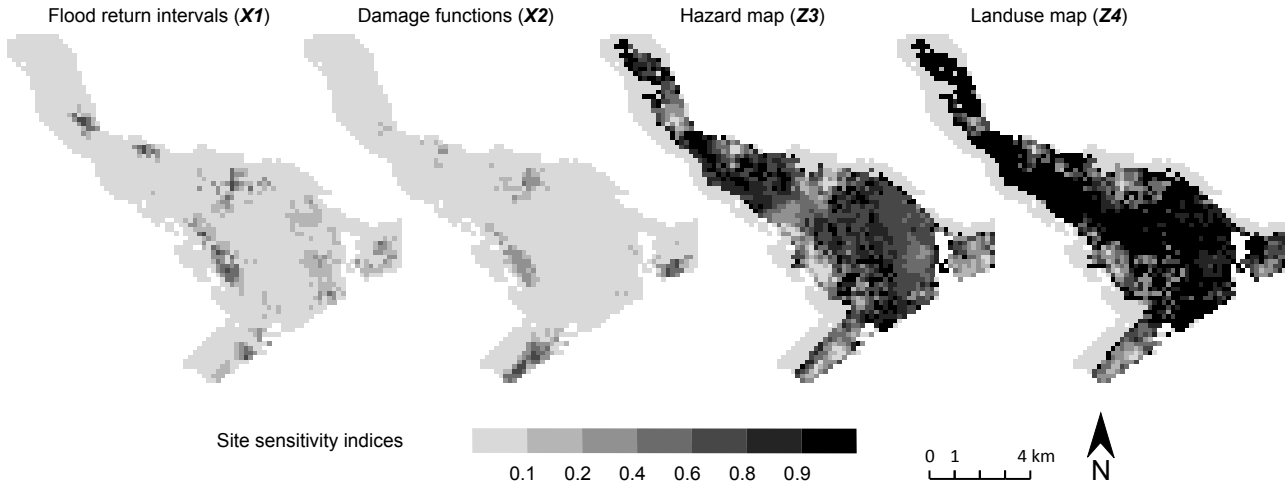
FIGURE 3. Maps $\mathscr{S}_{\{i\}}$ of site sensitivity indices for each model input.
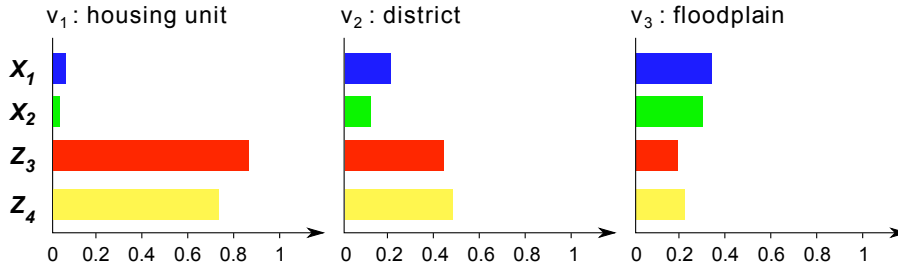

FIGURE 4. Block sensitivity indices $\mathrm{S}_{\{i\}}^T(\nu)$ on spatial supports ν_1 to ν_3.

hypothesis on $\mathcal{Z}$, site sensitivity indices $\mathrm{S}_{\{i\}}^T(s)$ do not depend on site $s \in \mathbb{R}^2$ and can simply be denoted by $\mathrm{S}_{\{i\}}^T$. For any support $\nu \subset \mathbb{R}^2$, the block sensitivity indices of $\mathbf{X}$ and $\mathcal{Z}$ over support ν satisfy :

$$\frac{\mathrm{S}_{\mathcal{Z}}^T(\nu)}{\mathrm{S}_{\mathbf{X}}^T(\nu)} \simeq \frac{\nu_c}{|\nu|} \quad \text{with} \quad \nu_c = A \cdot \frac{\mathrm{S}_{\mathcal{Z}}^T}{\mathrm{S}_{\mathbf{X}}^T} \tag{14}$$

where A depends only on the covariance function $C(\cdot)$. The symbol $\simeq$ means that the equality is valid to a first approximation [35, for a detailed proof]. Equation (14) shows that the ratio $\nu_c/|\nu|$ determines the relative contribution of the model inputs $\mathcal{Z}$ and $\mathbf{X}$ to the output variance $\mathrm{var}(Y_\nu)$. For a small ratio (i.e., when the area of the support ν is large compared with the critical size ν_c), the variability of $\mathcal{Z}$ is mainly *local*, and the spatial correlation of $\mathcal{Z}$ over ν is weak. This local variability averages over the support ν when the aggregated model output Y_ν is computed; hence, spatial input $\mathcal{Z}$ explains a small fraction of the output variance $\mathrm{var}(Y_\nu)$. This *change of support effect* is of great importance to better understand how the result of a sensitivity analysis may change under model upscaling or downscaling.

