# A filter-based approach for global sensitivity analysis of models with functional inputs

Sébastien Roux, Patrice Loisel, Samuel Buis


# A filter-based approach for global sensitivity analysis of models with functional inputs

Sébastien Roux, Patrice Loisel, Samuel Buis

PII: S0951-8320(17)30789-5
DOI: https://doi.org/10.1016/j.ress.2019.01.012
Reference: RESS 6364

To appear in: *Reliability Engineering and System Safety*

Received date: 30 June 2017
Revised date: 21 December 2018
Accepted date: 10 January 2019

**Highlights**

- A new method for GSA of models with functional inputs is introduced
- Filtering operators are used to perturb the functional inputs
- The response on model outputs is studied with Sobol Indices of boolean variables
- Relationships between Sobol Indices and error due to filtering are established
- The approach is generic and can improve model understanding or input acquisition

# A filter-based approach for global sensitivity analysis of models with functional inputs

Sébastien Roux [1a], Patrice Loisel[a], Samuel Buis[b]

[a] MISTEA, INRA, Montpellier SupAgro, Univ Montpellier, Montpellier, France
[b] EMMAH, INRA, Université d'Avignon et des Pays de Vaucluse, Avignon, France





[1] corresponding author email : sebastien.roux@inra.fr

*Preprint submitted to Elsevier* *January 11, 2019*

**Abstract**

We design a method called filter-based global sensitivity analysis (filter-based GSA) to analyze computer models with functional inputs. Understanding the impact of functional inputs is central in many applications like building energy or environmental studies. The present work is a step further in the analysis of the impact of functional inputs signal components onto model responses of interest. To perform filter-based GSA, the functional inputs are modified with filters in order to either enhance or suppress some components in the signal. The influence of filters on the model response is assessed by computing the Sobol' indices of Boolean factors that trigger the filters application. Two relationships between these indices and the error resulting from the filters application are established. The method is illustrated with smoothing filters applied on the climatic inputs of a toy model simulating crop yield. We show that the high frequencies for three out of the four climatic inputs are not important in the tested configuration. The present method does not require any hypotheses on the model or the type of filters used. It seems promising for model understanding, validation or simplification.

*Keywords*: Sobol' indices, Switching variables, Boolean factors, Filtering error, Crop model, Climatic inputs

## 1. Introduction

This work is devoted to the analysis of computer models having functional inputs. Functional inputs are related, for example, to temporally or spatially distributed inputs. They are commonly used in environmental and building energy studies in particular for studying the adaptations of physical systems to present or future climatic conditions [1, 2, 3, 4]. They play an important role in other domains such as nuclear energy [5], oil industry [6] or mechanical engineering [7]. However the complex structure due to inter and auto-correlation of temporal or spatial inputs as well as their large dimension make the deep analysis of their impact onto model predictions a challenging issue.

Several methods have been proposed to apply sensitivity analysis techniques to functional inputs. For instance, the so-called labeling [6, 8, 9] or joint metamodeling [10] approaches aims at computing Sobol' sensitivity indices [11] of functional input variables considered as a whole. In several studies [1, 2], scalar sensitivity analyses are repeated for different values of the functional inputs. The post-processing of the results may give access to more detailed information on the interactions between functional and non functional inputs than the above-cited methods, but at the price of a larger computational cost.

There is however a real need for complementary analyses to better understand which parts of these complex inputs are responsible for their global importance: is it their mean, their frequencies, their extreme values? Is the sensitive information localized around a critical instant (or space localization) or is it spread globally? The previous questions can be reformulated using the concept of component, which we use here to denote generically a term or group of terms in an additive decomposition of a functional variable (e.g some frequencies in a decomposition of a signal in a Fourier basis). The problem is thus to quantify the importance of such functional inputs components. Such knowledge can bring useful insights into the model behavior for understanding or validating model responses. It can also be used for simplifying the structure of model inputs in order to get simpler models or models with inputs easier to acquire.

To our knowledge, only few recent studies tackle the issue of detecting or characterizing sensitive components of functional inputs: Fruth et al. [7] introduces a decomposition into piecewise functions coupled with a screening algorithm, Picheny et al. [12] applied sparse functional regression techniques. Both approaches concentrate on one particular type of sensitive components: the detection of sensitive time intervals.

The present paper introduces a new method, called filter-based global sensitivity analysis (filter-based GSA), that allows to quantify the importance of *a priori* defined components of model functional inputs. It is less ambitious than methods that would automatically detect sensitive components but offers a very flexible and mathematically justified framework for testing hypothesis on their importance. Its main principles are i) to associate to a potentially sensitive component a filtering operator (or filter) that will transform a functional input such that the component is suppressed or enhanced and ii) to quantify in the model output the effect of applying the filter by using GSA. An application detailed in this article is to test using a smoothing filter the influence of fine details of temporal inputs, i.e. their high frequency content in a Fourier decomposition.

The paper is organized as follows: we first introduce the formalization of the problem and the filter-based GSA approach in Section 2. We analyze the link between the Total Sobol' Index (TSI) and errors due to filtering in Section 3. The method is illustrated on smoothing filters applied on the inputs of a toy crop model in Sections 4 and 5.

## 2. The filter-based GSA method

### *2.1. Context and notations*

The global sensitivity analysis (GSA) of a simulation model response $f(Z)$, where $Z$ is a set of independent scalar inputs $(Z_i)$, can be performed by estimating variance-based sensitivity indices (also called Sobol' indices) [13]. The first order and the total sensitivity indices are respectively defined as:

$$SI_1(Z_i) = \frac{Var[E[f(Z)|Z_i]]}{Var[f(Z)]}$$
$$TSI(Z_i) = \frac{E[Var[f(Z)|Z_{-i}]]}{Var[f(Z)]}$$

where $Var$ is the variance operator, $E$ the mathematical expectation, $Z_i \in Z$ and $Z_{-i} = Z \backslash Z_i$.

In this work, we consider a model response $f(X, p)$ whose inputs are split into two groups: $X = (x_1, .., x_N)$ a vector of functional inputs and $p = (p_1, .., p_M)$ a vector of scalar inputs. We suppose that we have a collection $(X^l)$ of $L$ datasets of the functional inputs: $(X^l)_{l=1,..,L} = (x_1^l, .., x_N^l)_{l=1,..,L}$. For example, $X^l$ may represent the weather data at a given year $l$. The sensitivity analysis is performed over the joint pdf of $p$ and $l$. We assume known the pdf of $p$, that $l$ is uniformly distributed within $[1, L]$ and that $p$ and $X$ are independent of each other. Therefore, the scalar model response of interest $y = f(X^l, p)$ is a random variable. We also suppose that, compared to $p$, the variability of the vector of functional inputs $X$ has a substantial effect on the model response $y$. This effect can be quantified by evaluating the sensitivity of the model response to the random factors $l$ and $p$ as in

[14, 9, 2]. Now, we would like to know to which components of the input signals the model response is more sensitive to.

*2.2. Principle of the method*

We propose an approach to test if some *a priori* defined components of the functional inputs $X$ explains the variations of the model output. The principles of the method are i) to transform the functional inputs with filtering operators in order to suppress or enhance some of their components and ii) to quantify the effect of applying the filters on the model output. The quantification procedure is a global sensitivity analysis applied to a reformulated model having additional Boolean factors that trigger the filters application. If the effect of applying a filter is high then it means that the associated component is important in the simulation of the model output. Otherwise, it means that the component is not important and that the model or its functional inputs can be simplified w.r.t. this component.

For illustration purposes, let us consider the crop yield model further detailed in Section 4. We are interested in the analysis of crop yield simulations with respect to several climatic variables measured over the cropping season at a daily time-step. The aim of the case study is to identify if the temporal resolution for climatic data acquisition could be decreased without significantly impacting crop yield prediction. This knowledge would eventually allow simplifying the model formulation and above all, alleviating the climatic data acquisition as less data would recorded. This problem is related to the question of influential frequency content of model functional inputs. To address this issue, the filter-based GSA is applied with a smoothing filter that removes fine details (high frequencies in the Fourier decomposition) of the different climatic inputs.

