How to perform global sensitivity analysis of a catchment-scale, distributed pesticide transfer model ? Application to the PESHMELBA model.

Emilie Rouzies [1], Claire Lauvernet [1], Bruno Sudret [2], and Arthur Vidard [3]

[1]INRAE, RiverLy, Lyon-Villeurbanne, 69625 Villeurbanne Cedex, France
[2]ETH Zurich, Chair of Risk, Safety and Uncertainty Quantification, Stefano-Franscini-Platz 5, CH-8093 Zurich, Switzerland
[3]Univ. Grenoble-Alpes, Inria, CNRS, Grenoble-INP, LJK, 38000 Grenoble, France

Correspondence: Emilie Rouzies (emilie.rouzies@inrae.fr)

Abstract. Pesticide transfers in agricultural catchments are responsible for diffuse but major risks to water quality. Spatialized pesticide transfer models are useful tools to assess the impact of the structure of the landscape on water quality. Before considering using these tools in operational contexts, quantifying their uncertainties is a preliminary necessary step. In this study, we explored how global sensitivity analysis can be applied to the recent PESHMELBA pesticide transfer model to

5   quantify uncertainties on transfer simulations. We set up a virtual catchment based on a real one and we ~~compared~~ combined different approaches for sensitivity analysis that could handle the specificities of the model: high number of input parameters, limited size of sample due to computational cost and spatialized output. ~~We compared~~ After a screening step, we calculated Sobol' indices obtained from Polynomial Chaos Expansion, HSIC dependence measures and feature importance measures obtained from Random Forest surrogate model ~~. Results showed the consistency of the different methods and they highlighted~~

10   ~~the relevance of Sobol' indices to capture interactions between parameters~~to get a comprehensive overview of PESHMELBA sensitivity. Sensitivity indices were first computed for each landscape element (site sensitivity indices). Second, we proposed to aggregate them at the hillslope and the catchment scale in order to get a summary of the model sensitivity and a valuable insight into the model ~~hydrodynamical~~ hydrodynamic behaviour. The methodology proposed in this paper may be extended to other modular and distributed hydrological models as there has been a growing interest in these methods in recent years.

15 ## 1   Introduction

Pesticide transfers from fields to water bodies is a major but also complex environmental concern. Significant efforts are required to assess risks for aquatic and human lives. However, this is made difficult due to the complexity of processes at stake ~~,~~and the diversity and the fragmentation of agriculture landscapes where pesticides are applied (Campbell et al., 2004). Pesticide transfer models are essential tools to support risk management as they make simulation of contamina-

20 tion and transfers possible. They also make it possible to explore and compare alternative scenarios. However, there are many challenges for developing and using such models, highlighted for example in Gascuel-Odoux et al. (2009). In particular, to support decision-making, it is important to use physically based models ~~that are simple enough to ensure flexibility~~

~~(Reichenberger et al., 2007; Dosskey et al., 2011)~~such as in Reichenberger et al. (2007) or Dosskey et al. (2011). A large range of models exist at a local scale (Adriaanse, 1997; Muñoz-Carpena et al., 1999; Carsel and Baldwin, 2000; Larsbo and Jarvis, 2003) but modelling approaches should also address the landscape scale. Indeed, the landscape configuration is of big influence on transfers and regulation and corrective actions can relevantly be set up at this scale. Building modular, process-based and distributed models is highly appropriate to address this scale and to take into account the landscape composition, (Buytaert et al., 2008; Kraft et al., 2012). This recent but promising approach ~~is based on the coupling of~~ consists in coupling different blocks representing one or several processes~~. This kind of models have flexible structures that are particularly promising for~~ ~~various risk assessment applications such as landscape management scenario exploration. Some models already exist following~~ ~~this idea based on the use of~~, often based on a modeling framework (Tortrat, 2005; Moussa et al., 2010; Branger et al., 2010; Rouzies et al., 2019). However, simulating at such scale also implies additional difficulties. First, catchment-scale models ~~often~~ that are spatially distributed require a high computational effort that cannot always be afforded (Herman et al., 2013). Second, spatialized process-based models may be characterized by a huge number of parameters that makes the parameter setting complex. Using such models thus raises ~~new~~ challenges about diagnosing model behaviour and uncertainty quantification (Gupta et al., 2008; van Griensven et al., 2006).

To address the issue of uncertainty quantification, ~~Global Sensitivity Analysis (GSA)~~ sensitivity analysis is a powerful tool that investigates how variations in the model outputs can be attributed to variations of its input factors (Saltelli et al., 2008; Pianosi et al., 2016). It aims at understanding the model complexity~~and~~, formulating realistic assumptions for its use ~~(Faivre et al., 2013)~~ ~~. A large range of sensitivity analysis methods exist and they can be classified according to their objectives. Among others,~~ ~~*screening* methods identify (if there are) parameters that have a negligible influence on the output variability. Screening thus~~ ~~makes it possible to decrease the model complexity by setting constant values or removing non-influential parameters. *Ranking*~~ ~~methods classify input factors according to their relative contribution to output variability. In the specific case of process-based~~ ~~models, GSA may give an insight into a whole process representation (Gascuel-Odoux et al., 2009) and paves the way for~~ ~~potential improvements. Therefore, the GSA~~ and paves the way for potential improvements (Faivre et al., 2013). In the field of pesticide transfer modelling, Dubus et al. (2003) first performed a local, one-at-a-time (OAT) sensitivity analysis on four pesticide models, meaning that each parameter influence was scrutinized individually while other parameters were set to their nominal values. Although simple and computationally cheap to set, results from OAT analysis may be inaccurate in case of non linear input/output relationship (Saltelli et al., 2004; Nossent and Bauwens, 2012). As a result, global sensitivity analysis (GSA) that vary all input factors simultaneously, on their entire ranges of definition, are favoured to address the sensitivity of environmental models. A large range of GSA methods exist which are not theoretically nor practically equivalent to set and the method should be carefully chosen depending on the ~~model characteristics and the analysis purpose~~ application characteristics (Song et al., 2015; Pianosi et al., 2016; Sarrazin et al., 2016). In the case of pesticide transfer modelling, variance decomposition and the Sobol' method (Sobol, 1993) have been widely used ~~(e.g. Metta, 2007; Fox et al., 2010; Lauvernet and Muñoz-Carpena, 2018; Gate~~ (e.g. Hong and Purucker, 2018; Gatel et al., 2019). Although very popular, such methods have several limitations. First, they are characterized by a high computational cost although it can be alleviated using metamodelling techniques. For instance, computing polynomial chaos expansion ~~(Ghanem and Spanos, 1991) directly gives~~ on the model (Ghanem and Spanos, 1991)

allows to compute Sobol' sensitivity indices ~~at a low computational price~~ directly, for a limited computational cost (Sudret, 2008; Fajraoui et al., 2011; Wang et al., 2015). Second, from a methodological point of view, analysing the impact of an input factor through a variance indicator may constitute a *restrictive summary of the distribution* (Da Veiga, 2015). It can be especially unsatisfactory for highly-skewed or multimodal output variables (Liu et al., 2005; Borgonovo et al., 2011; Pianosi et al., 2016). In order to overcome such limitation, another category of methods describes the dependence between the output and each input factor from a probabilistic point of view (Székely et al., 2007; Da Veiga, 2015). In particular, in Da Veiga (2015), the Hilbert-Schmidt Independence Criterion (HSIC) proposed by Gretton et al. (2005b) is introduced to quantify the covariance between non-linear transformations of the input factor and the output variable. Although it has been used in only a few application fields so far (mainly related to risk assessment on nuclear accident contexts, see De Lozzo and Marrel 2016; Meynaoui 2019), the HSIC measure is promising and may be extended to ~~hydrological and~~ pesticide transfer modelling where the variance decomposition may not be relevant. Additionally, the growing interest for machine learning techniques is paving the way for new approaches of GSA, such as the Random Forest method (RF). Indeed, its structure provides valuable information on feature importance that can be processed as sensitivity indices like in Harper et al. (2011) and Aulia et al. (2019) (see ~~Antoniadis et al. (2021)~~ Antoniadis et al. 2021 for a review on the use of random forests for sensitivity analysis).

Such approach can be of particular interest when only a preexisting sample is available without additional information on the input.

All previous methods can be computationally costly to set in case of a distributed highly parameterized model and a common approach for GSA in complex environmental models consists in first applying a *screening* step to identify (if there are) parameters that have a negligible influence on the output variability. Screening thus makes it possible to decrease the model complexity by setting constant values or removing non-influential parameters. In a second time, a *ranking* step, based for example on one of the previous methods, classifies remaining input factors according to their relative contribution to output variability. Most screening steps for pesticide models (e.g. Fox et al., 2010; Vanuytrecht et al., 2014; Lauvernet and Muñoz-Carpena, 2018; are performed with the Elementary Effect or Morris method (Morris, 1991). However, this qualitative method is based on a visual clustering that may be fuzzy for high dimension input spaces. In these cases, a high number of levels may be necessary to make clusters appear. Again, recent alternative methods based for example on an independence statistical tests using the HSIC measure (De Lozzo and Marrel, 2014) for the screening step may be worth exploring.

Last but not least, ~~pesticide transfer~~ environmental models are often fully spatialized, meaning that the interest area is divided into spatial units on which equations are solved locally. Such specificity should be carefully addressed to make the GSA step as informative and relevant as possible. Sensitivity can first be examined at a local scale, on each spatial unit~~, which can make it computationally extensive~~. It may turn the exercise computationally extensive but allows for a robust identification of influential parameters, especially in decision-making problems such as demonstrated in Smith et al. (2021). A second point of view consists in providing the user with synthetic measures that summarize the sensitivity over the whole spatial domain. Saint-Geours (2012) introduces the notion of *site* sensitivity indices to refer to local GSA and the notion of *aggregated* ~~sensitivity indices (sometimes called~~ (or *block* ~~sensitivity indices)~~) sensitivity indices to make reference to sensitivity indices ~~with respect to all spatial points~~ synthetized on the whole domain. Site sensitivity indices result in sensitivity maps that detail spatial contri-

butions of influential parameters (Herman et al., 2013; Abily et al., 2016). Aggregated sensitivity indices are built either on a scalar objective function built from the spatialized output (like sum, average or maximum value) as explored in Saint-Geours et al. (2014) or from a GSA performed on spatialized multidimensional output. In that case the extension of the GSA method to multidimensional output should be examined so as the interpretation of the results. For instance, the generalization of Sobol' indices to multidimensional output from Gamboa et al. (2013) is used in De Lozzo and Marrel (2016), to perform GSA on a spatialized radioactive material release model.

In this paper, we explore how the sensitivity of risk-assessment models can be relevantly tackled based on the example of the process-oriented, modular PESHMELBA model (PESticide and Hydrology: ModELling at the catchment scale). We perform a GSA of the PESHMELBA model ~~and we aim at identifying relevant tools and a feasible methodology that could be transposed to other complex, distributed models. The~~ based on a screening and a ranking step. As many ranking methods with different definitions of sensitivity exist, we assume that combining some of them is most likely to provide a comprehensive and robust ranking. We thus combine several GSA methods and we also investigate the spatialized aspect of the sensitivity analysis ~~is particularly investigated~~. A broader scientific purpose is to determine how the information got from GSA can be used to better understand the processes that drive transfers and fate of pesticides in the PESHMELBA model. The analysis is performed on a virtual scenario based on a real catchment in the Beaujolais region (France).

The paper is organized as follows: we describe the PESHMELBA model in Section 2.1 and the model setup in Section 2.2. ~~Then,~~ The input sampling is described in Section 2.3, then we introduce the different GSA ~~methods used and the case study in Section ??~~ approaches and the methodology used for landscape analysis ~~in Section 2.6. Input sampling is described in Section 2.3~~ from Section 2.4 to Section 2.6. Results are presented in Section 3 focusing successively on screening (Section 3.1), comparison of GSA methods (Section 3.2) and spatial analysis (Section 3.3).