5. SENSITIVITY ANALYSIS IN AGRO-ECOLOGY WITH LANDSCAPE AS AN INPUT FACTOR

Today, much research in agronomy and ecology involves complex dynamic processes that occur at the scale of an agricultural landscape. These processes include population genetics and population dynamics which depend on the composition and structure of the landscape. In this section, we consider theoretical dynamic models that

aim to study and better understand such processes. We focus on the case when the model is spatially explicit and the influence of landscape features is of interest. In the previous section, the emphasis was on uncertainties about a given landscape. In this section, we are interested in comparing the effects of different landscapes on the output (see Fig. 5).

5.1. **Models of coexistence and crop disease propagation**

5.1.1. *Models short description*

Similar models have been developed for two types of agricultural issues. The first type (see *e.g.* [22, 45]) deals with the coexistence between GM (genetically modified) and non-GM crop varieties. The objective is to determine the conditions (if any) for both types of crops to coexist within the same area, and to minimize the risk of non-GM fields to exceed given thresholds of GM content because of cross-pollination from GM fields. The second issue (see *e.g.* [32]) deals with crop disease propagation over an agricultural area. The propagation depends on the genetic composition of the pathotypes in the environment and on resistance genes which differ between crop varieties. The objective is to better manage the allocation of crop varieties to the different fields, in order to minimize the strength and speed of disease propagation.

In both cases, the problem involves dynamic processes and dispersal (of pollen or spores) between fields. The landscape is described by its composition (the proportions of the different varieties) and by its structure (the field pattern and the spatial allocation of the varieties). Both characteristics may have an influence on the output.

5.1.2. *Landscape input*

In the models developed in [22, 32, 45], the spatial support is an idealised agricultural landscape within a square area $\mathcal{A}$ of a few squared kilometers. The field pattern is a polygon tessellation in $\mathcal{A}$ with each polygon representing a crop field. It can be described by two categorical and spatialised inputs $\mathcal{Z}_1$ and $\mathcal{Z}_2$, where $Z_1(s)$ is the field containing location s and $Z_2(s)$ is the variety allocated to $Z_1(s)$.

5.1.3. *Main characteristics*

In [22, 32, 45], space and time are discretised at a fine scale. The models contain state variables $w(s,t)$ which represent population quantities in location s at time t. The quantity $w(s,t)$ depends on the quantities $w(s',t-1)$ through integro-differential equations that describe pollen or spore dispersal. The final output $\mathcal{Y}$ is a given function of the state variables $w(s,t)$. Like in Section 4, it can be written

$$\mathcal{Y} = \mathcal{G}(X_1, \cdots, X_p, \mathcal{Z}_1, \mathcal{Z}_2), \tag{15}$$

where X_i are scalar inputs, $\mathcal{Z}_i$ are spatially distributed inputs and $\mathcal{Y}$ is a spatially distributed output. Note that $\mathcal{G}$ is never a point-based model in the sense of Section 4, because pollen and spore dispersal generate dependency between what happens in any two locations s and s'.

Compared to Section 4, we are not interested in the output map $\mathcal{Y} = \{Y(s)|s \in \mathbb{R}^2\}$ *per se*, but rather in scalar summaries of $\mathcal{Y}$ which may also involve features of $\mathcal{Z}_1$ and $\mathcal{Z}_2$.

5.2. **Pragmatic sensitivity analysis with input factors related to landscape features**

The output $\mathcal{Y}$ depends on the particular field pattern associated with $\mathcal{Z}_1$ and on the variety allocation to fields associated with $\mathcal{Z}_2$. When performing sensitivity analyses of $\mathcal{G}$, we would like to account for and quantify this influence. This can be done more or less formally.