### 2.2.1. Filtering operators

We associate to a potentially sensitive component a filtering operator. In our context, a filtering operator is a signal transformation that can typically suppress or enhance some components of a signal, as defined in the introduction. No hypothesis is required on the filter definition (for instance no linearity of the operation), but its properties in terms of affected components are essential to interpret the results of the method. Two different cases can be considered: the case of a single filter and the case of several filters. In the latter case, a question of concern is the quantification of the different possible effects stemming from all possible combinations of the filters. To this aim, we introduce $q$ filters denoted by $(g_1, .., g_q)$ that apply on vectors of functional inputs. The filtered vector $\widetilde{X}^l$ when applying all filters writes:

$$\widetilde{X}^l = g_1 \circ .. \circ g_q(X^l)$$

### 2.2.2. Boolean switching factors

In order to quantify the effect of applying one or several filters, we perform a sensitivity analysis on a new model $\widetilde{f}$ that extends the initial model $f$ using additional Boolean inputs named switching factors. Switching factors have been introduced in the context of sensitivity analysis in [15] to assess the sensitivity to the presence of stochastic errors in spatially distributed inputs. They have been further analyzed in the general context of functional inputs in [10]. In the context of filter-based GSA, switching factors are used to trigger the filters application on the functional inputs $X^l$. When a single filter is considered ($q = 1$), we use a single switching factor $\eta$ to trigger the application of filter $g$. This defines a new model response $\widetilde{f}(\eta, l, p)$ such as:

$$\widetilde{f}(\eta, l, p) = \begin{cases} f\left(X^l, p\right) & \text{if } \eta = 0 \\ f\left(g(X^l), p\right) & \text{if } \eta = 1 \end{cases} \quad (1)$$

When $q$ filters $g_1, .., g_q$ are studied, we use $q$ switching variables $\eta_1, .., \eta_q$ and a new model $\widetilde{f}$ so that:

$$\widetilde{f}\,(\eta_1, .., \eta_q, l, p) = f\left(\bar{g}_1 \circ .. \circ \bar{g}_q(X^l), p\right) \quad (2)$$

$$\text{with } \bar{g}_i(X^l) = \begin{cases} X^l & \text{if } \eta_i = 0 \\ g_i(X^l) & \text{if } \eta_i = 1 \end{cases}$$

Using this new model $\widetilde{f}$, we quantify the effect of applying a filter and thus the influence of the associated component, using the total sensitivity indices of the switching factor when performing a global sensitivity analysis on factors $(\eta, l, p)$ (resp. $(\eta_1, .., \eta_q, l, p)$ for $q \geq 1$). The distribution associated to the switching factor $\eta$ for GSA purposes is the Bernoulli distribution $B(1/2)$ (resp. $q$ independent Bernoulli distributions for $q$ switching factors).

*2.3. Steps of the approach*

The numerical computations required to apply the method can be summarized by the following steps:

1. Application of the filter(s) on the functional data
2. Implementation of the new model $\tilde{f}(\eta, l, p)$ from $f(X^l, p)$
3. Global sensitivity analysis of $\tilde{f}(\eta, l, p)$ using a classical method that handles continuous or discrete factors $p$ and discrete factors $\eta$ and $l$
4. Analysis of Sobol' indices of $\tilde{f}(\eta, l, p)$

These steps are also presented in Figure 1.

## 3. Analysis of total sensitivity indices (TSI) of Boolean factors in filter-based GSA

The filter-based GSA method can be used for model simplification. Indeed, if the effect of a filter, and thus the influence of the associated component, is weak, then the model inputs can be simplified w.r.t. this component.

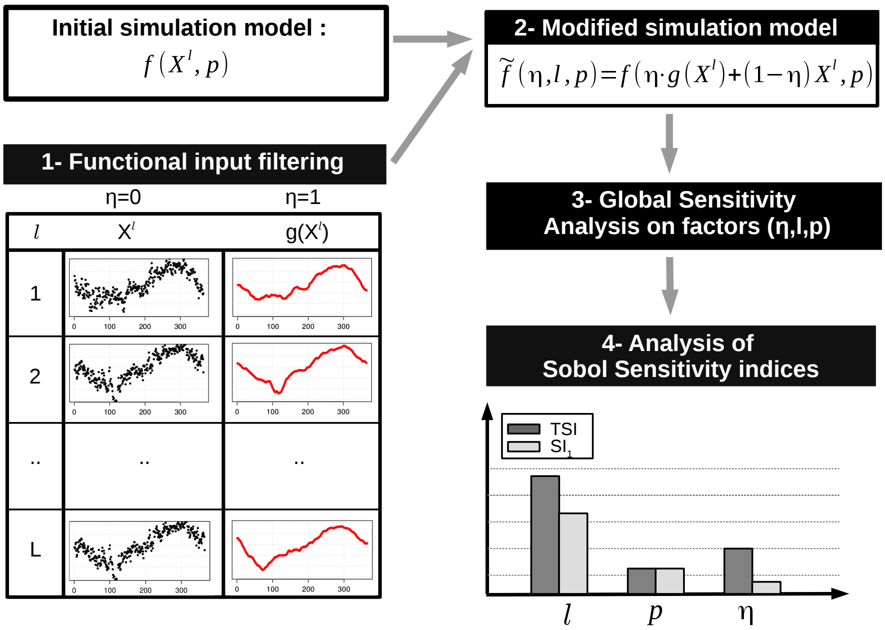


Figure 1: The main steps of the filter-based GSA method (here with a single functional input, a single filter (i.e. $q = 1$)). First the functional inputs are filtered and stored. Then a modified simulation model $\tilde{f}$ is defined using an additional Boolean factor. The global sensitivity analysis of the model response to each input factor is performed by estimating the first order sensitivity index ($SI_1$) and the total sensitivity index ($TSI$). The impact of the filter is studied using the sensitivity indices of the Boolean factor $\eta$, as explained in Section 3.

This is one way to reduce the dimensionality of a model with functional inputs. The filtered inputs may be for example easier to acquire or facilitate the mathematical analysis of the model.

We first remark that we can apply to the switching variable $\eta$ the classical result on the factor fixing problem [16] which states that a factor can be fixed without significantly impacting the variance of the output if its TSI is small. This leads to conclude that if $TSI(\eta)$ is small enough, then it is equivalent to fix $\eta = 0$ or $\eta = 1$, which means that it is equivalent to use the model with the initial functional inputs $X$ or with the filtered inputs $\tilde{X}$. But, as noted by

Sobol' et al [17] in the context of fixing non-influential scalar factors, "a limit with factor fixing is that of fixing unessential factors without knowing the magnitude of the approximation error that is being produced from sensitivity indices". These authors have established theoretical relationships between approximation errors and TSI to overcome this pitfall. In the following, we extend their results to the filter-based GSA and present relationships between TSI of switching variables and approximation errors when $q = 1$ and $q > 1$. In the latter case, we study to what extent the $2^q - 1$ possible errors corresponding to using any subset of the $q$ filters can be directly deduced from the TSI of the $q$ Boolean factors.

*3.1. Error of filtering*

We introduce the mean squared error criterion $\mathcal{E}_{g_1,..,g_q}$ to quantify the error on the model output due to the combined application of $q$ filters $g_1, .., g_q$. This error is computed relatively to the output variability induced by the variations of all the selected model inputs $X$ and $p$ as in [11, 17] where an error due to fixing scalar model parameters is defined. Let $V$ be the variance of $f$: $V = Var_{l,p}[f]$. The error of filtering $\mathcal{E}_{g_1,..,g_q}$ is defined as:

$$\mathcal{E}_{g_1,..,g_q} = \frac{1}{V} E_{l,p}\left[\left(f(X^l, p) - f(g_1 \circ .. \circ g_q(X^l), p)\right)^2\right] \tag{3}$$

where $E_{l,p}$ denotes the expectation over factors $l$ and $p$. We can rewrite the error with $\widetilde{f}$ as follows:

$$\mathcal{E}_{g_1,..,g_q} = \frac{1}{V} E_{l,p}\left[\left(\widetilde{f}(1, .., 1, l, p) - \widetilde{f}(0, .., 0, l, p)\right)^2\right] \tag{4}$$

*3.2. TSI of Boolean factors*

We first present the expression of the TSI of Boolean factors of model $\widetilde{f}$ (the proof is given in Appendix A). Let $\widetilde{V}$ denote the total variance of $\widetilde{f}$: $\widetilde{V} = Var_{\eta,l,p}[\widetilde{f}]$.