2 Material and methods

2.1 The PESHMELBA model

The PESHMELBA model represents a catchment as a set of interconnected components that stand for landscape elements such as plots, Vegetative Filter Strips (VFSs), ditches, hedges or rivers (Rouzies et al., 2019). In order to respect the spatial organization and the heterogeneity of the landscape, it deals with mesh elements that can be surfaces or lines. Surface mesh elements are called Homogeneous Units (HUs). A HU is a portion of landscape that is homogeneous in terms of ~~hydrodynamical~~ hydrodynamic processes and agricultural practices. Linear mesh elements are called reaches. A reach is characterized by its nature (so far ditch, river or hedge) and by its neighbouring components: it is at most in contact with one elementary mesh element on each bank. In addition to its geometric or hydrodynamic properties, each mesh element is characterized by its one-way connections with the neighbouring components that stand at a lower altitude. One or several processes are represented on each element depending on its nature. Lateral transfers at surface and in subsurface between elements are also integrated. Independent codes called units are used to simulate the different processes, depending on the knowledge the user has on the

targeted catchment functioning. Then, the OpenPALM coupler (Fouilloux and Piacentini, 1999; Buis et al., 2006) is used to couple the units and to build the complete application. OpenPALM has adapted features to easily deal with spatial and temporal aspects of the model. For example, synchronization tools make it possible to couple processes with different time steps. The final structure is highly modular and process representations can easily be added, upgraded or removed depending on the landscape description. These features make PESHMELBA particularly suitable for scenario exploration.

PESHMELBA focuses on surface and subsurface transfers of water and pesticides. An extensive description of elements and processes already included can be found in Rouzies et al. (2019). The PESHMELBA version used in this study integrates a representation of water and pesticide transfers on plots, VFSs and rivers. Each plot or VFS is represented by a unique column of soil divided into vertical cells. In such a column, vertical infiltration is simulated using a solution of the 1D Richards equation proposed by Ross (Ross, 2003). An adapted set of parameters makes it possible to represent high infiltration rate, surface runoff reduction and enhanced adsorption and degradation on VFSs. Root-water uptake is integrated based on Varado et al. (2006). Surface runoff routing is represented based on the kinematic wave (Lighthill and Whitham, 1955) and the Darcy law (Darcy, 1857) is used for lateral subsurface transfers. In addition to shallow groundwater tables, PESHMELBA also represents shallowly perched water tables and associated lateral transfers. Finally, reactive transfer of solutes is represented: advection, degradation based on a first order law and adsorption, based on linear or Freundlich isotherms are integrated. Each river or ditch reach is represented by a unique tank. The River1D module (Branger et al., 2010) solves the kinematic wave equation for water routing and pesticide advection in the network. Groundwater-river exchanges are represented by the Miles formulation adapted by Dehotin et al. (2008).

2.2 Model setup

A virtual scenario of limited size ~~was~~ is set from a portion of la Morcille catchment (France) in order to explore different GSA methods and to ease interpretation of spatialized results. The chosen portion ~~was~~ is selected so as to remain representative of the global composition of La Morcille catchment in terms of soil, slope, type and size of elements as well as interface length between them. The chosen scenario ~~was~~ is composed of 10 vineyard plots, 4 vegetative filter strips and ~~6~~ 5 river reaches that delimit a left and a right slope (see Figure 1).

Figure 1. Left: portion of La Morcille catchment selected to perform sensitivity analysis. Yellow units stand for vineyard plots while green units stand for vegetative filter strips. Brown stars denote locations of pesticide application. Right: slopes and connections between elements.

Soils types on the catchment are mainly sandy (Peyrard et al., 2016). We ~~used~~ use the classification from Frésard (2010) that groups soil types into 3 main Soil Units (SUs). Each SU is defined by the vertical succession of 3 or 4 soil layers, also called soil horizons: one surface horizon, 1 or 2 intermediary horizons and one deep horizon as depicted in Figure 2. Note that interface depths can vary from one SU to another. The reader may refer to Rouzies et al. (2019) for further details on how soil types and soil horizons are represented in PESHMELBA. At the catchment scale, the classification ~~resulted~~ results in the following SU (see Figure 2): sandy soil (SU1), sandy soil on clay on the right bank plateau (SU2) and heterogeneous sandy soils on lower slopes and thalwegs (SU3).

~~Soil type locations for the case study. Green contours show the vegetative filter strips.~~

~~Spatial arrangement was~~ Spatial arrangement is set in order to be as realistic as possible in terms of possible interfaces between SUs. Each SU ~~was~~ is set at least on one vineyard plot and one VFS on the virtual scenario. ~~Bulk density~~
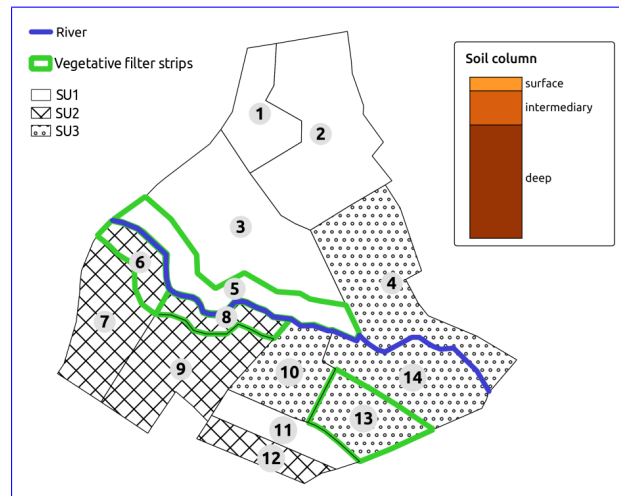
Figure 2. Soil type locations for the case study. Green contours show the vegetative filter strips.

Considering the different soil horizons whose hydrodynamic behavior and texture must be parametrized, the two types of vegetation (grassland and vineyard), and the different landscape element types (plots, VFSs and river reaches) that are simulated, the scenario results in 145 input parameters to be considered for sensitivity analysis. They are described in Table 1 so as the spatial level on which they are set and their values for the nominal simulation.

For the nominal scenario, values for bulk density bd and organic carbon content ~~data were~~ moc are available from Van den Bogaert (2011) and Randriambololohasinirina (2012). ~~Hydrodynamic~~ They are, as well as hydrodynamic parameters for each soil type~~are~~, described in Table 2. Retention values measured by Van den Bogaert (2011) ~~were~~ are used to fit retention curve using SWRCfit tool (Seki, 2007). A Schaap-Van Genuchten conductivity curve ~~was~~ is used (Schaap and van Genuchten, 2006; Ross, 2006) whose matching point at saturation Ko and empirical pore-connectivity L ~~were~~ are derived from conductivity data and retention parameters from retention curve fitting. Surface organic carbon content ~~was~~ are set equal to that of the first soil horizon on plots and VFSs. For each SU, only the first soil horizon on VFSs differs from vineyard plots so as to highlight enhanced infiltration capacities. Surface horizon on VFSs ~~was~~ are characterized by a 2.8%-organic carbon content (Randriambololohasinirina, 2012) and a saturated hydraulic conductivity of 150 mm.h^{-1} (or 4.31·10^{-5} m.s^{-1}) following Catalogne et al. (2018). In the absence of data to characterize potential anisotropy of vertical and horizontal saturated conductivities Ks_v and Ks_h, isotropy ~~was~~ is considered, thus the ratio Ks_h/Ks_v ~~was~~ is set to 1 on the catchment.

Input factor [units]	Description	Spatial level definition	Nominal value
Soil parameters			
$thetas$ [m^{3}m^{-3}]	Saturated water content	Soil horizon	see Table 2
$thetar$ [m^{3}m^{-3}]	Residual water content	Soil horizon	see Table 2
Ks [ms^{-1}]	Saturated hydraulic conductivity	Soil horizon	
hg [m]	Air-entry pressure in Van Genuchten retention characteristic curve	Soil horizon	see Table 2
mn	Deduced parameter from Van Genuchten retention characteristic curve n: $mn = n$-1	Soil horizon	see Table 2
Ko [ms^{-1}]	Matching point at saturation in modified Mualem Van Genuchten conductivity curve (Schaap and van Genuchten, 2006)	Soil horizon	see Table 2
L	Empirical pore-connectivity parameter	Soil horizon	see Table 2
bd [gcm^{-3}]	Bulk density	Soil horizon	see Table 2
moc [gg^{-1}]	Organic carbon content	Soil horizon	see Table 2
Pesticide parameters			
Koc [mLg^{-1}]	Freundlich sorption coefficient	Pesticide type	769
$DT50$ [d]	Half life	Pesticide type	47.1
Vegetation parameters			
$manning$ [sm$^{-1/3}$]	Manning's roughness	Vineyard plot-VFS	0.033-0.2
Zr [m]	Rooting depth	Vineyard plot-VFS	2.62-0.9
$F10$	Fraction of the root length density in the top 10% of the root zone	Vineyard plot-VFS	0.370-0.335
$LAImin$	Min LAI value	Vineyard (plot)	0.01
$LAImax$	Max LAI value	Vineyard (plot)	2.5
$LAIharv$	LAI value at harvest time	Vineyard (plot)	0.01
LAI	Constant LAI value	Grassland (VFS)	5
River parameters			
$hpond$ [m]	Ponding height in the river bed	River reach	0.01
di [m]	Distance between the river bed and the limit of impervious saturated	River reach	1.5

	Horizon	depth [m]	$theta s$ [m^{3}m^{-3}]	$theta r$ [m^{3}m^{-3}]	hg [m]	n [-]	Ks [ms^{-1}]	Ko [ms^{-1}]	L [-]
SU1	11-14	0.05	0.34	0.04	$-9.69 \cdot 10^{-2}$	1.27	$3.93 \cdot 10^{-5}$-$4.31 \cdot 10^{-5}$	$2.86 \cdot 10^{-7}$	-8.43
	2	0.5	0.34	0.05	$-3.29 \cdot 10^{-2}$	1.20	$8.64 \cdot 10^{-5}$	$2.28 \cdot 10^{-7}$	-6.52
	3	0.65	0.32	0.08	$-2.09 \cdot 10^{-2}$	1.20	$5.39 \cdot 10^{-5}$	$7.47 \cdot 10^{-7}$	-4.24
	4	4	0.28	0.07	$-5.99 \cdot 10^{-2}$	1.23	$3.11 \cdot 10^{-5}$	$1.47 \cdot 10^{-6}$	-0.14
SU2	12-15	0.035	0.34	0.04	$-9.69 \cdot 10^{-2}$	1.27	$3.93 \cdot 10^{-5}$-$4.31 \cdot 10^{-5}$	$2.86 \cdot 10^{-7}$	-8.43
	6	0.4	0.35	0	$-6.60 \cdot 10^{-2}$	1.13	$2.16 \cdot 10^{-5}$	$3.19 \cdot 10^{-7}$	9.66
	7	0.55	0.32	0	$-7.18 \cdot 10^{-2}$	1.08	$9.60 \cdot 10^{-6}$	$1.67 \cdot 10^{-7}$	-10
	8	4	0.42	0	$-0.30 \cdot 10^{-2}$	1.08	$3.98 \cdot 10^{-6}$	$9.72 \cdot 10^{-8}$	10
SU3	13-16	0.06	0.34	0.04	$-9.69 \cdot 10^{-2}$	1.27	$3.93 \cdot 10^{-5}$-$4.31 \cdot 10^{-5}$	$2.86 \cdot 10^{-7}$	-8.43
	9	0.45	0.33	0.08	$-6.72 \cdot 10^{-2}$	1.26	$3.05 \cdot 10^{-5}$	$3.36 \cdot 10^{-7}$	0.42
	10	4	0.32	0.06	$-3.56 \cdot 10^{-2}$	1.18	$2.38 \cdot 10^{-5}$	$3 \cdot 10^{-7}$	1.05

Table 2. Hydrodynamics parameters for SU1, 2 and 3 based on Van Genuchten retention curve and on Schaap-Van Genuchten conductivity curve fitting. Parameters are described in Table 1. Values for the surface horizon are explicitly distinguished between plot and VFS when they are different. Horizons 11, 12, 13 are surface horizons for plots whereas horizons 14, 15, 16 are surface horizons for VFSs.