A simple and pragmatic approach consists in generating a finite collection of landscapes based on a few input factors. In [45], six real-world field patterns were used and $\mathcal{Z}_1$ was defined through a categorical input factor called MAP at six levels (see Figure 5). The variety allocation $\mathcal{Z}_2$ depended on $\mathcal{Z}_1$ and on three additional input factors : MAIZ, the maize area proportion in $\mathcal{A}$; GM, the GM-maize area proportion; AGGR, the degree of aggregation between GM fields.
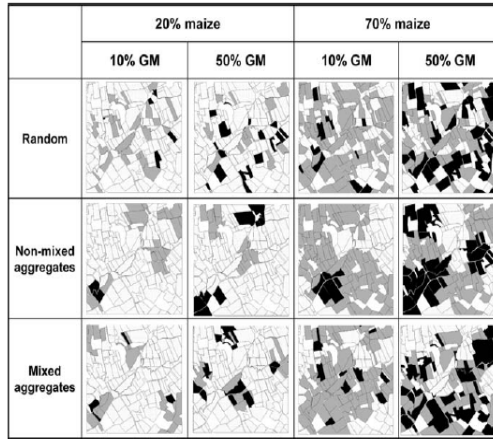
FIGURE 5. A set of maps generated in [45] with the same field pattern $\mathcal{Z}_1$ and different variety allocations $\mathcal{Z}_2$.

In [45] all factors, including MAP, MAIZ, GM and AGGR, were given a finite number of levels and considered as independent. Simulations were run according to an unreplicated complete factorial design. In each simulation, the field pattern $\mathcal{Z}_1$ was entirely determined by the level of factor MAP. For each combination of levels of MAP, MAIZ, GM and AGGR, the variety allocation $\mathcal{Z}_2$ was generated by a simulated annealing algorithm. An analysis of variance was then performed as described in [27]. Note that this can be considered as a special case of the variance-based sensitivity analysis described in Section 1, where all input factors X_i are given a symmetric multinomial distribution.

5.3. Landscape representation by stochastic processes

An alternative is to model the spatial variability (or uncertainty) of interest by defining $\mathcal{Z}_1$ and $\mathcal{Z}_2$ as two stochastic processes, as proposed in Section 4. We shortly review recent works on agricultural field patterns and variety allocation modelling by stochastic processes.

5.3.1. *Field pattern*

Stochastic field patterns ($\mathcal{Z}_1$ input factor) can be modelled by Voronoi tessellations based on stochastic point processes. In [22], it was shown that for more realistic field patterns, the seeds of the Voronoi tessellations should be distributed as a pairwise-interaction point process rather than a Poisson process. Up to a scalar, the density function of a pairwise-interaction process is equal to

$$\lambda^n \prod_{i<j} \gamma_\theta(\|s_i - s_j\|),$$

where λ is the process intensity, the s_i are the locations of the n centroids, $\|s - s'\|$ is the Euclidean distance between s and s', and $\gamma_\theta(\cdot)$ is the interaction function with interaction parameter θ. Parameter λ allows to control the expected number of points (and so the number of fields) in the area, while θ determines the distribution of pairwise distances between the centroids (and so, indirectly, the spatial distribution of field sizes).

The polygons associated with a Voronoi tessellation based on random seeds do not look like actual agricultural fields. To overcome this drawback, rectangular tessellations due to Mackisack and Miles [26] were investigated by Le Ber *et al.* [22]. More recently, Kiêu *et al.* [18, 19] have investigated the T-tessellations introduced by Arak *et al.* [1]. In a T-tessellation, each vertex lies at the intersection of three edges, two of which are aligned. In [18], the definition of a completely random T-tessellation is proposed. Based on this definition, more specific

T-tessellation processes can be defined by using Gibbsian density functions $h(T)$ with $-\log h(T) = \sum_k \theta_k g_k(T)$, where T denotes a T-tessellation , $g_k(T)$ is a feature of interest measured on T (number of fields, size variability, angle values, etc.), and θ_k is a parameter. For sensitivity analysis, it is possible either to fix the density function and its parameters or to fix the density function and vary its parameters.

5.3.2. *Variety allocation*

Conditional on $\mathcal{Z}_1$, variety allocation ($\mathcal{Z}_2$ input factor) is a stochastic function from the set of fields to the set of varieties. In [45], the allocation is generated by a simulated annealing algorithm. Alternatively, it can be defined as a Gibbs process with a density function allowing to control the variety proportions and their degree of spatial aggregation (see also [32]).

5.4. **Consequences on the sensitivity analysis**

When the variance-based sensitivity analysis of Section 1 includes the agricultural landscape among the input factors, or indeed any similarly complex input, several options arise. In particular, for each spatial input $\mathcal{Z}_i$,

- the stochastic process $\mathcal{Z}_i$ may be considered as a landscape input factor *per se*;
- the stochastic process $\mathcal{Z}_i$ may be generated conditionally on some of its parameters that are considered as the actual landscape input factors. To be useful, these parameters must correspond to landscape features of interest. Examples are the parameters *e.g.* λ or θ in the tessellation processes described above or the MAP, MAIZ, GM and AGGR values for the allocation of varieties in [45].