**Proposition 3.1.** *The TSI of a Boolean factor $\eta_i$ when studying a model $\widetilde{f}(\eta_1,..,\eta_q,l,p)$ can be expressed as a sum of $2^{q-1}$ mean squared differences of model evaluations differing on the i-th Boolean factor:*

$$TSI_{\eta_i} = \frac{1}{2^{q+1}\cdot\widetilde{V}} \sum_{\alpha\in\{0,1\}^{q-1}} E_{l,p}\left[\left(\begin{array}{c}\widetilde{f}(\alpha_1,..,\alpha_{i-1},1,\alpha_{i+1},..,\alpha_q,l,p)\\ -\widetilde{f}(\alpha_1,..,\alpha_{i-1},0,\alpha_{i+1},..,\alpha_q,l,p)\end{array}\right)^2\right] \tag{5}$$

### 3.3. Case of a single switching factor

When a single filter $g$ and so a single switching factor $\eta$ is considered, a simple relation between the mean squared error $\mathcal{E}_g$ and $TSI_\eta$ can be obtained:

**Proposition 3.2.** *$TSI_\eta$ and $\mathcal{E}_g$ are directly linked when $q=1$ by:*

$$\mathcal{E}_g = 4\cdot TSI_\eta \cdot \frac{\widetilde{V}}{V} \tag{6}$$

*Proof.* The result is a direct application of Eq.(5) and Eq.(4) with $q=1$:

$$TSI_\eta = \frac{1}{4\widetilde{V}} E_{l,p}\left[\left(\widetilde{f}(1,l,p)-\widetilde{f}(0,l,p)\right)^2\right] = \frac{\mathcal{E}_g}{4}\frac{V}{\widetilde{V}}$$

In the case of a single switching factor, we have a direct relation between the error due to filtering and the TSI of the switching factor. A global sensitivity analysis on $\widetilde{f}(\eta,l,p)$ completely characterizes the impact of applying a filter on the functional inputs. Both the error level and its interactions with other factors can be obtained from the sensitivity indices.

### 3.4. Case of $q$ switching factors ($q>1$)

When considering $q$ filters, $(2^q-1)$ different errors can be defined, each corresponding to a subset $(g_{i_1},..,g_{i_s})$ of $(g_1,..,g_q)$, with $s\le q$. The error of filtering associated with the application of the composed filter $g_{i_1}\circ..\circ g_{i_s}$ is defined as follows:

$$\mathcal{E}_{g_{i_1},..,g_{i_s}} = \frac{1}{V}E_{l,p}\left[\left(f(X^l,p)-f(g_{i_1}\circ..\circ g_{i_s}(X^l),p)\right)^2\right] \tag{7}$$

In the following proposition (see proof in Appendix B), we relate each of these errors to the TSI of the $q$ switching factors:

**Proposition 3.3.** *Each* $(2^q - 1)$ *error of filtering* $\mathcal{E}_{g_{i_1},..,g_{i_s}}$ *corresponding to using a composed filter* $g_{i_1},..,g_{i_s}$ *can be related to a partial sum of TSI associated to the analysis of model* $\widetilde{f}(\eta_1,..,\eta_q,l,p)$ *using the following inequality:*

$$\mathcal{E}_{g_{i_1},..,g_{i_s}} \leq \bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}} \tag{8}$$

$$\text{with } \bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}} = 2^{q+1} \cdot \frac{\widetilde{V}}{V} \cdot \sum_{i=i_1}^{i_s} TSI_{\eta_i}$$

A practical interest of Eq.(8) is that from a single sensitivity analysis performed on the model with $q$ switching variables, one can identify among the $2^q - 1$ possible combinations of filters those who have no impact on the model response. For instance, if $\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}} \simeq 0$ then one can conclude that the composed filter $g_{i_1} \circ .. \circ g_{i_s}$ has no impact on the model response. Compared to the result of Eq.(6) obtained with $q = 1$, we have lost the equality in the relation between the approximation error and TSI. Note also that given the way Eq.(8) was obtained in Appendix B using large upper bounds, it might happen that $\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}}$ largely overestimate $\mathcal{E}_{g_{i_1},..,g_{i_s}}$. This is illustrated in the following from numerical experiments.

## 4. Description of the numerical experiments

We illustrate our method on a simplified crop model. Our aim is to better understand the impact of climate on crop yield and to study the possibility of simplifying the model w.r.t. its climatic inputs. A low-pass (i.e. smoothing) filter has been applied on its climatic inputs to quantify the impact of their high frequency fluctuations on crop yield. This example aims at i) illustrating the link between error due to filtering and TSI of switching factors (Eq.(6) and Eq.(8)) and ii) checking whether the climatic data could be simplified without introducing significant level of error on the simulated crop yield.

### *4.1. The ToyCrop model*

We have defined a simple crop model called ToyCrop for the purpose of this study. Three objectives motivated the development of this toy model: i) the requirement of analysing a model with a complex structure of inputs (vector of functional inputs), ii) the need to work with a model having a very low computational cost (order of magnitude: milliseconds per simulation), iii) the need to have a model structure with very few equations allowing qualitative validation of sensitivity results.

ToyCrop main output $y$ represents the crop yield. The growth is simulated over a period corresponding to several months and is affected by two limiting factors: stress to high temperatures and stress due to water scarcity. The complete definition of the model is given in Appendix C. For the purpose of illustrating the filter-based GSA approach, we consider a case study where ToyCrop response of interest is of the form:

$$y = f_{\text{toyC}}\left(X^l, p\right)$$

$$\text{with: } \begin{cases} X^l = \left(x^l_{rain}, x^l_{rad}, x^l_{temp}, x^l_{et0}\right) \\ p = \left(\tau_{tmoy}, TTSW, k_c\right) \end{cases}$$

The vector of functional inputs $X^l$ is made of four climatic variables discretized at a daily time step and sampled for year $l$ during several months: rain $x^l_{rain}(t)$, incoming solar radiation $x^l_{rad}(t)$, mean temperature $x^l_{temp}(t)$ and evapotranspiration $x^l_{et0}(t)$. The vector $p$ of independent scalar inputs is made of three parameters that affect the sensitivity to temperature and water stress: $\tau_{tmoy}$ (threshold for high temperature stress), $TTSW$ (Total Transpirable Soil Water) and $k_c$ (crop coefficient). More details are provided in Appendix C.

*4.2. Global impact of climatic variables*

We studied the ToyCrop model and the impact of its climatic inputs on a domain defined in Table 1. As for many crop models, it is expected that

| Factor | Definition | Distribution |
|---|---|---|
| $p_1(=\tau_{tmoy})$ | Threshold for high temperature stress | $U[20, 30]$ |
| $p_2(=TTSW)$ | Total Transpirable soil water | $U[100, 250]$ |
| $p_3(=k_c)$ | Cultural coefficient (Transpiration rate) | $U[0.5, 0.8]$ |
| $l$ | Label for climatic year | $DU(42)$ |

Table 1: Definition of the ToyCrop input factors probability distributions. $U$ stands for Uniform distribution and $DU$ for Discrete Uniform distribution. The label $l$ has a Discrete Uniform distribution $DU(42)$: it is used to take samples in a set of 42 historical climatic years.

the climatic variables have a strong global impact on the model output, as they do on crop yield in real life. This property is checked by preliminarily computing the TSI of the labeling variable $l$ (see Section 5.1).