175 The pesticide chosen in this study is the tebuconazole as it is a fungicide widely used on la Morcille catchment. It is a slightly mobile molecule and we ~~used~~ use a Freundlich isotherm to describe its adsorption equilibrium. Adsorption parameters ~~were~~ are obtained from Lewis et al. (2016) (Koc = 769 mL.g^{-1}, Freundlich isotherm exponent = 0.84). Surface degradation coefficient ~~was~~ is also taken from Lewis et al. (2016) ($DT50$ = 47.1 days) and a decreasing degradation rate in function of depth ~~was~~ is set as in FOCUS (2001). A 500g-application ~~was~~ is considered at the beginning of the simulation on plots 2 and 7 (see Figure

180 1). Most of the transformation and adsorption of tebuconazole ~~was~~ is supposed to happen on plots and VFSs at this modelling scale. Therefore, no adsorption or degradation ~~was~~ is simulated in the river.

A no-flux boundary condition ~~was~~ is applied on all sides except on surface where rain and potential evapotranspiration ~~were~~ are considered. Rain data ~~were~~ are extracted from BDOH database (Gouy et al., 2015). A 3-month simulation ~~was~~ is performed on a ~~winter~~ time period characterized by long and intense rain events (670 mm cumulated) allowing for significant

185 water and pesticide transfers both by surface runoff and subsurface saturated transfers. Potential evapotranspiration (PET) data ~~were~~ are obtained from MeteoFrance for the neighbouring site of Liergues (MeteoFrance, 2008). Data ~~were~~ are averaged over 10-day periods and corrected by a factor -11 % to match La Morcille site ~~(Durand, 2014; Caisson, 2019)~~as recommended in Durand (2014) and Caisson (2019). Rain and PET data for the simulation are summarized in Figure 3.
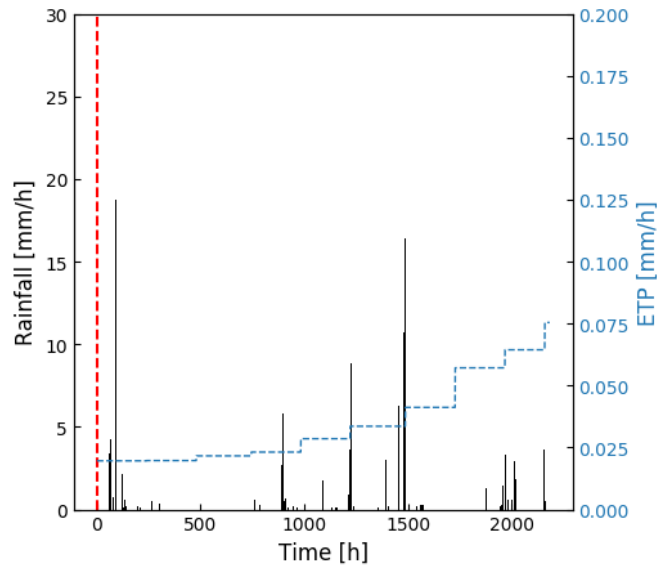
Figure 3. Climatic forcing (rain and potential evapotranspiration) for the simulation. The dotted red line stands for the one-shot pesticide application.

Although virtual, we ~~aimed~~ aim at setting initial conditions as plausible as possible for this scenario. ~~Initial~~ As running the model on a warm-up period is not possible due to data availabilitiy and computational cost limitation, initial water table levels ~~were~~ are deduced from piezometric data on a neighbouring hillslope and all soil columns are supposed to be in hydrostatic equilibrium at the beginning of the simulation. Data from several piezometers ~~were~~ are available on a transect, perpendicular to the river. Data ~~were~~ are extrapolated over the virtual hillslope width on both sides of the river. ~~All soil columns were supposed to be in hydrostatic equilibrium at the beginning of the simulation.~~ An upstream 0.177 m^{3}s^{-1}-flow ~~was~~ is considered in the river based on local measurements (Gouy et al., 2015).

Two types of vegetation ~~were~~ are represented in this scenario. Vineyard cover ~~was~~ is considered on plots while permanent grassland ~~was~~ is simulated on VFSs. Considering the period of simulation (3 months), a fixed root depth (Zr=2.62 m) and a fixed root density in the first 10 % of the root depth ($F10$=37 %) ~~were~~ are considered for vineyards following values reported in Smart et al. (2006) and confirmed by expert knowledge in the area. The root depth (Zr) ~~was~~ is set to 0.9 m and the root density in the first 10 % of the root depth ($F10$) ~~was~~ is set to 33.5 % for grassland (Brown et al., 2007). For vineyards, the Leaf Area Index (LAI) ~~was~~ is assumed to increase from a minimul value $LAImin$ from leaves formation until a maximum value $LAImax$ before declining until harvest $LAIharv$. Dates and associated values for this development cycle ~~were~~ are taken from Brown et al. (2007) ~~. They are summarized~~ and they are detailed in Appendix **??**~~so as~~. On grassland, the LAI is assumed to remain constant and a nominal value of 5 is chosen based on Brown et al. (2007). In Table 1 the reader should note that the LAI parameter relates to the constant LAI value ~~set for grassland on VFSs, based on Brown et al. (2007)~~for grassland while $LAImin, LAImax$ and $LAIharv$ relate to vineyard. The remaining parameters for root extraction model ~~were~~ are also fixed

to nominal values proposed by Varado et al. (2006); Li et al. (2001). Manning coefficients ~~were~~ are set from data reported in Arcement and Schneider (1989). A mature row crop value (0.033 s.m$^{-1/3}$) ~~was~~ is chosen for vineyard while a high grass pasture value (0.2 s.m$^{-1/3}$) ~~was~~ is set for VFS cover.

210   Finally, ponding height ~~was~~ is set to 0.01 m on vineyard plots while an increased value ~~was~~ is set on VFSs (0.05 m). According to ~~Gao et al. (2004); Walter et al. (2007)~~Gao et al. (2004) and Walter et al. (2007), the surface mixing layer thickness ~~was~~ is set to 0.01 m on both plot and VFS domains. In the river, the distance between the river bed and the limit of impervious saturated zone (di) ~~was~~ is set to 1.5 m (ERT field measure, personal communication) and the saturated hydraulic conductivity (Ks) ~~was~~ is set to 2.38·10^{-5} ms^{-1} accordingly to local saturated conductivities in the neighbouring soil. The ponding height is

215 set to 0.01m while the Manning coefficient is chosen equal to 0.079 s.m$^{-1/3}$ as suggested in Arcement and Schneider (1989) for channels with limited obstruction.

2.3   Input sampling

The input factor distributions are set to be as representative as possible of the available data on the catchment and the associated uncertainties. Mean values are taken from the nominal scenario described in Section 2.2. Distributions and standard deviations

220 are assigned based on experimental measurements from the catchment of application, available scientific literature or expert knowledge. All assigned distributions and corresponding statistics are summarized in Appendix **??**.
    As commonly found in the literature (Coutadeur et al., 2002; Fox et al., 2010; Schwen et al., 2011; Dairon, 2015; Dubus et al., 2003; Du , a lognormal distribution is assigned to the saturated hydraulic conductivity K_s. A 20 % coefficient of variation (CV) is used so as to remain representative of each soil horizon hydrodynamic behavior. Distributions for Schaap-Van Genuchten parameters

225 could not be found in the literature, thus empirical pore-connectivity parameter L is assigned a uniform distribution +/- 20% around the mean value (Zajac, 2010). As the matching point at saturation in modified Mualem Van Genuchten conductivity curve Ko has the same physical meaning than Ks, a log-normal distribution is also assigned to this parameter and a 20 % CV is set. Saturated water content $thetas$, residual water content $thetar$, Van Genuchten parameter mn and air-entry pressure hg are assigned normal distributions (Schwen et al., 2011; Alletto et al., 2015; Dairon, 2015; Gatel et al., 2019). A 10% CV is

230 set to $thetas$ (Gatel et al., 2019; Lauvernet and Muñoz-Carpena, 2018) and $thetar$ is assigned a 25% CV (Gatel et al., 2019) . A 10 % CV is set for mn and hg (Schwen et al., 2011; Alletto et al., 2015; Gatel et al., 2019). A uniform distribution is assigned to organic carbon content moc (Lauvernet and Muñoz-Carpena, 2018). A triangular distribution is assigned to the Freundlich sorption coefficient Koc (Lauvernet and Muñoz-Carpena, 2018) and a normal distribution is assigned to the half life $DT50$. 60% CV are assigned to Koc and $DT50$ distributions as such parameters are considered relatively uncertain

235 (Dubus et al., 2003). Triangular distributions are assigned to Manning's roughness $manning$ on plots and for the river bed (Lauvernet and Muñoz-Carpena, 2018; Gatel et al., 2019). A uniform distribution with a +/- 20 % range around the mean value is assigned to remaining input factors as little information could be found in the literature (Zajac, 2010).
    Using a fully distributed model such as PESHMELBA raises the issue of sampling strategy. Indeed, in this case study, even if the site is only composed of 14 surface units, the large number of soil horizons on the catchment, considering the hydrodynamic

240 distinction between plots and VFSs, also dramatically increases the number of parameters. Sampling all parameters on each

spatial unit leads to a huge number of simulations that could not be computationally afforded. Moreover, such independent sampling on a very large number of parameters may lead to misinterpretation of the sensitivity analysis results as the influence of physical processes could not be distinguished from spatial arrangement. For each sample, one set of soil parameters is therefore sampled for each soil horizon and those parameters are applied to all spatial units that contain this horizon which mantains the number of parameters to be considered in the GSA to 145.

2.4 ~~Global Sensitivity Analysis methods~~ Methodology for global sensitivity analysis in the PESHMELBA model

~~We explored several Global Sensitivity Analysis methods in order to identify the most suitable approach to deal with the specificities of the targeted application: heterogeneous and coupled structure of the model, highly non linear processes and spatialized aspect. In what follows, we denote $Y \in \mathbb{R}$ a given model output. Y is function of a multivariate input random vector $\underline{X} = (X_1, ..., X_M) \in \mathbb{R}^M$ such as $Y = \mathcal{M}(X_1, ..., X_M)$, with $\mathcal{M}$ the computational model that maps $\underline{X}$ onto the output space. The following three sections detail the mathematical theory under each method investigated to compute sensitivity indices for the PESHMELBA model: the Sobol' indices based on variance decomposition using generalized polynomial chaos expansion, the HSIC dependence measures and the feature importance measures~~ The full workflow is summarized in Figure 4. Considering the high number of input parameters, a screening step is first performed to decrease the dimension of the problem. Screening is performed with a statistical independence test based on the HSIC measure (see Section 2.4.2) on an initial 4,000-point Latin Hypercube Sample (LHS, McKay et al. 1979).

Second, a new 1,000-point LHS obtained from the reduced set of input parameters is computed to perform ranking. Sensitivity indices are computed from the 1,000-point sample based on 1/ variance decomposition (Section 2.4.1), 2/ HSIC dependence measure (Section 2.4.2) and 3/ feature importance measure obtained from Random Forest ~~metamodelling technique. For each method, extension to multidimensional output $\underline{Y} \in \mathbb{R}^d$ is also investigated. Readers in a hurry may directly jump to Section 2.6 that providesa qualitative summary and the technical details on the implementation for~~ (Section 2.4.3). These methods are all global, model-free and all suit to non-linear, non-monotonic models. However, they may not be equally costly to set and they define the notion of sensitivity in contrasted ways: in the variance decomposition method, it is assumed that a parameter is influential if it is responsible for a large portion of the output variance. From the HSIC measure point of view, a parameter is influential if any non linear transformation of the targetted output variable is highly correlated to any non linear transformation of that parameter. Finally, a parameter is identified as influential by the RF feature importance measure if making an error on that parameter implies a large decrease in accuracy of a pre-built RF surrogate model. However, as implementing several methods may not be possible in every case studies, we also compare these methods regarding the information it provides, accuracy and ease of implementation to provide future users with guidelines to choose the most adapted approach for their case study.