In the first case, the map labelling approach of Section 4 can be applied. In the second case, the parameters of interest of the process are usually included among the independent input factors X_i. The sensitivity analysis is performed as described in Section 1, with $\mathcal{Z}_i$ generated using the parameter values X_i^j given by the experimental simulation design.

In fact, the complete input structure is often more complicated than appears at first sight. For example, consider the input $\mathcal{Z}_2$ discussed in Section 5.2. It is an additional input factor whose levels are the field-and-variety pattern after varieties have been allocated to the fields. Let's call this factor VARMAP. Then VARMAP is a categorical factor nested in MAP, MAIZ, GM and AGGR. A complete variance decomposition would require replications of VARMAP for given combinations of these nesting factors. If $\mathcal{G}$ is stochastic, replicated runs of $\mathcal{G}$ for a given level of VARMAP would be required as well. For the sake of conciseness, this will not be developed any further in this paper.

References

[1] T. Arak, P. Clifford, and D. Surgailis. Point-based polygonal models for random graphs. *Advances in Applied Probability*, 25:348–372, 1993.

[2] O. Bonin. Sensitivity analysis and uncertainty analysis for vector geographical applications. In *7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Lisbon, Portugal, July 5-7, 2006.

[3] C. Ciric, P. Ciffroy, and S. Charles. Use of sensitivity analysis to identify influential and non-influential parameters within an aquatic ecosystem model. *Ecological Modelling*, 246:119–130, 2012.

[4] E. de Rocquigny, N. Devictor, and S. Tarantola. *Uncertainty in Industrial Practice.* Wiley, New York, 2008.

[5] B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.

[6] B. Efron and R.J. Tibshirani. *An introduction to the bootstrap.* Chapman & Hall/CRC, 1993.

[7] K. Erdlenbruch, E. Gilbert, F. Grelot, and C. Lescouliers. Une analyse coût-bénéfice spatialisée de la protection contre les inondations : application de la méthode des dommages évités à la basse vallée de l'Orb. *Ingénieries Eau-Agriculture-Territoires*, 53:3–20, 2008.

[8] P.F. Fisher. Modelling soil map-unit inclusions by Monte Carlo simulation. *Int. J. Geogr. Inf. Sci.*, 5(2):193–208, 1991.

[9] M.A. Grepl and A.T. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *Mathematical Modelling and Numerical Analysis*, 39(1):157–181, 2005.

[10] T. Hengl. Finding the right pixel size. *Comput. Geosci.*, 32:1283–1298, 2006.

[11] G.B.M. Heuvelink, D.J. Brus, and G. Reinds. Accounting for spatial sampling effects in regional uncertainty propagation analysis. In *Proceedings of the Ninth International Symposium on Spatial Accuracy Assessment in Natural Ressources and Environmental Sciences*, pages 85–88, Leicester, UK, July 20-23, 2010.

[12] W. F. Hoeffding. A class of statistics with asymptotically normal distributions. *Annals of Mathematical Statistics*, 19:293–325, 1948.

[13] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, 52(1):1–17, 1996.

[14] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two sobol index estimators. Preprint available at hal-00665048, 2012.

[15] A. Janon, M. Nodet, and C. Prieur. Certified reduced-basis solutions of viscous burgers equations parametrized by initial and boundary values. Preprint available at inria-00524727, 2010, accepted in *Mathematical modelling and Numerical Analysis*.

[16] A. Janon, M. Nodet, and C. Prieur. Uncertainties assessment in global sensitivity indices estimation from metamodels. Preprint available at inria-00567977, 2011, *accepted in International Journal for Uncertainty Quantification*.

[17] A. Janon, M. Nodet, and C. Prieur. Goal-oriented error estimation for reduced basis method, with application to certified sensitivity analysis. Preprint available at hal-00721616, 2012.

[18] K. Kiêu, K. Adamczyk, H. Monod, and R.S. Stoica. A completely random t-tessellation model and gibbsian extensions. Preprint available at hal-00785980, 2013.

[19] K. Kiêu, K. Adamczyk, R.S. Stoica, and H. Monod. A new model for random t-tessellations. In *14th Workshop on Stochastic Geometry, Stereology and Image Analysis*, page 29, Neudietendorf, 2007.

[20] J.P.C. Kleijnen. *Design and analysis of simulation experiments*. Springer Publishing Company, Incorporated, 2007.

[21] D.J. Knezevic, N.C. Nguyen, and A.T. Patera. Reduced basis approximation and a posteriori error estimation for the parametrized unsteady boussinesq equations. *Mathematical Models and Methods in Applied Sciences*, 2010.