*4.3. Low-pass filter*

The low-pass filter employed averages the temporal inputs over a sliding window of size $2d + 1$ (d=15 days in the numerical application). Its application on a single temporal input $x(t)$ gives:

$$g_{\text{lowP}}(x)(t) = \frac{1}{2d+1} \sum_{k=-d}^{d} x(t+k)$$

The effect of the application of this filter is to remove the high frequency fluctuations of the functional input. We define a low-pass filter acting on the vector $X = [x_1(t), .., x_N(t)]$ of functional variables in the following way:

$$g^i_{\text{lowP}}(X)(t) = [x_1(t), .., x_{i-1}(t), g_{\text{lowP}}(x_i(t)), x_{i+1}(t), .., x_N(t)]$$

The application of filter $g_{\text{lowP}}$ to each climatic variable is shown in Figure 2. The impact of the filter depends on the type of climatic variable on which it is applied: the filtered temperature is relatively close to the original series. This is also globally true for evapotranspiration and radiation but with more dispersion around the filtered signals. Rain however is poorly approximated due to its characteristic in Mediterranean climates (rare and stormy rainfall events). Due to these approximations and to the form of the model equations (See Appendix C), it is thus expected that the filter effect will depend on the climatic year and that the effect of filtering rain will prevail over the other effects. One objective of the numerical experiments is to show that we are able to quantify these expected properties. We design two different numerical experiments on the ToyCrop model to test the impact of applying the low-pass filter on the climatic inputs and to illustrate the link between TSI and normalized mean squared errors.

### *4.4. First numerical experiment ($q = 1$)*

In the first experiment, we use a single switching factor $\eta_{all}$ to trigger the simultaneous application of the low-pass filter $g_{lowP}$ on the four climatic variables. This is obtained using the composed filter $g^{all}_{lowP} = g^1_{\text{lowP}} \circ g^2_{\text{lowP}} \circ g^3_{\text{lowP}} \circ g^4_{\text{lowP}}$. We thus define the model $\widetilde{f}_{\text{toyC}}(\eta_{all}, l, p)$:

$$\widetilde{f}_{\text{toyC}}(\eta_{all}, l, p) = \begin{cases} f_{\text{toyC}}\left(X^l, p\right) & \text{if } \eta_{all} = 0 \\ f_{\text{toyC}}\left(g^{all}_{lowP}(X^l), p\right) & \text{if } \eta_{all} = 1 \end{cases}$$

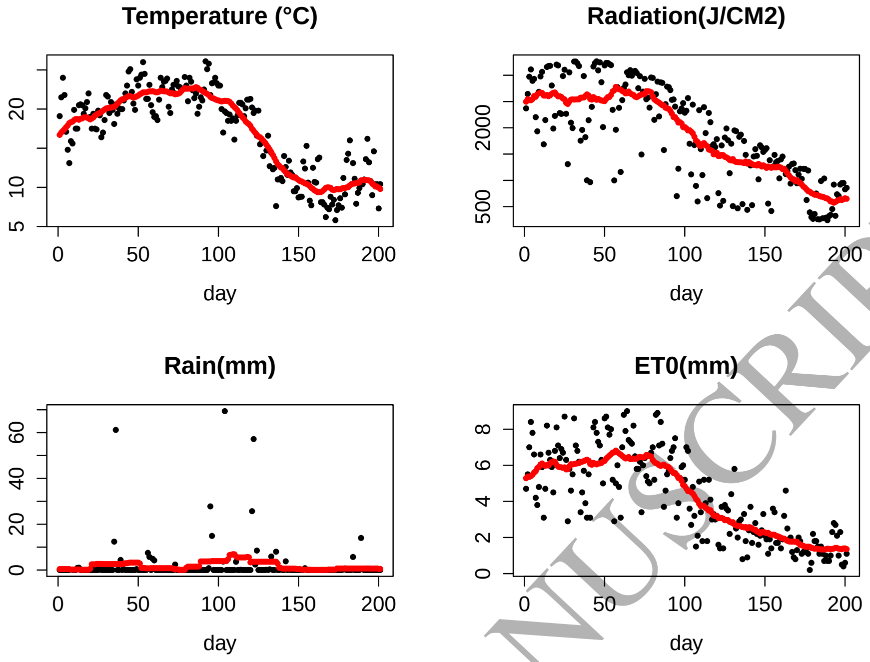


Figure 2: Result of the application of the low-pass filter (pure averaging over a $2d+1 = 31$ days) on four climatic variables: temperature, radiation, rain, evapotranpiration ($ET_0$). The filtered variables are plotted in red, the original in black.

### *4.5. Second numerical experiment ($q = 4$)*

In the second experiment we use four switching factors to trigger the application of low-pass filters independently on the four climatic variables and compute their TSI. The new model on which sensitivity analysis is performed in that case is the following:

$$\widetilde{f}_{\text{toyC}}\left(\eta_{rain}, \eta_{rad}, \eta_{temp}, \eta_{et0}, l, p\right) = f_{\text{toyC}}\left(\bar{g}_1 \circ .. \circ \bar{g}_4(X^l), p\right) \qquad (9)$$

$$\text{with } \bar{g}_i(X^l) = \begin{cases} X^l & \text{if } \eta_i = 0 \\ g^i_{\text{lowP}}(X^l) & \text{if } \eta_i = 1 \end{cases}$$

*4.6. Details of the numerical computations*

We follow the algorithmic steps described in section 2.3. The first and the total sensitivity indices are estimated using a Sobol' algorithm (SobolJansen from the Sensitivity package [18]) with an input size $N = 5000$ and 100 replicates for the bootstrap step.

We also perform additional computations to gain more insights into the theoretical results presented in Section 3. We estimate independently of the sensitivity experiment several errors of filtering as defined in Eq.(3) and Eq.(7). For the first experiment ($q = 1$) there is only one such error to compute denoted by $\mathcal{E}_{g^{all}}$ . For the second ($q = 4$) we compute 7 errors among the $2^4 - 1 = 15$ possible combinations. We select i) the 4 errors associated to filtering a single variable: $\mathcal{E}_{g_1}, \mathcal{E}_{g_2}, \mathcal{E}_{g_3}, \mathcal{E}_{g_4}$; ii) the error associated to filtering the variables 2 and 4: $\mathcal{E}_{g_2,g_4}$; iii) the error associated to filtering the last three variables: $\mathcal{E}_{g_2,g_3,g_4}$ , and iv) the error of filtering all the four variables: $\mathcal{E}_{g_1,g_2,g_3,g_4}$. These errors are computed using a direct Monte Carlo estimation based on 42000 samples and ensuring equal representation of each climatic year.

We compute the right hand side of Eq.(6) and Eq.(8) which are bounds of the filtering error (with equality for $q = 1$) deduced from the TSI values. These estimates are:

$$\bar{\mathcal{E}}_{g^{all}} = 4 \cdot TSI_{\eta_{all}} \cdot \frac{\tilde{V}}{V}$$
$$\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}} = 2^{q+1} \sum_{i=i_1}^{i_s} TSI_{\eta_i} \cdot \frac{\tilde{V}}{V} \tag{10}$$

We also analyze more precisely Eq.(8) by computing all quantities involved in the decomposition of $\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}}$ : i) the partial sum of TSI, ii) the ratio of variances $\widetilde{V}/V$ and iii) the quantity noted $C_{i_1,..,i_s}$ that should be multiplied by $\sum_{i=i_1}^{i_s} TSI_{\eta_i} \cdot \widetilde{V}/V$ to exactly obtain the mean squared error $\mathcal{E}_{g_{i_1},..,g_{i_s}}$:

$$C_{i_1,..,i_s} = \frac{\mathcal{E}_{g_{i_1},..,g_{i_s}}}{\sum_{i=i_1}^{i_s} TSI_{\eta_i} \cdot \frac{\widetilde{V}}{V}} \tag{11}$$

## 5. Results of the numerical experiments

### *5.1. Global effect of climatic inputs*

The result of the global sensitivity analysis on non filtered data is presented in Figure 3. We can see that the TSI of factor $l$ corresponding to the label of the climatic variables is high ($\approx 0.7$), which confirms that the global effect of climatic variables on the model output is high and that a complementary analysis of this global effect makes sense on this model.