Although the PESHMELBA model is dynamic, model outputs considered in this paper are scalar quantities rather than temporal series. In order to investigate PESHMELBA abilities to properly represent transfers in a heterogeneous landscape, sensitivity analysis is performed on spatialized hydrological and quality variables including cumulated water volume and pesticide mass transferred in the subsurface (saturated lateral transfers) and on surface (surface runoff). However, these quantities are
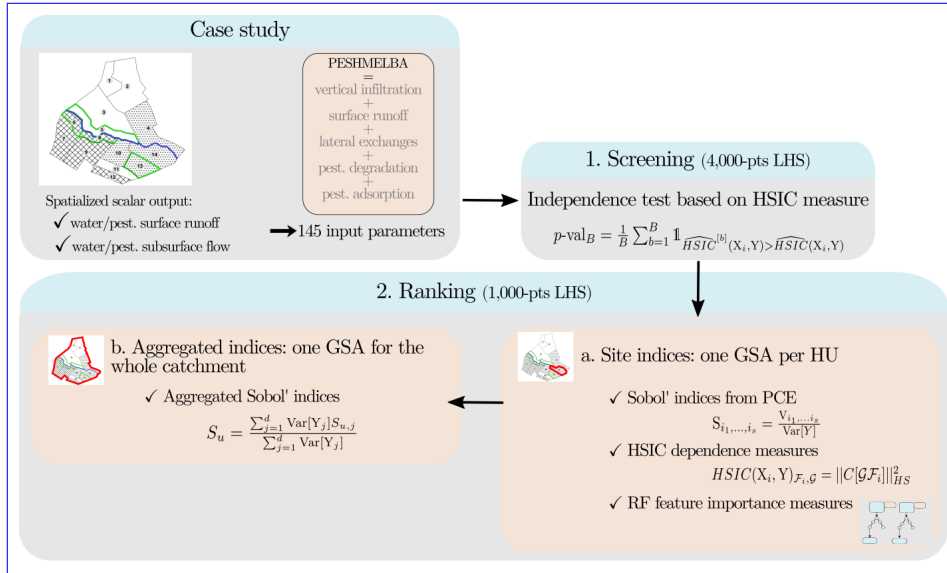
Figure 4. Full workflow used to perform GSA on the PESHMELBA model in 3 steps: 1/ Screening, 2.a/ Ranking at a local scale, 2.b/ Ranking at the catchment scale.

spatialized leading to multidimensional outputs. To deal with the spatialized aspect, GSA is first performed on scalar quantities, on each landscape element and sensitivity indices are aggregated in a second time providing catchment-scale sensitivity indices (see Section 2.6).

In what follows, we denote $Y \in \mathbb{R}$ a given scalar output from PESHMELBA. Y is function of a multivariate input random vector $Y = \mathcal{M}(\mathbf{X})$ where $\mathbf{X} = (X_1, ..., X_M) \in \mathbb{R}^M$ contains the 145 input parameters considered in this case study and where $\mathcal{M}$ is the PESHMELBA model.

2.4.1 Variance decomposition

~~Definition~~

Variance-based methods aim at determining how input factors contribute to the output variance (Faivre et al., 2013). One of the most popular variance-based method is the Sobol' method (Sobol, 1993). ~~It is based on the decomposition of the output variable Y into summands of increasing dimension:~~

$$Y = \mathcal{M}(x_1, ..., x_M) = m_0 + \sum_{s=1}^{M} \sum_{i_1 < ... < i_s}^{M} m_{i_1, ... i_s}(x_{i_1}, ..., x_{i_s}),$$

~~where $1 \leq i_1 < ... < i_s \leq M$ and where summands $m_{i_1, ... i_s}(x_{i_1}, ..., x_{i_s})$ are defined so that their integral over any of its independent variables is zero. This last property implies that the summands are orthogonal, which makes this decomposition unique~~ The Sobol' indices capture the direct impact of any input and also accurately describe the impact of input parameters in

interaction with others. It is ~~referred to as ANOVA for ANalysis Of VAriances (Archer et al., 1997). If Y = $\mathcal{M}(\mathbf{X})$ is square integrable, all summands $m_{i_1,...,i_s}$ are also square integrable and one can square and integrate Eq. (??) leading to:~~

$$\int \mathcal{M}^2(\mathbf{X})\mathrm{d}\mathbf{X} = m_0^2 + \sum_{s=1}^{M}\sum_{i_1<...<i_s}^{M}\int m_{i_1,...,i_s}^2 \,\mathrm{d}x_{i_1}...\mathrm{d}x_{i_s}.$$

~~$\int \mathcal{M}^2(\mathbf{X})\mathrm{d}\mathbf{X} - m_0^2$ is~~ based on the decomposition of the total variance of ~~$\mathcal{M}(\mathbf{X})$ denoted~~ Var[Y] ~~in what follows. If one introduces the notion of partial variances $V_{i_1,...,i_s}$ so that:~~

$$\mathrm{V}_{i_1,...i_s} = \int m_{i_1,...,i_s}^2(x_{i_1},...x_{i_s})\mathrm{d}x_{i_1},...,\mathrm{d}x_{i_s} \text{ for } 1 \le i_1 < ... < i_s \le M, s = 1,...,M,$$

~~Then, Eq. (??) reads:~~ the output:

$$\mathrm{Var}[Y] = \sum_{s=1}^{M}\sum_{i_1<...<i_s}^{M}\mathrm{V}_{i_1,...,i_s}, \tag{1}$$

~~In such formulation,~~ where $\mathrm{V}_{i_1,...i_s}$ indicates the portion of variance that can be attributed to interactions between input parameters $X_i, i \in i_1,...i_s$. ~~Single-index variance V_i indicates the portion of variance that can be attributed to the main effect of X_i taken alone.~~ From the above, one can define Sobol' indices as:

$$\mathrm{S}_{i_1,...,i_s} = \frac{\mathrm{V}_{i_1,...i_s}}{\mathrm{Var}[Y]}. \tag{2}$$

By definition, $0 \le \mathrm{S}_{i_1,...,i_s} \le 1$. In particular, first order sensitivity indices $\mathrm{S}_i = \frac{\mathrm{V}_i}{\mathrm{Var[Y]}}$ only account for main effects of parameter X_i. They can be interpreted as the decrease in the total output variance that could be obtained when removing the uncertainty about X_i when setting X_i to a fixed value (Tarantola et al., 2002). These indices are usually calculated as a first step as they often account for a large portion of the variance (Faivre et al., 2013). Total sensitivity indices $\mathrm{S}_{\mathrm{T}_i}$ evaluate the total effect of an input factor X_i on the output by taking into account its main effect S_i and all interaction terms that involve it:

$$\mathrm{S}_{\mathrm{T}_i} = \sum_{\mathcal{I}_i}\mathrm{S}_{i_1,...,i_s}, \quad \mathcal{I}_i = \{(i_1,...,i_s) \mid \exists k, 1 \le k \le s, i_k = i\}. \tag{3}$$

The total sensitivity index S_{T_i} stands for the portion of total output variance that remains as long as X_i stays unknown (Tarantola et al., 2002).

~~Sobol' indices computation using generalized polynomial chaos expansion~~

~~In the original paper, Sobol (1993) uses Monte Carlo simulations to compute these sensitivity indices. This method is based on two Monte Carlo samples of size N. However, such estimation of sensitivity indices requires 2^N estimations of the model. Saltelli et al. (2008) recommend to choose $N = k(M+2)$ where M is the number of input factors and k a number between 500 and 1000. Using such a sample size may end up being unfeasible in case of computationally expensive models. The use of a metamodel is then an interesting alternative, in particular with the Polynomial Chaos Expansions (PCE) that have the~~

advantage of providing direct estimation of Sobol' indices direct computation requires a large sample size that can not be afforded in this case study. As a result, we compute Sobol' indices . PCE theory is briefly introduced in what follows, mainly based on Le Gratiet et al. (2017).

from a limited sample size, based on Polynomial Chaos Expansion metamodelling technique is based on homogeneous chaos theory (Wiener, 1938). It (PCE, Sudret 2008) in order to circumvect such difficulty. This approach consists in building a surrogate model which analytical polynomial expression is directly related to Sobol' indices. Building a PCE and deducing the associated Sobol' indices thus only requires a training sample of limited size and knowledge about each input parameter marginal distribution. More precisely, PCE provides a functional approximation of the computational model based on the projection of the model output on a suitable basis of stochastic polynomial functions in the random inputs (Ghanem and Spanos, 1991). For any square integrable scalar output random variable $Y = \mathcal{M}(\underline{X}), Y \in \mathbb{R}$ Y, its polynomial chaos expansion is expressed as follows:

$$Y = \sum_{\boldsymbol{\alpha} \in \mathbb{N}^M} \gamma_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\underline{X}), \tag{4}$$

where the $\Psi_{\boldsymbol{\alpha}}$'s are multivariate orthonormal polynomials that constitute the basis and $\gamma_{\boldsymbol{\alpha}}$'s the associated coordinates. Multivariate polynomials are built as tensor products of univariate polynomials. For example, in the case of a multivariate polynomial of degree $\boldsymbol{\alpha}$, one defines the multi-index $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_M)$ with $|\boldsymbol{\alpha}| = \sum_{i=1}^{M} \alpha_i$ and $\Psi_{\boldsymbol{\alpha}}$ is expressed as:

$$\Psi_{\boldsymbol{\alpha}}(\mathbf{X}) = \prod_{i=1}^{M} \Psi_{\alpha_i}(x_i),$$

where $\Psi_{\alpha_i}(x_i)$'s is the univariate polynomial of degree α_i from the orthonormal family associated to variable x_i. Univariate polynomials are chosen among classical families of polynomials, depending on the input factor distributions (Xiu and Karniadakis, 2002) built according to the marginal probability density functions of each input factor and $\gamma_{\boldsymbol{\alpha}}$ are the associated coordinates. Expansion from Eq. (4) is usually truncated to a finite sum for practical computation:

$$Y \approx \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \gamma_{\boldsymbol{\alpha}} \Psi_{\boldsymbol{\alpha}}(\mathbf{X}),$$

where $\mathcal{A} \in \mathbb{N}^M$ is the subset of selected multi-indices of multivariate polynomials. Different truncation schemes exist, the most obvious being the standard truncation scheme that only keeps multivariate polynomials of total degree less or equal than p: $\mathcal{A}^{M,p} = \{\boldsymbol{\alpha} \in \mathbb{N}^M : |\boldsymbol{\alpha}| \leq p\}$.