[22] F. Le Ber, C. Lavigne, K. Adamczyk, Angevin, N. Colbach, J.-F. Mari, and H. Monod. Neutral modelling of agricultural landscapes by tessellation methods – Application for gene flow simulation. *Ecological Modelling*, 220:3536–3545, 2009.

[23] E. Lebon, V. Dumas, P. Pieri, and H.R. Schultz. Modelling the seasonal dynamics of the soil water balance of vineyards. *Functional Plant Biology*, 30:699–710, 2003.

[24] C. Van Leeuwen, O. Tregoat, X. Choné, B. Bois, D. Pernet, and J.-P. Gaudillère. Vine water status is a key factor in grape ripening and vintage quality for red bordeaux wine. how can it be assessed for vineyard management purposes? *J. Int. Sci. Vigne Vin*, 43:121–134, 2009.

[25] L. Lilburne and S. Tarantola. Sensitivity analysis of spatial models. *Int. J. Geogr. Inf. Sci.*, 23(2):151–168, 2009.

[26] M. Mackisack and R. Miles. Homogeneous rectangular tessellations. *Advances in Applied Probability*, 28:993–1013, 1996.

[27] H. Monod, C. Naud, and D. Makowski. Uncertainty and sensitivity analysis for crop models. In D. Wallach, D. Makowski, and J. W. Jones, editors, *Working with Dynamic Crop Models: Evaluation, Analysis, Parameterization, and Applications*, pages 55–100. Elsevier, 2006.

[28] M.D. Morris. Factorial sampling plans for preliminary computational experiments. *Technometrics*, 33:161–174, 1991.

[29] N.C. Nguyen, K. Veroy, and A.T. Patera. Certified real-time solution of parametrized partial differential equations. *Handbook of Materials Modeling*, pages 1523–1558, 2005.

[30] J.E Oakley and A. O'Hagan. Probabilistic sensitivity analysis of complex models: a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(3):751–769, 2004.

[31] United States Department of Agriculture. Chapter 10 - estimation of direct runoff from storm rainfall. *National Engineering Handbook*, 2004.

[32] J. Papaïx, O. David, C. Lannou, and H. Monod. Dynamics of adaptation in spatially heterogeneous metapopulations. *PLoS One*, 8(2):e54697, 2013.

[33] J.B. Passioura. Simulation models: science; snake oil, education, or engineering? *Agronomy Journal*, 88:690–694, 1996.

[34] A. Pellegrino, E. Lebon, M. Voltz, and J. Wery. Relationships between plant and soil water status in vine (vitis vinifera l.). *Plant and Soil*, 266:129–142, 2004.

[35] N. Saint-Geours. *Sensitivity analysis of spatial models: application to cost-benefit analysis of flood risk management plans*. PhD thesis, Université Montpellier 2, 2012.

[36] A. Saltelli, K. Chan, and M. Scott. *Sensitivity Analysis*. John Wiley & Sons, 2000.

[37] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. *Global Sensitivity Analysis: The Primer*. Wiley, New York, 2008.

[38] T.J. Santner, B.J. Williams, and W. Notz. *The design and analysis of computer experiments*. Springer Verlag, 2003.

[39] L. Sirovich. Turbulence and the dynamics of coherent structures. part i-ii. *Quarterly of applied mathematics*, 45(3):561–590, 1987.

[40] SMVOL. Programme d'actions de prévention des inondations sur les bassins de l'Orb et du Libron. Technical report, Syndicat Mixte des Vallées de l'Orb et du Libron, 2011.

[41] I. M. Sobol'. Sensitivity analysis for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414, 1993.

[42] B. Sudret. Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, 93(7):964–979, 2008.

[43] J.Y. Tissot and C. Prieur. Estimating sobol' indices combining monte carlo estimators and latin hypercube sampling. *Preprint available at hal-00743964*, 2012.

[44] K. Veroy and A.T. Patera. Certified real-time solution of the parametrized steady incompressible navier–stokes equations: rigorous reduced-basis a posteriori error bounds. *International Journal for Numerical Methods in Fluids*, 47(8-9):773–788, 2005.

[45] V. Viaud, H. Monod, C. Lavigne, F. Angevin, and K. Adamczyk. Spatial sensitivity of maize gene-flow to landscape pattern: a simulation approach. *Landscape Ecology*, 23:1067–1079, 2008.

[46] S.P. Wechsler. Uncertainties associated with digital elevation models for hydrologic applications: a review. *Hydrol. Earth Syst. Sci.*, 11(4):1481–1500, 2007.