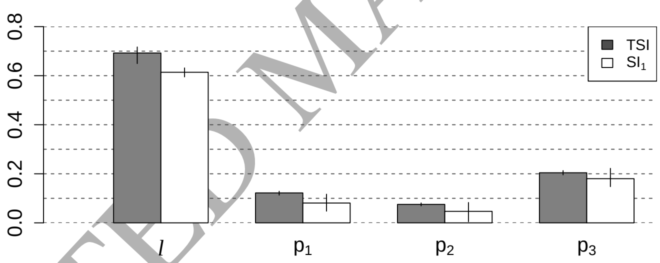


Figure 3: First-order ($SI_1$) and total sensitivity indices (TSI) obtained using a labeling technique to characterize the global effect of climatic variables. This effect is quantified using the sensitivity indices of factor $l$ representing the index of the weather series.

### *5.2. Results of the low-pass filter application with one switching factor*

The estimated sensitivity indices with the low-pass filter applied on all climatic variables are presented in Figure 4. The switching factor $\eta_{all}$ has a rather high TSI $0.166 \pm 0.01$, the highest after the TSI of the climate label factor $l$ which dominates the sensitivity results ($TSI_l \approx 0.6$). This suggests that the high frequency content, i.e. rapid fluctuations, of climatic inputs has a non negligible impact on crop yield. This result is not surprising because of the strong effect of the filter on some variables particularly on ($x_{rain}$, $x_{et0}$) and of the fact that the water stress definition is strongly influenced by the $x_{rain}$ (see Appendix C). A qualitative analysis based on the value of $TSI_{\eta_{all}}$ would lead to reject the simultaneous simplification of all ToyCrop climatic inputs by a smooth approximation: there is an influential information in the high frequency fluctuations of the climatic inputs that should be preserved to keep precise predictions. However, it is not possible to guess from this numerical experiment to which climatic variable these important components belong to.

The second objective of this numerical experiment is to illustrate and verify Eq.(6). This is done in Table 2. The two independent estimations of the mean squared error, the direct one $\mathcal{E}_{g_{all}}$ and the other $\bar{\mathcal{E}}_{g_{all}}$ deduced from the TSI value, are virtually equal as expected from Eq.(6). The filtering error is about 0.79 which is a very large value confirming that climatic inputs should not all be replaced simultaneously by a smooth approximation.

| $\mathcal{E}_{g_{all}}$ | $TSI$ | $\tilde{V}/V$ | $\bar{\mathcal{E}}_{g_{all}}$ |
|---|---|---|---|
| 0.797 | 0.166 $\pm$0.01 | 1.141 | 0.759 $\pm$0.044 |

Table 2: Numerical verification of the link (Eq.6) between the normalized mean squared error $\mathcal{E}_{g_{all}}$ and the TSI of the single switching factor for first numerical experiment. Confidence intervals are indicated when possible.

### *5.3. Results of low-pass filter application with four switching factors*

The sensitivity indices for the second numerical experiment involving combined filters are presented in Figure 5. They allow to differentiate the effect of filtering the different variables and illustrate the result of Eq.(8).

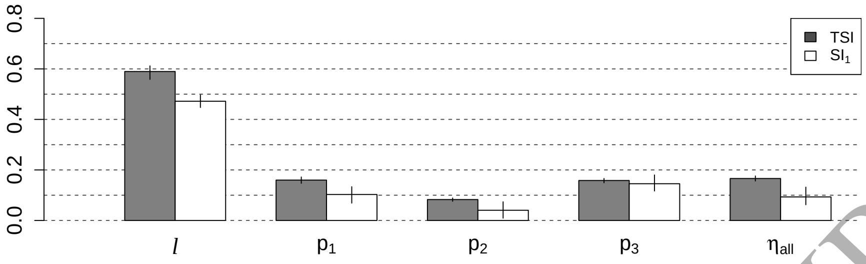


Figure 4: First-order ($SI_1$) and total sensitivity indices (TSI) obtained in the first numerical experiment when using a single switching factor to trigger the application of a single filter on the four climatic variables.

Unlike the previous experiment, we obtain very small TSI for $\eta_{rad}$, $\eta_{temp}$ and $\eta_{et0}$, the largest value being associated to $\eta_{temp}$ ($0.0066 \pm 0.0007$). On the other hand $\eta_{rain}$ has a much larger TSI ($0.1481 \pm 0.0124$). This result strongly suggests that high frequency fluctuations of solar radiation, temperature and evapotranspiration have weak impact on crop yield and that on the contrary the high frequency content of rain is important. We can compute, by using Eq.(8) the filtering error upper bounds $\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}}$ of all combined filters $g^{i_1}_{\mathrm{lowP}} \circ .. \circ g^{i_s}_{\mathrm{lowP}}$ in order to get a more quantitative view of the errors that would be made by simplifying the climatic variables using a smooth approximation. These values, which are directly obtained from the TSI of the sensitivity experiment, are also compared with direct independent estimations of the filtering errors $\mathcal{E}_{g_{i_1},..,g_{i_s}}$ in Table 3. As $\mathcal{E}_{g_{i_1},..,g_{i_s}} \leq \bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}}$, we can conclude from the sole knowledge of the TSI that filtering the second and fourth variables ($x_{rad}$ and $x_{et0}$) alone or together leads to a low error on yield prediction, as $\bar{\mathcal{E}}_{g_2}, \bar{\mathcal{E}}_{g_4}$ and $\bar{\mathcal{E}}_{g_2,g_4}$ are low (Table 3). The error bounds associated to filtering the third variable $x_{temp}$ alone or simultaneously with $x_{rad}$ and $x_{et0}$ are moderately high (0.243 and 0.256 respectively). Basing conclusions only on the sensitivity experiment and computations of $\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}}$, we cannot be sure that the effect of filtering these three variables is small enough to simplify them by smooth approximations. However the error bounds allow to make the hypothesis that the filtering error is very low since we know that the bound in Eq.(8) is likely pessimistic in practice. Actually, for this experiment and using the independent computation of the errors, we find that the effective ratio between error and partial sums of TSI (taking into account the ratio $\widetilde{V}/V = 1.14$), denoted by $C_{i_1,..,i_s}$ (see Eq.(11)), is very small compared to $2^{q+1} = 32$: it lies between 3 and 6 (see Table 3). The additional numerical experiment required to validate the hypothesis that three variables can be simplified simultaneously is a single-switch GSA with filter $g^2_{\mathrm{lowP}} \circ g^3_{\mathrm{lowP}} \circ g^4_{\mathrm{lowP}}$ . The result of this experiment (not shown) is a negligible error level of $\mathcal{E}_{g_2,g_3,g_4} = 0.021$ obtained using Eq.(6) and the TSI of the switching factor ($TSI = 0.005 \pm 0.001$).

*5.4. Conclusion on the numerical experiments*

Using filter-based GSA we were able to show that high frequency fluctuations of mean temperature, solar radiation and evapotranspiration have a negligible effect on the simulation of crop yield. It is not the case for rain as expected. Using the theoretical link between TSI and model error developed in section 3, we concluded from the 4-switch experiment that solar radiation and evapotranspiration can be replaced by a smooth approximation without increasing significantly the level of model error. An additional experiment with a single switch was required to confirm the hypothesis that the combined filtering of $(x_{temp}, x_{rad}, x_{et0})$ also induces a very low error on crop yield. Apart from a better understanding on the behavior of the ToyCrop model and a reduction of its dimensionality, a practical application of this result is related to the missing data problem. It shows that missing data could be replaced by smooth interpolation data without impacting model results. Indeed, smoothing only alters high frequency content of a signal which was proven not sensitive.