PCE truncated decomposition can be rearranged into summands of increasing order. Such formulation is based on sets $\mathcal{A}_{i_1,...,i_s}$ that contain all $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_M)$ tuples such that only indices $(i_1, ..., i_s)$ are nonzero:

$$\mathcal{A}_{i_1,...,i_s} = \left\{ \boldsymbol{\alpha} : \begin{array}{l} \alpha_k > 0, \forall k = 1, ..., M \mid k \in (i_1, ..., i_s) \\ \alpha_k = 0, \forall k = 1, ..., M \mid k \notin (i_1 ... i_s) \end{array} \right\}$$

As detailed in Sudret (2008), the truncated expansion in Eq. (??) can be written as:

$$Y = \gamma_0 + \sum_{i=1}^{M} \sum_{\boldsymbol{\alpha} \in \mathcal{A}_i} \gamma_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_i) + \sum_{1 \leq i_1 < i_2 \leq M} \sum_{\boldsymbol{\alpha} \in \mathcal{A}_{i_1, i_2}} \gamma_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_{i_1}, x_{i_2}) + ... +$$

$$\sum_{1 \leq i_1 < ... < i_s \leq M} \sum_{\boldsymbol{\alpha} \in \mathcal{A}_{i_1, ..., i_s}} \gamma_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_{i_1}, ..., x_{i_s}) + ... +$$

$$\sum_{\boldsymbol{\alpha} \in \mathcal{A}_{1,2,...,M}} \gamma_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(x_1, ..., x_M).$$

By unicity of the ANOVA decomposition, Eq. (??) is therefore the ANOVA of Y. As a result, Sobol' indices can be deduced from gathering PCE coefficients according to the dependency on each basis polynomial. Because of the orthonormality of the multivariate polynomials, the partial variance are nothing but the sum of the squares of the PCE coefficients. After square-summing and normalizing, one gets:

$$S_{i_1, ..., i_s} = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_{i_1, ..., i_s}} \frac{\gamma_{\boldsymbol{\alpha}}^2}{\mathrm{Var}[Y]}$$

where $\mathrm{Var}[Y] = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \gamma_{\boldsymbol{\alpha}}^2$ is the total variance. In particular, first-order indices read:

$$S_i = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_i} \frac{\gamma_{\boldsymbol{\alpha}}^2}{\mathrm{Var}[Y]}, \quad \mathcal{A}_i = \{\boldsymbol{\alpha} \in \mathbb{N}^M \mid \alpha_i > 0, \alpha_j = 0, \forall j \neq i\}$$

, using for example a truncation scheme based on least angle regression (Blatman and Sudret, 2011). The Sobol' indices can then be obtained analytically from the coefficients $\gamma_{\boldsymbol{\alpha}}$ (see Sudret 2008 for a demonstration of the relation between PCE and total-order indices are expressed as:

$$S_i^T = \sum_{\boldsymbol{\alpha} \in \mathcal{A}_i^T} \frac{\gamma_{\boldsymbol{\alpha}}^2}{\mathrm{Var}[Y]}, \quad \mathcal{A}_i^T = \{\boldsymbol{\alpha} \in \mathbb{N}^M \mid \alpha_i > 0\}.$$

Extension to multidimensional outputs

In this paper, Sobol' indices for multidimensional outputs are calculated following the formulation by Gamboa et al. (2013) for generalized). In our study, the UQLab Matlab software (Marelli and Sudret, 2014) is used to computed Sobol' indices . Denoting the output variable $\underline{\mathbf{Y}} = \mathcal{M}(\underline{\mathbf{X}}) \in \mathbb{R}^d$ such as $\mathbb{E}[\|\underline{\mathbf{Y}}\|]^2 < \infty$, for any input parameter X_i, generalized first order Sobol' indices are based on the Hoeffding decomposition of the model function $\mathcal{M}$:

$$\mathcal{M}(\underline{\mathbf{X}}) = c + \mathcal{M}_i(\underline{\mathbf{X}}_i) + \mathcal{M}_{\sim i}(\underline{\mathbf{X}}_{\sim i}) + \mathcal{M}_{i, \sim i}(\underline{\mathbf{X}}_i, \underline{\mathbf{X}}_{\sim i}),$$

where $c = \mathbb{E}[\underline{\mathbf{Y}}], \mathcal{M}_i = \mathbb{E}[\underline{\mathbf{Y}}|\underline{\mathbf{X}}_i] - c, \mathcal{M}_{\sim i} = \mathbb{E}[\underline{\mathbf{Y}}|\underline{\mathbf{X}}_{\sim i}] - c$ and $\mathcal{M}_{i, \sim i} = \underline{\mathbf{Y}} - \mathcal{M}_{\sim i} - \mathcal{M}_i - c$ where $\sim i$ is the subset $\{1 \leq k \leq M, k \neq i$ After computing the covariance and scalarizing both sides of Eq. (??) by applying the trace operator Tr, one gets:

$$\mathrm{Tr}(\Sigma) = \mathrm{Tr}(C_i) + \mathrm{Tr}(C_{\sim i}) + \mathrm{Tr}(C_{i, \sim i})$$

 where $\Sigma, C_i, C_{\sim i}$ from the 1,000-point LHS (step 2.a, Figure 4). We use a q-norm- and $C_{i,\sim i}$ respectively denote the covariance matrices of $\underline{\mathbf{Y}}, \mathcal{M}_i(\underline{\mathbf{X}}_i), \mathcal{M}_{\sim i}(\underline{\mathbf{X}}_{\sim i})$ and $\mathcal{M}_{i,\sim i}(\underline{\mathbf{X}}_i, \underline{\mathbf{X}}_{\sim i})$. As soon as $\underline{\mathbf{Y}}$ is not constant, one can divide both sides of Eq. (??) by $\mathrm{Tr}(\Sigma)$, leading to:

$$S_i = \frac{\mathrm{Tr}(C_i)}{\mathrm{Tr}(\Sigma)} = \frac{\sum_{j=1}^{d} \mathrm{Var}[\mathrm{Y}_j] S_{i,j}}{\sum_{j=1}^{d} \mathrm{Var}[\mathrm{Y}_j]},$$

where $\mathrm{Var}[\mathrm{Y}_j]$ stands for the variance of the scalar j^{th} component of $\underline{\mathbf{Y}}$ and $S_{i,j}$ is the first order Sobol' indice of Y_j associated to X_i. The Hoeffding decomposition can be generalized to any subset $u \in \{1,..,M\}$ and Sobol' indices at any order can be then computed the same way. degree-adaptive sparse PCE based on Least Angle Regression Scheme (LARS, Blatman and Sudret 2011) with q-norm $\in [0.1, 0.2, ..., 1.0]$ and a maximum degree of 3.

2.4.2 HSIC dependence measure

The following sections briefly describe the HSIC dependence measure and introduce a method for computing an estimator. Although the description below considers 1d-input factor and 1d-output, the HSIC theory can be easily extended to multidimensional variables (see De Lozzo and Marrel 2016 for example) Sensitivity measures based on Hilbert-Schmidt Independence Criterion (HSIC, Da Veiga 2015) belong to the category of dependence measures that quantify, from a probabilistic point of view, the dependence between each input and the output. The greater the dependency between the input factor and the output, the greater the associated sensitivity measure. The Hilbert-Schmidt Independence Criterion used for GSA is based on the cross-correlation between any non-linear transformations of some input factor X_i and the output Y (De Lozzo and Marrel, 2016). Such dependence measure simultaneously captures a very broad spectrum of forms of dependency between the variables (Meynaoui et al., 2018). These indices can be estimated from small samples (a few hundreds of points) and do not depend on the number of inputs, which is a huge advantage.

The HSIC theory relies on Reproducing Kernel Hilbert Space (RKHS) and kernel functions. The Hilbert space $\mathcal{H}$ of functions from $\mathcal{X}$ to $\mathbb{R}$ associated to the scalar product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is a RKHS if for all $x \in \mathcal{X}$, the application $h \in \mathcal{H} \mapsto h(x)$ is a continuous linear form. The Riesz representation theorem stands that to each value x in the space $\mathcal{X}$ corresponds a unique function $\phi_x \in \mathcal{H}$ so that $\forall h \in \mathcal{H}, h(x) = \langle h, \phi_x \rangle_{\mathcal{H}}$. Such RKHS thus provides us with a mapping from $\mathcal{X}$ to $\mathcal{H}$. This mapping and thus the RKHS $\mathcal{H}$ is defined by its unique kernel function k such that $\forall (x, x') \in \mathcal{X}^2, \langle \phi_x, \phi_{x'} \rangle_{\mathcal{H}} = k(x, x')$

Definition

Let $\mathcal{F}_i$ denote the RKHS composed of all continuous bounded functions of input X_i with values in $\mathbb{R}$ and $\mathcal{G}$ the RKHS composed of real-valued continuous bounded functions of output Y with values in $\mathbb{R}$. $\langle \cdot, \cdot \rangle_{\mathcal{F}_i}$ (resp. $\langle \cdot, \cdot \rangle_{\mathcal{G}}$) is the inner product on $\mathcal{F}_i$ (resp. $\mathcal{G}$) and k_{X_i} (resp. k_Y) is the corresponding kernel function .

The main idea behind the Hilbert-Schmidt Independence Criterion (HSIC) used for GSA is to calculate the cross-correlation between any non-linear transformations of some input factor X_i and the output Y (De Lozzo and Marrel, 2016). It means that such dependence measure simultaneously captures a very broad spectrum of forms of dependency between the variables

(Meynaoui et al., 2018). ~~Such non-linear transformations of X_i and Y are described by the elements of $\mathcal{F}_i$ and $\mathcal{G}~~$that defines such scalar product. The HSIC measure ~~then corresponds to the cross-covariance between any functions f and g in these RKHS, namely cov$(f(X_i), g(Y))$. To do so, one may define the cross-covariance operator $C[\mathcal{G}\mathcal{F}_i] : \mathcal{G} \to \mathcal{F}_i$ which is the unique operator from $\mathcal{G}$ to $\mathcal{F}_i$ such that:~~

$$\forall f_i \in \mathcal{F}_i, \forall g \in \mathcal{G}, \langle f, C[\mathcal{G}\mathcal{F}_i](g)\rangle_{\mathcal{F}_i} = \text{cov}(f_i(X_i), g(Y)). \tag{400}$$

~~Finally, the HSIC measure~~ corresponds to the square of the Hilbert-Schmidt norm of the ~~cross-correlation operator~~ cross-covariance operator $C[\mathcal{G}\mathcal{F}_i] : \mathcal{G} \to \mathcal{F}_i$, which is:

$$HSIC(X_i, Y)_{\mathcal{F}_i, \mathcal{G}} = ||C[\mathcal{G}\mathcal{F}_i]||_{HS}^2 = \sum_{j,k} \langle u_j^i, C[\mathcal{G}\mathcal{F}_i](v_k)\rangle_{\mathcal{F}_i} = \sum_{j,k} \text{cov}(u_j^i(X_i), v_k(Y)), \tag{5}$$

where $(u_j^i)_{j \geq 0}$ and $(v_k)_{k \geq 0}$ are orthogonal bases of $\mathcal{F}_i$ and $\mathcal{G}$ respectively.