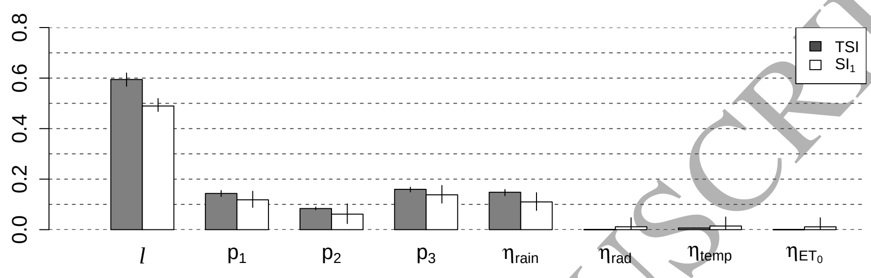


Figure 5: First-order ($SI_1$) and total sensitivity indices ($TSI$) obtained in the second numerical experiment: four switching factors are used to trigger the application of 4 filters on the 4 climatic variables.

| $i_1, .., i_s$ | $\mathcal{E}_{g_{i_1},..,g_{i_s}}$ | $\sum_{i=i_1}^{i_s} TSI_{\eta_i}$ | $\bar{\mathcal{E}}_{g_{i_1},..,g_{i_s}}$ | $C_{i_1,..,i_s}$ |
|---|---|---|---|---|
| $\{1\}$ | 0.6546 | 0.1481 ±0.0124 | 5.373 ±0.338 | 3.91 ±0.25 |
| $\{2\}$ | 0.0012 | 0.0003 ±0.0000 | 0.010 ±0.001 | 4.12 ±0.27 |
| $\{3\}$ | 0.0228 | 0.0066 ±0.0007 | 0.243 ±0.027 | 3.05 ±0.34 |
| $\{4\}$ | 0.0006 | 0.0001 ±0.0000 | 0.004 ±0.000 | 4.51 ±0.35 |
| $\{2, 4\}$ | 0.0017 | 0.0004 ±0.0000 | 0.014 ±0.001 | 4.07 ±0.28 |
| $\{2, 3, 4\}$ | 0.0212 | 0.0070 ±0.0008 | 0.256 ±0.028 | 2.68 ±0.30 |
| $\{1, 2, 3, 4\}$ | 0.7971 | 0.1551 ±0.0131 | 5.629 ±0.366 | 4.55 ±0.30 |

Table 3: Link between $\mathcal{E}_{g_{i_1},..,g_{i_s}}$ and $TSI$ for the second numerical experiment. $(i_1, .., i_s) \in \{1, .., 4\}$ are indices of a set of variables inside vector $(x_{rain}, x_{rad}, x_{temp}, x_{et0})$. Confidence intervals are indicated when possible.

## 6. Discussion and Conclusion

We propose an approach to analyze models with functional inputs. The method is designed to test and quantify the importance of *a priori* defined components of these functional inputs. This might be particularly useful for model understanding, simplification or validation in research areas using models with functional inputs. The method is based on the use of filters applied on the functional inputs to suppress or enhance some of their components. Sensitivity analysis of the model response to these components can then be conducted by triggering the application of the filters with Boolean factors. This approach was illustrated on a toy model simulating crop yield. In the numerical exercise, high frequencies of climatic inputs were found non-important for three out of four climatic inputs.

The proposed method is global, model free and generic in the sense that it does not depend on the type of filter applied. We used smoothing filters in the numerical example but other types of modifications can be handled in the same way. For instance filters acting specifically on high signal values or at a specific time (or space) position can be used to assess the importance of different kind of components of functional inputs.

We have shown that the mean squared error resulting from the filters application is related to the total sensitivity indices (TSI). This new result can be seen as a detailed analysis of the relationship between TSI and approximation error in a factor fixing approach ([11],[17]) when applied to a Boolean factor. A direct relation between TSI and approximation error was derived when a single filter was used, whereas only an upper bound involving a partial sum of TSI was found in the case of multiple filters. This result is particularly useful in a perspective of model simplification.

Filters allow to modify functional inputs in a very flexible way and thus to quantify the importance of diverse types of functional input components. Practitioners may however define filters that lead to unrealistic transformed datasets. This was for example the case in the presented application since the chosen filter did not respect the correlation structure between climatic inputs. It is important to note that this issue does not impact the conclusions if the method is used for model simplification as in the presented example. It is however an aspect to keep in mind when interpreting high effects of a filter application if the method is used for model behavior exploration, since the filtered functional input may be outside of the model validity domain. Concerning our numerical application, to overcome such situations, one option could be to use constrained filters that preserve the functional inputs correlation structure.

## 7. Acknowledgments

This work was part of the EA-SENSICLIM project funded by the French National Institute for Agricultural Research (INRA). The authors thank the reviewers for providing many comments and suggestions that helped improving the article.

## 8. References

## References

[1] P. Casadebaig, B. Zheng, S. Chapman, N. Huth, R. Faivre, K. Chenu, Assessment of the Potential Impacts of Wheat Plant Traits across Environments by Combining Crop Modeling and Global Sensitivity Analysis., PloS one 11 (1).

[2] E. I. Teixeira, G. Zhao, J. de Ruiter, H. Brown, A.-G. Ausseil, E. Meenken, F. Ewert, The interactions between genotype, management and environment in regional crop modelling, European Journal of Agronomy 88 (2017) 106 – 115.

[3] J. Goffart, T. Mara, E. Wurtz, Generation of stochastic weather data for uncertainty and sensitivity analysis of a low-energy building, Journal of Building Physics.

[4] M. Ruiz-Ramos, R. Ferrise, A. Rodríguez, I. J. Lorite, M. Bindi, T. R. Carter, S. Fronzek, T. Palosuo, N. Pirttioja, P. Baranowski, et al., Adaptation response surfaces for managing wheat under perturbed climate and co2 in a mediterranean environment, Agricultural Systems (2017) In Press.

[5] S. Nanty, C. Helbert, A. Marrel, N. Prot, C. Prieur, Sampling, metamodeling, and sensitivity analysis of numerical simulators with functional stochastic inputs, SIAM/ASA Journal on Uncertainty Quantification 4 (1) (2016) 636–659.

[6] P. Ruffo, L. Bazzana, A. Consonni, A. Corradi, A. Saltelli, S. Tarantola, Hydrocarbon exploration risk evaluation through uncertainty and sensitivity analyses techniques, Reliability Engineering & System Safety 91 (10) (2006) 1155–1162.

[7] J. Fruth, O. Roustant, S. Kuhnt, Sequential designs for sensitivity analysis of functional inputs in computer experiments, Reliability Engineering & System Safety 134 (2015) 260–267.

[8] L. Lilburne, S. Tarantola, Sensitivity analysis of spatial models, Inter-

national Journal of Geographical Information Science 23 (2) (2009) 151–168.

[9] F. Anstett-Collin, J. Goffart, T. Mara, L. Denis-Vidal, Sensitivity analysis of complex models: Coping with dynamic and static inputs, Reliability Engineering and System Safety 134 (2014) 268–275.

[10] B. Iooss, M. Ribatet, Global sensitivity analysis of computer models with functional inputs, Reliability Engineering & System Safety 94 (7) (2009) 1194–1204.

[11] I. M. Sobol, Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiments 1 (4) (1993) 407–414.

[12] V. Picheny, R. Servien, N. Villa-Vialaneix, Interpretable sparse SIR for functional data, Statistics and Computing (2018) 1–13.

[13] A. Saltelli, K. Chan, E. M. Scott, et al., Sensitivity analysis, Wiley New York, 2000.

[14] G. Baroni, a. Facchi, C. Gandolfi, B. Ortuani, D. Horeschi, J. C. van Dam, Uncertainty in the determination of soil hydraulic parameters and its influence on the performance of two hydrological models of different complexity, Hydrology and Earth System Sciences 14 (2) (2010) 251–270.

[15] M. Crosetto, S. Tarantola, Uncertainty and sensitivity analysis: tools for gis-based model implementation, International Journal of Geographical Information Science 15 (5) (2001) 415–437.

[16] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, Sensitivity analysis in practice: a guide to assessing scientific models, John Wiley & Sons, 2004.

[17] I. Sobol, S. Tarantola, D. Gatelli, S. Kucherenko, W. Mauntz, Estimating the approximation error when fixing unessential factors in global sensitivity analysis, Reliability Engineering & System Safety 92 (7) (2007) 957 – 960.