The resulting sensitivity indexes proposed by Da Veiga (2015) are defined for each input factor $X_i, i \in \{1, ..., M\}$ as:

$$S_{X_i}^2 = \frac{HSIC(X_i, Y)_{\mathcal{F}_i, \mathcal{G}}}{\sqrt{HSIC(X_i, X_i)_{\mathcal{F}_i, \mathcal{F}_i} HSIC(Y, Y)_{\mathcal{G}, \mathcal{G}}}}. \tag{6}$$

~~The HSIC measure fully depends on the choice of the RKHS $\mathcal{F}_i$ and $\mathcal{G}$ and especially on the choice of the scalar product that defines the relation between elements from these RKHS. The kernel function defines such scalar product (Meynaoui, 2019). A large range of kernel functions exists and in this paper we focus on universal kernels (i.e. a~~ Based on Gretton et al. (2005a), an estimator of HSIC can be computed from an N-sample $(x_i^j, y^j), j \in \{1, ..., N\}$ of (X_i, Y):

$$\widehat{HSIC}(X_i, Y)_{\mathcal{F}_i, \mathcal{G}} = \frac{1}{(N-1)^2} Tr(KHLH), \tag{7}$$

where $H \in \mathbb{R}^{N \times N}$ is the centering matrix $H_{ij} = \delta_{ij} - \frac{1}{N}$ and $K \in \mathbb{R}^{N \times N}$ and $L \in \mathbb{R}^{N \times N}$ are the Gram matrices defined as $K_{ij} = k_{X_i}(x_i, x_j)$ and $L_{ij} = k_Y(y^i, y^j)$ where k_{X_i} and k_Y are the ~~kernel that is dense in the space of continuous functions with respect to the infinity norm). Indeed, it is necessary to use~~ functions associated to each RKHS. In this study, and following De Lozzo and Marrel (2014, 2016) and Da Veiga (2015), we choose a Gaussian kernel as it is a universal kernel ~~to~~ that can fully characterize the independence of ~~the variables . If the RKHS $\mathcal{F}_i$ and $\mathcal{G}$ are universal, HSIC(X_i, Y) is equal to 0 if and only if X_i~~ variables ~~and Y are independent (Gretton et al., 2005b; Da Veiga, 2015). Following De Lozzo and Marrel (2014, 2016) and Da Veiga (2015) we chose a Gaussian kernel as they often perform well (Gretton et al., 2005a) and~~ that can be used for scalar or vectorial variables. For a vectorial variable $\mathbf{x} \in \mathbb{R}^q$, it is expressed as follows:

$$k(\mathbf{x}, \mathbf{x}') = \exp(-\lambda ||\mathbf{x} - \mathbf{x}'||_2^2), \tag{8}$$

with $||.||_2$ is the Euclidian norm in $\mathbb{R}^q$ and where the hyperparameter λ is called the bandwidth parameter of the kernel. In this study, the bandwith λ is estimated from the inverse of the empirical standard deviation of the sample.

$$HSIC(\mathrm{X}_i,\mathrm{Y})_{\mathcal{F}_i,\mathcal{G}} = \mathbb{E}[k_{\mathrm{X}_i}(\mathrm{X}_i,\mathrm{X}'_i)k_{\mathrm{Y}}(\mathrm{Y},\mathrm{Y}')]$$
$$+ \mathbb{E}[k_{\mathrm{X}_i}(\mathrm{X}_i,\mathrm{X}'_i)]\mathbb{E}[k_{\mathrm{Y}}(\mathrm{Y},\mathrm{Y}')]$$
$$- 2\mathbb{E}[\mathbb{E}[k_{\mathrm{X}_i}(\mathrm{X}_i,\mathrm{X}'_i)|\mathrm{X}_i]\mathbb{E}[k_{\mathrm{Y}}(\mathrm{Y},\mathrm{Y}')|\mathrm{Y}]],$$

Ranking and screening and based on HSIC measure

~~The HSIC dependence measure can be computed for each input factor X_i and directly used as a sensitivity index for ranking. In De Lozzo and Marrel (2014) a statistical test approach is also proposed for screening. It is based on bootstraps and it is then adapted for non asymptotic contexts~~ When using a universal kernel, the HSIC indices can also be statistically used for screening purpose (De Lozzo and Marrel, 2014). A statistical test can be set with the null hypothesis "X_i *and Y are independent*". Considering an experimental design of N points $(\cancel{X_i^1,...X_i^N}x_i^1,...,x_i^N)$ and the associated output points $(\cancel{Y_1,...,Y_N), \text{De Lozzo and Marrel (2014)} \text{aim at testing the hypothesis "}X_i \text{ and Y are independent". An }} y_1,...,y_N),$ an estimator $\widehat{HSIC}(\mathrm{X}_i,\mathrm{Y})$ of the dependence measure $HSIC(\mathrm{X}_i,\mathrm{Y})$ is firstly computed. Then B bootstrap versions $\underline{\mathbf{Y}}^{[1]},...,\underline{\mathbf{Y}}^{[B]}$ are resampled from the original output sample $(\cancel{Y_1,...,Y_N}y_1,...,y_N)$ with replacement so as to contain the same number of points N. For each $\underline{\mathbf{Y}}^{[B]}$ the input points associated to X_i are not resampled. Indeed, under the independence hypothesis, any values of Y can be associated to X_i. For each bootstrap version b, an estimator $\widehat{HSIC}^{[b]}(\mathrm{X}_i,\mathrm{Y})$ is computed. Then, the associated bootstrapped p-value is given by :

$$p\text{-val}_B = \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}_{\widehat{HSIC}^{[b]}(\mathrm{X}_i,\mathrm{Y})>\widehat{HSIC}(\mathrm{X}_i,\mathrm{Y})} \tag{9}$$

Finally, denoting α the significance level, if $p\text{-val}_B<\alpha$, the independence hypothesis is rejected, otherwise it is accepted. In this study, such statistical test is used to perform screening based on 100 bootstrap replicates and a 1% significance level. The

450 R code provided in De Lozzo and Marrel (2016) (see supplementary material) to compute HSIC measure has been adapted in Python to perform both screening and ranking.

2.4.3 Random Forest

This section briefly introduces the random forest metamodelling technique and details how it can be used to derive sensitivity measures. Again, it is assumed that the output Y is scalar but the concepts and algorithms below can be extended to multidimensional

455 variables.

General formulation

Random forest (Breiman, 2001) belongs Random Forests (Breiman, 2001) belong to ensemble machine learning techniques. It consists in creating a surrogate by averaging results from an ensemble of K decision trees created independently. A decision tree is composed of an ensemble of discriminatory binary conditions contained in nodes. Such conditions are hierarchically

460 applied from a root node to a terminal node (tree leaf) (Rodriguez-Galiano et al., 2014). The input space is therefore successively partitioned into smaller groups that correspond to the nodes according to a response variable. Such splitting goes on until reaching a minimum threshold of members per node. In this study, we consider regression trees that focus on continuous response variables.

  As each individual decision tree is very sensitive to the input dataset, bagging is used to avoid correlations between them

465 and to ensure model stability. It consists in training each decision tree from a different training dataset smaller than the original one. Such subsets are built from the original one by resampling with replacement making some members be used more than once while others may not be used. Such a technique makes the random forests more robust when facing slight variations in the input space and increases accuracy of the prediction (Breiman, 1996, 2001). The samples that are not used to grow a tree are called "out-of-bag" (OOB) data and will be used for the test step. The RF workflow is summarized in Figure 5.
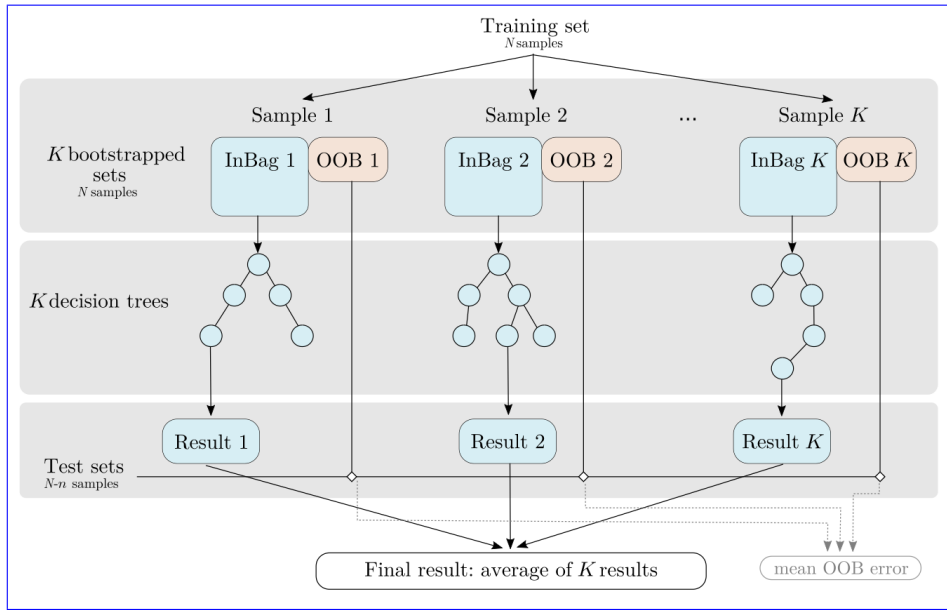
Figure 5. RF workflow (adapted from Rodriguez-Galiano et al. 2014). K bootstrapped sets are extracted from the original training set. Part of each set ~~"~~"InBag~~"~~" is used to grow an ~~independent~~ independant decision tree and the final regression value is the average of all trees. The remaining portion of each tree "out-of-bag" (OOB) is used as a test set.

470 ~~**Feature importance for the sensitivity analysis**~~

RF structure can be ~~cleverly~~ used to provide knowledge about how influential each input factor is. This measure is referred to as *feature importance* in the RF formalism. The random forest is first trained on the targeted output variable Y using a N-points sample $(\underline{\mathbf{X}}^j, \mathrm{Y}^j)$ for $j \in \{1, ..., N\}$. Once the forest has been trained, each input factor X_i is permuted individually so as to break the link between X_i and Y. The effect of such permutation on the model accuracy is then investigated. A large

475 decrease in accuracy indicates that the input factor is highly influential whereas a small decrease in accuracy indicates that it has little influence. Different algorithms exist to compute such Mean Decrease in Accuracy (MDA) (see Bénard et al. (2021) for an extensive review of the different formulations in the existing R and Python packages) and we focus here on the original formulation from Breiman paper (Breiman, 2001). The decrease in accuracy is originally computed from the mean square error between predictions from OOB data with and without permutation for each tree. Results are then averaged over all trees to

480 get the MDA. The algorithm for feature importance calculation is extensively described in various papers (Soleimani, 2021; Bénard et al., 2021, e.g.) and it is reminded in what follows:

1. For each tree k:

    – Estimate $\hat{\epsilon}_{OOB_k}$ the error from the OOB sample $\mathcal{L}_k$:

$$\hat{\epsilon}_{OOB_k} = \frac{1}{|\mathcal{L}_k|} \sum_{j \setminus (\underline{\mathbf{x}}^j, \mathrm{y}^j) \in \mathcal{L}_k} (\underline{\mathrm{y}}^j - \hat{\mathcal{M}}_{RF}(\underline{\mathbf{X}\mathbf{x}}^j))^2. \tag{10}$$

Where $\hat{\mathcal{M}}_{RF}$ is the estimated RF metamodel.

- For each input factor X_i:

  - Randomly permute ~~X_i in $\{\underline{\mathbf{X}}^j \in \mathcal{L}_k\}$~~ x_i in $\{\mathbf{x}^j \in \mathcal{L}_k\}$ to generate a new input set ~~$\{\underline{\mathbf{X}}^{j*} \in \mathcal{L}_k\}$~~ $\{\mathbf{x}^{j*} \in \mathcal{L}_k\}$.