[18] G. Pujol, B. Iooss, A. Janon, Sensitivity: Global sensitivity analysis of model outputs, R package https://cran.r-project.or/package=sensitivity.

[19] N. Brisson, J. Wery, K. Boote, Fundamental concepts of crop models illustrated by a comparative approach, Working with Dynamic Crop Models, Elsevier, Amsterdam (2006) 257–279.

[20] J. Monteith, Solar radiation and productivity in tropical ecosystems, Journal of Applied Ecology 9 (3) (1972) 747–766.

[21] T. Sinclair, Water and nitrogen limitations in soybean grain production i. model development, Field Crops Research 15 (2) (1986) 125–141.

[22] N. Brisson, C. Gary, E. Justes, R. Roche, B. Mary, D. Ripoche, D. Zimmer, J. Sierra, P. Bertuzzi, P. Burger, et al., An overview of the crop model stics, European Journal of agronomy 18 (3) (2003) 309–332.

[23] E. Coucheney, S. Buis, M. Launay, J. Constantin, B. Mary, I. García de Cortázar-Atauri, D. Ripoche, N. Beaudoin, F. Ruget, K. S. Andrianarisoa, C. Le Bas, E. Justes, J. Léonard, Accuracy, robustness and

behavior of the STICS soilcrop model for plant, water and nitrogen outputs: Evaluation over a wide range of agro-environmental conditions in France, Environmental Modelling & Software 64 (2015) 177–190.

[24] E. Lebon, V. Dumas, P. Pieri, H. R. Schultz, Modelling the seasonal dynamics of the soil water balance of vineyards, Functional Plant Biology 30 (6) (2003) 699–710.

[25] S. Roux, X. Delpuech, G. Daudin, F. Brun, J. Wery, D. Wallach, Providing user-oriented uncertainty information with a vineyard model used for irrigation decisions, Practical Applications of Agricultural System Models to Optimize the Use of Limited Water (2014) 183–208.

[26] NRCS, National Engineering Handbook. Part 630 Hydrology, National Resources Conservation Service, USDA, Washington, DC., 2004.

## A. Proof of Eq.(5)

To establish Eq.(5), we will use two properties of Boolean random variables:

- the variance of $Z \sim B(1/2)$ taking values $z_0$ and $z_1$:

$$Var[Z] = \left(\frac{z_0 - z_1}{2}\right)^2 \tag{A.1}$$

- the expectation of a random variable $h(\eta_1, .., \eta_n)$ where all $\eta_i$ are independent and have a Bernoulli distribution $B(1/2)$:

$$E\left[h(\eta_1, .., \eta_n)\right] = \frac{1}{2^n} \sum_{(\alpha_1,..,\alpha_n)\in\{0,1\}^n} h(\alpha_1, .., \alpha_n) \tag{A.2}$$

We write $\eta_{-i} = (\eta_1, .., \eta_{i-1}, \eta_{i+1}, .., \eta_q)$ and $X_{-\eta_i} = (\eta_{-i}, l, p)$. Then starting from the definition of the TSI, we can write:

$$\begin{aligned} TSI_{\eta_i} &= \frac{1}{\widetilde{V}} E\left[Var(Y|X_{-\eta_i})\right] \\ &= \frac{1}{\widetilde{V}} E_{\eta_{-i},l,p}\left[Var_{\eta_i}(\widetilde{f}(\eta_1, .., \eta_q, l, p)\right] \end{aligned} \quad \text{(A.3)}$$

Using Eq.(A.1) in Eq.(A.3), with $Z = \widetilde{f}(\eta_1, .., \eta_q, l, p)$ we can write:

$$\begin{aligned} TSI_{\eta_i} &= \frac{1}{\widetilde{V}} E_{\eta_{-i},l,p}\left[\left(\begin{array}{c} \frac{1}{2}(\widetilde{f}(\eta_1, .., \eta_{i-1}, 1, \eta_{i+1}, .., \eta_q, l, p) \\ -\widetilde{f}(\eta_1, .., \eta_{i-1}, 0, \eta_{i+1}, .., \eta_q, l, p)) \end{array}\right)^2\right] \\ &= \frac{1}{\widetilde{V}} E_{\eta_{-i}}\left[E_{l,p}\left[\left(\begin{array}{c} \frac{1}{2}(\widetilde{f}(\eta_1, .., \eta_{i-1}, 1, \eta_{i+1}, .., \eta_q, l, p) \\ -\widetilde{f}(\eta_1, .., \eta_{i-1}, 0, \eta_{i+1}, .., \eta_q, l, p)) \end{array}\right)^2\right]\right] \end{aligned}$$

Finally, we use Eq.(A.2) to simplify the previous expression by taking:

$$h(\eta_1, .., \eta_{i-1}, \eta_{i+1}, .., \eta_q) = E_{l,p}\left[\left(\begin{array}{c} \frac{1}{2}(\widetilde{f}(\eta_1, .., \eta_{i-1}, 1, \eta_{i+1}, .., \eta_q, l, p) \\ -\widetilde{f}(\eta_1, .., \eta_{i-1}, 0, \eta_{i+1}, .., \eta_q, l, p)) \end{array}\right)^2\right]$$

This yields exactly to Eq.(5).

## B. Proof of Eq.(8)

We start with the case $s = q$. The general case $s < q$ can be directly deduced from it. The main idea comes from the comparison of $\mathcal{E}_{g_1,..g_q}$ (Eq. (4)) and the expression of $TSI_{\eta_i}$ (Eq.(5)). The error $\mathcal{E}_{g_1,..g_q}$ is defined using the difference $\widetilde{f}(1, .., 1, l, p) - \widetilde{f}(0, .., 0, l, p)$ where $q$ bits are different, whereas $TSI_{\eta_i}$ is defined as a sum of terms differing only on one bit. So the principle of the proof is to decompose the difference on $q$ bits as several sums of one-bit differences. Such decompositions should also use a maximum of these one-bit differences in order to make a efficient use of the TSI definition.

Let us define the $q^2$ differences $(\Delta_{ij}\widetilde{f}(l,p))_{i,j\in\{1..q\}}$ by:

$$\Delta_{ij}\widetilde{f}(l,p) = \begin{cases} \widetilde{f}(1,..,1,\eta_j = 1,1,..,1,l,p) \\ \qquad -\widetilde{f}(1,..,1,\eta_j = 0,1,..,1,l,p) & \text{if } j = i \\ \widetilde{f}(1,..,1,\eta_i = 0,0,..,0,\eta_j = 1,1,..,1,l,p) \\ \qquad -\widetilde{f}(1,..,1,\eta_i = 0,0,..,0,\eta_j = 0,1,..,1,l,p) & \text{if } j > i \\ \widetilde{f},(0,..,0,\eta_j = 1,1,..,1,\eta_i = 0,0,..,0,l,p) \\ \qquad -\widetilde{f}(0,..,0,\eta_j = 0,1,..,1,\eta_i = 0,0,..,0,l,p) & \text{if } j < i \end{cases}$$

where the only difference if $i \leq j$ (resp. $j < i-1$) is on the j-th Boolean variable, where Boolean factors for $k = 1,..,i-1$ and $k = j+1,..,q$ are set to 1 (resp. $k = j+1,..,i-1$). We can use these one-bit differences to build $q$ decompositions of $\widetilde{f}(1,..,1,l,p) - \widetilde{f}(0,..,0,l,p)$:

$$\widetilde{f}(1,..,1,l,p) - \widetilde{f}(0,..,0,l,p) = \sum_{k=i}^{i+q-1} \Delta_{i,1+mod(k-1,q)}\widetilde{f}(l,p) \text{ for } i = 1..q$$

where $mod(k-1,q)$ is equal to $k-1$ modulo $q$. The sum of the previous decompositions yields:

$$q\left(\widetilde{f}(1,..,1,l,p) - \widetilde{f}(0,..,0,l,p)\right) = \sum_{i=1}^{q}\sum_{k=i}^{i+q-1} \Delta_{i,1+mod(k-1,q)}\widetilde{f}(l,p)$$

Each difference $\Delta_{ij}\widetilde{f}(l,p)$ occurs only in the ith decomposition, hence occurs exactly once in the sum of the decompositions. So we are able to re-number these $q^2$ differences as $\Delta_1\widetilde{f}(l,p),..,\Delta_{q^2}\widetilde{f}(l,p)$ and to write:

$$q\left(\widetilde{f}(1,..,1,l,p) - \widetilde{f}(0,..,0,l,p)\right) = \sum_{k=1}^{q^2} \Delta_k\widetilde{f}(l,p)$$

Then we have:

$$\left(\widetilde{f}(1,..,1,l,p)-\widetilde{f}(0,..,0,l,p)\right)^2=\frac{1}{q^2}\left[\sum_{k=1}^{q^2}(\Delta_k\widetilde{f})^2+2\sum_{k=1}^{q^2}\sum_{j>k}\Delta_k\widetilde{f}.\Delta_j\widetilde{f}\right](l,p)$$

$$\leq\frac{1}{q^2}\left[\sum_{k=1}^{q^2}(\Delta_k\widetilde{f})^2+\sum_{k=1}^{q^2}\sum_{j>k}((\Delta_k\widetilde{f})^2+(\Delta_j\widetilde{f})^2)\right](l,p)=\sum_{k=1}^{q^2}(\Delta_k\widetilde{f})^2(l,p)$$

Each $q^2$ square difference $(\Delta_k\widetilde{f})^2(l,p)$ occurs once in the sum:

$$\sum_{i=1}^{q}\sum_{\eta_{-i}\in\{0,1\}^{q-1}}\left(\widetilde{f}(\eta_1,..,1,..,\eta_q,l,p)-\widetilde{f}(\eta_1,..,0,..,\eta_q,l,p)\right)^2$$

We deduce the relation:

$$E_{l,p}\left[\left(\widetilde{f}(1,..,1,l,p)-\widetilde{f}(0,..,0,l,p)\right)^2\right]\leq E_{l,p}\left[\sum_{i=1}^{q^2}(\Delta_i\widetilde{f})^2(l,p)\right]$$

$$\leq\tilde{V}\cdot 2^{q+1}\sum_{k=1}^{q}TSI_{\eta_k}$$

We can then conclude that:

$$\mathcal{E}_{g_1,..,g_q}=\frac{1}{V}E_{l,p}\left[\left(f(X^l,p)-f(g_1\circ..\circ g_q(X^l),p)\right)^2\right]$$

$$\mathcal{E}_{g_1,..,g_q}\leq 2^{q+1}\cdot\frac{\tilde{V}}{V}\cdot\sum_{i=1}^{q}TSI_{\eta_i}$$

If we want to consider the mean squared error $\mathcal{E}_{g_{i_1},..,g_{i_s}}$ due to input simplification of the switching variables $i_1,..,i_s$, we build the decompositions associated to the restriction of $\widetilde{f}$ to variables $i_1,..,i_s$ and make a similar reasoning to obtain:

$$\mathcal{E}_{g_{i_1},..,g_{i_s}}\leq 2^{q+1}\cdot\frac{\tilde{V}}{V}\cdot\sum_{i=i_1}^{i_s}TSI_{\eta_i}$$

## C. ToyCrop Model complete definition

The ToyCrop main output that is, crop yield, is affected by two limiting factors: high temperature and water scarcity. The model has four climatic inputs discretized at a daily time step and sampled during several months and over several years $l = 1, 2, .., 42$. These inputs are rain $x^l_{rain}(t)$, solar radiation $x^l_{rad}(t)$, mean temperature $x^l_{temp}(t)$ and evapotranspiration $x^l_{et0}(t)$. Crop yield in this model is defined as a weighted sum over each day of the simulated period of the solar radiation $x^l_{rad}(t)$. Compared to the formalisms of classical complex crop models [19], it can be noted that ToyCrop makes use of the Radiation Use Efficiency (RUE) concept [20, 21] used in many crop models, as for example the STICS crop model [22, 23]. Moreover, ToyCrop uses a water balance equation to compute a water stress index as in some vineyards studies [24, 25]. The detailed definition of model outputs, inputs and equations are presented in the following.

*C.1. Model inputs*

| Name | Type | Definition | Unit |
|---|---|---|---|
| $x_{rad}$ | functional | Daily incoming radiation | $Jcm^{-2}$ |
| $x_{rain}$ | functional | Daily rain | mm |
| $x_{tmoy}$ | functional | Daily mean air temperature | °C |
| $x_{et0}$ | functional | Daily evapotranspiration demand | mm |
| $t_1$ | scalar | Starting date for biomass growth | Day of Year |
| $t_2$ | scalar | Harvest date | Day of Year |
| $\tau_{tmoy}$ | scalar | Threshold for heat stress | °C |
| $\tau_{FTSW}$ | scalar | Threshold for water stress | no unit |
| $k_c$ | scalar | Crop coefficient | no unit |
| TTSW | scalar | Total transpirable soil water | mm |
| $RS$ | scalar | Runoff strength | no unit |
| $b_0$ | scalar | Biomass at simulation start | g |
| $atsw_0$ | scalar | Water at simulation start | mm |

*C.2. State variables*

| Name | Definition |
|---|---|
| $\Delta b^t$ | Daily biomass increment |
| $RUE^t$ | Daily radiation use efficiency |
| $ATSW^t$ | Daily available transpirable soil water |
| $FTSW^t$ | Daily fraction of transpirable soil water |

*C.3. Model equations*

$$b(t) = \sum_{s=t_1}^{t} \Delta b^s$$

$$y = b(t_2)$$

$$\Delta b^t = RUE^t \cdot x_{rad}^t$$

$$RUE^t = \mathbb{1}_{x_{tmoy}^t < \tau_{tmoy}} \cdot \min\left(1, \frac{FTSW^{t-1}}{\tau_{FTSW}}\right)$$

$$ATSW_{tmp}^t = ATSW^{t-1} + x_{rain}^t - Q\left(RS, x_{rain}^{t-5,.,t}\right) - k_c \cdot x_{et0}^t \cdot \min\left(1, \frac{FTSW^{t-1}}{\tau_{FTSW}}\right)$$

$$ATSW^t = \min\left(TTSW, \max(0, ATSW_{tmp}^t)\right)$$

$$FTSW^t = \frac{ATSW^t}{TTSW}$$

The only definition remaining for a complete description of the ToyCrop model is the runoff function $Q\left(RS, x_{rain}^{t-5,.,t}\right)$. This function is inspired by the Curve Number method [26] used in many crop model applications. The principle is to define from the Runoff Strength (RS) parameter three values for the soil retention capacity ($S$) corresponding to wet, medium and dry antecedent runoff conditions. The choice between these states depends on the rainfall accumulated during the five previous days denoted by $x_{rain}^{t-5,.,t}$. Runoff $Q$ is determined from $S$ and $x_{rain}$ by using the following expressions [26]:

$$Q = \begin{cases} 0 & \text{if } x_{rain} < 0.2S \\ \frac{(x_{rain}-0.2S)^2}{x_{rain}+0.8S} & \text{if } x_{rain} \geq 0.2S \end{cases}$$

The computation of $S$ from $x_{rain}^{t-5,.,t}$ and $RS$ is given by:

$$S = \begin{cases} 254(100/cn1 - 1) & \text{if } x_{rain}^{t-5,.,t} < 12.7 \\ 254(100/cn3 - 1) & \text{if } x_{rain}^{t-5,.,t} \geq 28 \\ 254(100/cn - 1) & \text{otherwise} \end{cases}$$

with:

$$\begin{cases} cn & = 95\ RS + 30\ (1 - RS) \\ cn1 & = 4.2\ cn/(10 - 0.058\ cn) \\ cn3 & = 23\ cn/(10 + 0.13\ cn) \end{cases}$$

Even if this definition is less intuitive than the others, introducing a non linear runoff function in the ToyCrop model is interesting to test the effect of smoothing temporally the rain input $x_{rain}(t)$.