  - Estimate $\hat{\epsilon}^*_{OOB_k}(i)$ using the permuted input set:

$$\hat{\epsilon}^*_{OOB_k}(i) = \frac{1}{|\mathcal{L}_k|} \sum_{\substack{~~j \setminus (\underline{\mathbf{X}}^j, Y^j) \in \mathcal{L}_k~~ \; j \setminus (\mathbf{x}^j, y^j) \in \mathcal{L}_k}} (\underline{Y} y^j - \hat{\mathcal{M}}_{RF}(\underline{\mathbf{X}} \mathbf{x}^{j*}))^2. \tag{11}$$

2. For each input factor X_i

   - Compute the mean decrease in accuracy MDA_i:

$$MDA_i = \frac{1}{K} \sum_{k=1}^{K} \hat{\epsilon}_{OOB_k} - \hat{\epsilon}^*_{OOB_k}(i). \tag{12}$$

   Where K is the total number of trees

Despite the "black-box" aspect of RF building, recent works theoretically ~~established~~ establish a link between Mean Decrease in Accuracy and Sobol Total Indices when input parameters are assumed independent. Indeed, Gregorutti et al. (2017) ~~established~~ establish that for all input parameter X_i:

$$ST_i = \frac{MDA_i}{2\text{Var}[Y]}. \tag{13}$$

2.5 ~~**GSA strategy for the landscape analysis**~~

~~In this study, several GSA approaches that define the notion of sensitivity in contrasted ways were tested and compared. These methods are described from a methodological point of view in the previous sections and qualitatively summarized in what follows. A variance decomposition method was first used and the associated Sobol' indices (Sobol, 1993) were computed. These indices measure the impact of each input parameter on the variance of the output. They capture the direct impact of any input and also accurately describe the impact of input parameters when they are interacting ones with others. Sobol' indices direct computation requires a large sample size that could not be afforded in this case study. As a result, we computed Sobol' indices from a limited sample size, based on Polynomial Chaos Expansion (PCE, Sobol 1993; Sudret 2008) in order to circumvect such difficulty. This approach consists in building a surrogate model which analytical polynomial expression is directly related to Sobol' indices. Building a PCE and deducing the associated Sobol' indices thus only requires a training sample of limited size and knowledge about each input parameter probability distribution. In this study, we used the UQLab Matlab software (Marelli and Sudret, 2014) and we computed Sobol' indices from a 1,000-point Latin Hypercube Sample (LHS, McKay et al. 1979). We used a q-norm- and degree-adaptive sparse PCE based on Least Angle Regression Scheme (LARS, Blatman and Sudret 2011) with q-norm $\in [0.1, 0.2, ..., 1.0]$ and a maximum degree of 3. Second, we computed sensitivity measures based on Hilbert-Schmidt Independence Criterion (HSIC, Da Veiga 2015). HSIC sensitivity indices belong to the category of dependence measures that quantify, from a probabilistic point of view, the dependence between each input~~

~~and the output. More precisely, the HSIC indice calculates the distance between the input/output joint pdf and the product~~
~~of their marginal pdfs when they are embedded in a Reproducing Kernel Hilbert Space. HSIC-based measures were used~~
~~both for ranking and for screening purposes as they allow for fully characterizing the independence between two variables. In~~
~~addition, these indices can be estimated from small samples (a few hundred of points) and do not depend on the number~~
~~of inputs.~~ In this study, ~~screening was performed on a 4,000-point LHS to ensure a reasonable simulation time using a~~
~~statistical test for independence. The power of the test α was set to 1% and 100 bootstrap replicates were used. Then,~~
~~ranking was performed on the new 1,000-point LHS obtained from the reduced set of input parameters. The R code provided~~
~~in De Lozzo and Marrel (2016) to compute HSIC measure was adapted in Python to perform both screening and ranking.~~
~~Finally, sensitivity indices deduced from Random Forest (RF, Breiman 2001) surrogate model were also computed. A Random~~
~~Forest model is an empirical input/output relationship based on an ensemble of decision trees. The relative importance of~~
~~each input parameter in the RF building (called *feature importance measure*) can be easily deduced from it. More precisely,~~
~~an input parameter X_i is considered important if when breaking the link between X_i and the output Y by permutation, the~~
~~RF prediction error increases. The metric for measuring prediction error can be defined in different ways and we chose the~~
~~original formulation of Mean Decrease Accuracy from Breiman (2001). These feature importance measures constitute a kind~~
~~of sensitivity indices that can be calculated (almost) for free as soon as the RF structure is available. In this study, we used~~ the
randomForestSRC R package (Ishwaran and Kogalur, 2020) is used to obtain feature importance measures once again from
the 1,000-point LHS. The number of trees used to train the RF ~~was~~ is set to 500.

~~Full workflow used to perform GSA on the PESHMELBA model.~~

2.5 Robustness assessment of the sensitivity indices

~~In order to investigate PESHMELBA abilities to properly represent transfers in a heterogeneous landscape, sensitivity analysis~~
~~was performed on spatialized hydrological and quality variables including cumulated water volume and pesticide mass transferred~~
~~in the subsurface (saturated lateral transfers) and on surface (surface runoff). The full workflow is summarized in Figure ??.~~
~~GSA was first performed at a site scale using each landscape element individually and considering Sobol' indices, HSIC~~
~~dependence measures and RF feature importance measures.~~ 95% confidence intervals ~~were also~~ are calculated for all methods
and an additional 200-point test ~~set allowed for assessing~~ is used to assess PCE and RF metamodel performances. The boot-
strap resampling procedure provided in UQLab ~~was~~ is used to calculate confidence intervals on PCE-based Sobol indices (see
Marelli and Sudret 2018 for justification). As RF building already includes a bootstrap step to build each tree, applying another
bootstrap could thus affect the consistency of the OOB sample. We then ~~used~~ use a subsampling approach without replacement
with subsample size set to 800 to get error bounds on feature importance measures. As deducing the RF structure from a re-
duced dataset may lead to limited performances, we ~~primarily~~ first checked that RF performances were reasonably decreased
when shifting the training dataset size from 1,000 to 800. The same procedure ~~was~~ is applied to estimate error bounds on HSIC
indices as 800 points ~~was~~ is supposed to be far enough to compute consistent estimators. ~~100 replications were used~~ As the
HSIC measure calculation is more computationally demanding than PCE and RF, we limit the number of replications to 100 for
all resampling methods. ~~In a second time, aggregated sensitivity indices were computed to summarize all *site*/local sensitivity~~

indices. In that case, the objective function was no longer a scalar but rather a vector that gathered a given cumulated variable for all HUs over the catchment. Aggregated

2.6 Aggregated sensitivity indices

Catchment-scale sensitivity indices are computed considering a multidimensional output $\mathbf{Y} \in \mathbb{R}^d$ that gathers scalar outputs on each landscape unit. Sobol' indices were calculated following Gamboa et al. (2013) on are aggregated at the catchment scale following the formulation by Gamboa et al. (2013) for generalized Sobol' indices (Section ??).

2.7 Input sampling

Definition of uncertainty in input factors is critical to make sure that GSA results are representative of the model performance (Sobol, 2001; Muñoz-Carpena et al., 2010; Nossent and Bauwens, 2012; Herman et al., 2013). Such an uncertainty is represented by attaching to each input factor a probability density function (pdf) that reflects its variability in the simulation context (Gatel et al., 2019). GSA performed in this paper was fully oriented toward the application case described in the previous sections. Therefore, the input factor distributions were chosen as being as representative as possible of the available data on the catchment and the associated uncertainties. Mean values were taken from the standard scenario described in Section 2.2. Distributions and standard deviations were assigned based on experimental measurements from the catchment of application, available scientific literature or expert knowledge (Table 1).

Input factor unitsDescription Pdf $thetas$ m³m⁻³Saturated water content N$thetar$ m³m⁻³Residual water content TNKs ms⁻¹Saturated hydraulic conductivity LNhg mAir-entry pressure in Van Genuchten retention characteristic curve N mn Deduced parameter from Van Genuchten retention characteristic curve n: $mn = n$-1 NKo ms⁻¹Matching point at saturation in modified Mualem Van Genuchten conductivity curve (Schaap and van Genuchten, 2006)LN L Empirical pore-connectivity parameter Ubd gcm⁻³Bulk density Umoc gg⁻¹Organic carbon content U $Kfoc$ mLg⁻¹Freundlich sorption coefficientT$DT50$ dHalf life N$manning$ sm⁻¹ᐟ³Manning's roughness TZr mRooting depth U$F10$ Fraction of the root length density in the top 10% of the root zone U$LAImax$ Max LAI value U$hpond$ mPonding height in the river bed U di mDistance between the river bed and the limit of impervious saturated zone U Ks ms

$$S_u = \frac{\sum_{j=1}^{d} \mathrm{Var}[Y_j]S_{u,j}}{\sum_{j=1}^{d} \mathrm{Var}[Y_j]},$$

(14)

where u is a subset of $\{1,...,M\}$, $\mathrm{Var}[Y_j]$ is the variance of the scalar j⁻¹ᵗʰ Saturated conductivity of the river bed U $manning$ sm⁻¹ᐟ³River bed Manning's roughness T $hpond$ mPonding height U $adsorpthick$ mMixing layer thickness U Input factor description and corresponding pdf. U:uniform, N:normal, TN:truncated normal, LN:log-normal, T:triangular. Distribution parameters (mean, standard deviation, bounds) are not indicated in this table as they differ from one soil type to another and from one vegetation type to another. Statistics of each assigned distribution are given in detail in Appendix ??.

As commonly found in the literature (Coutadeur et al., 2002; Fox et al., 2010; Schwen et al., 2011; Dairon, 2015; Dubus et al., 2003; Du , lognormal distribution was assigned to saturated hydraulic conductivity. A 20 % coefficient of variation (CV) was used so as to

remain representative of each soil horizon hydrodynamical behavior. Distributions for Schaap-Van Genuchten parameters could

580 not be found in the literature, thus L was assigned a uniform distribution +/- 20% around the mean value (Zajac, 2010). As Ko has the same physical meaning than Ks, a log-normal distribution was also assigned to this parameter and a 20 % CV was set. $thetas, thetar$ and Van Genuchten parameters hg were assigned normal distributions (Schwen et al., 2011; Alletto et al., 2015; Dairon, 20... A 10% CV was set to $thetas$ (Gatel et al., 2019; Lauvernet and Muñoz-Carpena, 2018) and $thetar$ was assigned a 25% CV (Gatel et al., 2019). A 10 % CV was set for mn and hg (Schwen et al., 2011; Alletto et al., 2015; Gatel et al., 2019). A

585 uniform distribution was assigned to moc (Lauvernet and Muñoz-Carpena, 2018). Bounds from data were used when several measurements were available. A mean distribution spread was calculated from horizons with several data and was used for setting max and min bounds on horizons when a single measurement was available. A triangular distribution was assigned to Koe (Lauvernet and Muñoz-Carpena, 2018) and a normal distribution was assigned to $DT50$. 60% CV were assigned to Koe component of $\mathbf{Y}$ and $DT50$ distributions as such parameters are considered relatively uncertain (Dubus et al., 2003). Triangular

590 distributions were assigned to Manning' s roughness on plots and for the river bed (Lauvernet and Muñoz-Carpena, 2018; Gatel et al., 2019). A uniform distribution with a +/- 20 % range around the mean value was assigned to remaining input factors as little information could be found in the literature (Zajac, 2010).

Using a fully distributed model such as PESHMELBA raised the issue of sampling strategy. Indeed, in this case study, even if the site was only composed of 14 surface units, the large number of soil horizons on the catchment, considering

595 the hydrodynamical distinction between plots and VFSs, also dramatically increased the number of parameters. Sampling all parameters on each spatial unit led to a huge number of simulations that could not be computationally afforded. Moreover, such independent sampling on a very large number of parameters may lead to misinterpretation of the sensitivity analysis results as the influence of physical processes could not be distinguished from spatial arrangement. For each sample, one set of hydrodynamical parameters was therefore sampled for each soil horizon and those parameters were applied to all spatial

600 units that contained this horizon leading to a total of 145 parameters to be sampled. $S_{u,j}$ is the Sobol' indice of subset u on Y_j. Eq. (14) formulates catchment-scale indices as an average of Sobol' indices on each landscape unit weighted by local output variances. First and total Sobol' indices can be notably computed this way. Their computation can be made in a second step after performing local Sobol' analysis on each landscape unit but a direct estimator is also proposed in Gamboa et al. (2013) avoiding numerous local analysis in case of high dimensional model output. For HSIC and RF sensitivity indices, the definitions

605 for scalar output remain valid for vectorial output. Catchment-scale indices based on both HSIC and RF can thus be directly computed on $\mathbf{Y}$ when moving to multidimensional case.

3 Results

3.1 Screening

Screening was is performed on the 4,000-point LHS. Simulations are run on the HIICS cluster (26 nodes, 692 cores, 64 to 256

610 GO of RAM per server) available at INRAE, France, for a simulation time of 4 days. Screening is performed at a site scale, on each HU individually, to remain as conservative as possible. Influential parameters at the catchment scale were are then de-

duced from the union of influential parameters for each site. After screening and union, 42 influential parameters are selected for water subsurface flow, 54 parameters are selected for pesticide subsurface flow, 43 parameters are selected for water surface runoff while 45 parameters are retained for pesticide surface runoff. The remaining parameters are given in Appendix **??**.

615 The number of input parameters retained after screening remains quite high proving that performing screening on PESH-MELBA variables is a challenging task. ~~It is commonly stated than only a few parameters explain most of the variability but such conclusion does not apply in this case as more than 40 parameters are identified as influential for each output variable considered (see Figure 6).~~ This can be partially explained by the methodology that may not be discriminating enough but it can also be a consequence of the many physical processes interacting in PESHMELBA in a spatially-distributed way, each of them

620 with its own set of characteristic parameters. ~~Additionally~~Furthermore, spatial heterogeneities in terms of number of influential parameters are noticed depending on the studied output state variable, as shown in Figure 6. More influential parameters are retained on the right bank (bottom part of the catchment), which is mainly composed of HUs from soil 2 and soil 3 (see Figure 2) for all output variables.~~Additionally, screening~~

Screening on pesticide variables results on a higher number of influential parameters on HUs situated on the right bank, close

625 to the outlet. Once again, such spatial heterogeneities in the influential parameters ~~identified~~ may be related to the physical processes at stake in the different parts of the catchment. To be noted that no strong correlations are noticed between the locations where pesticides were applied and the number of screened parameters (see Figure 1). ~~Such heterogeneities may rather be related to the climate forcing chosen for the simulation.Additional studies in contrasted climate contexts should be carried out to check whether or not these conclusions can~~ One should also remain aware that all conclusions about influential parameters

630 and heterogeneities are specific to the context of this case study (soils, climate,...) and cannot be generalized.
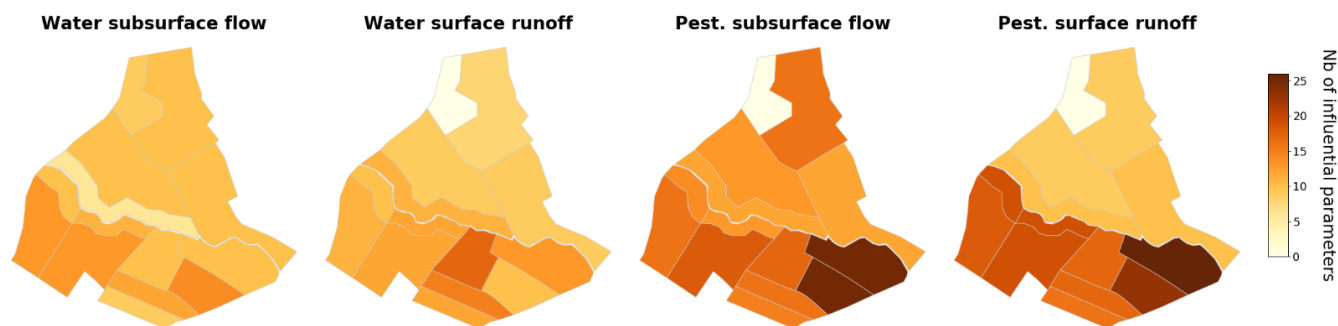
Figure 6. Number of influential parameters retained after screening for the different state variables on each HU. From left to right: cumulated water lateral transfers in subsurface, cumulated lateral pesticide transfers in subsurface, cumulated water surface runoff, cumulated pesticide surface runoff.

3.2 GSA on a single HU

The three methods (Sobol' indices from PCE, HSIC and RF) ~~were~~ are applied on each HU based on the 1,000-point ~~sample~~ LHS generated from the set of screened input parameters. Again, simulations are performed on the HIICS cluster, for a simulation

time of around 24 hours. First column of Figure 7 shows the top parameters ranked from Sobol' total-order indices for all
output variables on one HU close to the outlet, HU14. The identified influential parameters highly depend on the output variable considered. They are linked to distinct physical processes that may interact with each others and they bring knowledge about the way PESHMELBA represents the hydrological functioning of the virtual catchment. Water subsurface flow (top line) is driven by deep soil ~~hydrodynamical~~ hydrodynamic parameters both related to vertical infiltration and subsurface saturated transfers. Water surface runoff (line 2) is also mainly influenced by deep soil parameters. Overland flow is therefore identified as being mostly due to saturation rather than to rainfall excess. Subsurface exchanges with the river is also identified as an influential process as the river bed saturated conductivity (Ks_river) is part of the most influential parameters. Such finding is consistent with the position of HU14, which is directly connected to the river but also to many plots (see Figure 1).

Pesticide variables (line 3 and 4) are influenced by a higher diversity of parameters that characterize contrasted and interactive physical processes. Both pesticide subsurface flow and surface runoff are mostly influenced by the pesticide adsorption coefficient. Moreover, unlike the hydrological variables, parameters for surface and intermediary soil horizons rank among the top influential parameters. Such parameters may be linked to vertical infiltration but also to pesticide adsorption, as organic carbon content (moc) and saturated water content ($thetas$) are involved in the calculation of adsorption equilibrium. Additionally, the roughness coefficient ($manning\_plot$) and the ponding height ($hpond\_plot$) that are related to surface runoff calculation are also ranked as highly influential on pesticide surface runoff.

Sobol' first-order indices (Figure 7, column 2) reveal that first-order effects explain more than 95% of the water subsurface flow variance and that interactions (defined as $S_T - S_i$) contribute relatively little to the output variance for this variable. The conclusions for remaining variables are much more contrasted as interactions explain more than 40% of the output variance for water surface runoff, more than 70% of pesticide subsurface flow variance and reach more than 80% of explained variance in the case of pesticide surface runoff. Such strong interactions are not only linked to the numerous parameters used to simulate a given physical process in PESHMELBA but also reflect interactions between physical processes in the model.

Columns 3 and 4 in Figure 7 show HSIC and RF sensitivity indices for Sobol' top-ten parameters. On the whole, the rankings obtained from Sobol' total-order, HSIC and RF sensitivity indices are consistent giving confidence in the robustness of these methods. The most influential parameters identified from Sobol' indices are also captured by the other methods. Additional results (not shown here) also show that the top-ten rankings based on Sobol' total indices at least contain the five most influential parameters based on HSIC and RF rankings. Then, Figure 7 does not miss any preponderant parameters for HSIC and RF. Rankings match perfectly well for water subsurface flow while there are slight differences for water surface runoff. Differences between rankings are more pronounced for pesticide variables. In that case, most parameters have zero or very low first order Sobol' indices characterizing nearly purely interacting effects.

To summarize on method comparison, the three methods roughly identify the same top-ten of influential parameters for each variables but the rankings exhibit slight differences for pesticide variables. In De Lozzo and Marrel (2016), the authors call for caution when drawing general conclusion about HSIC and Sobol' indices comparison as they are based on different mathematical concepts. In their paper, HSIC indices were found to be more similar to Sobol' first-order than to total-order
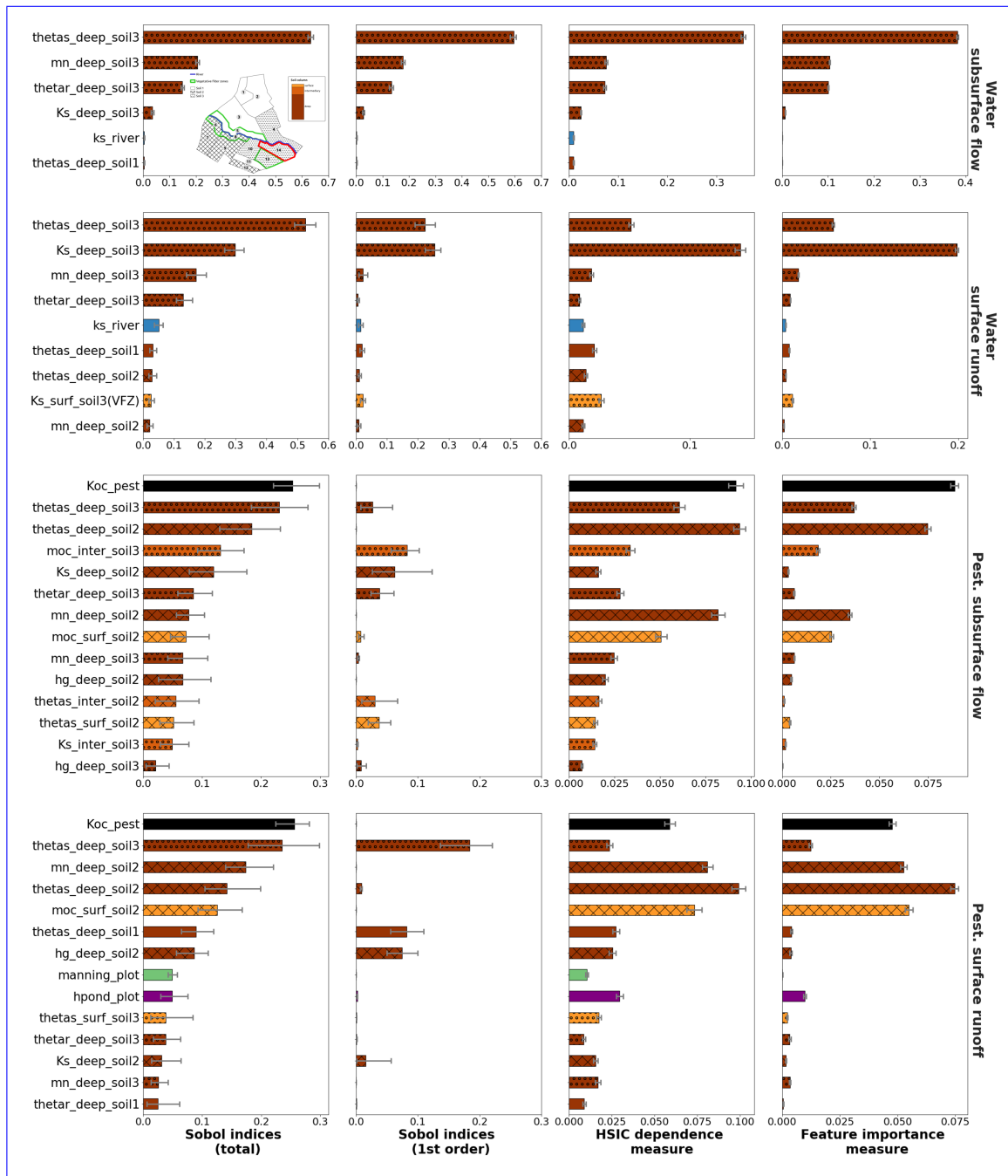
Figure 7. Sobol' total and first order indices computed using PCE, HSIC and RF site sensitivity indices for all output variables on HU14 with associated 95% confidence intervals. RF feature importance measures are normalized by $2\mathrm{Var}[Y]$ following Eq. (13). HU14 is displayed with a red contour on top left figure. For all methods, displayed parameters are the ~~10~~ most influential parameters regarding Sobol' total indices. The bar colours are related to physical processes: brown is related to soil parameters and the darker the brown, the deeper the parameter, blue is related to river parameters whereas green is related to vegetation parameters. Filling in brown bars refers to the soil type of the parameter: soil 1 is not filled, soil 2 is cross hatched whereas soil 3 is filled with circles.

indices but our findings reach the opposite conclusion. ~~These results confirm guidelines on the use of these indices stating that they should not be used for general comparison, being based on different mathematical concepts. While Sobol' indices scrutinize only the variance of the output, the HSIC indices consider the whole dependence between an input and the output. In addition,~~ HSIC indices are not intrinsically built to capture interacting effects. Our results show that they obviously capture some interactions but not all of them and not always ~~.~~ which may explain some observed differences. However, combining HSIC dependence measure and Sobol' total indices can also be of interest as it allows to identify parameters that are influential in other quantities than the output variance. For example, results show a high HSIC value and an intermediate Sobol' total indice value for mm_deep_soil2 in pesticide subsurface flow ranking or for $thetas\_deep\_soil2$ in pesticide surface runoff ranking. It may indicate that the influence of such parameters is not predominantly involved in the output variance. Similarly, the slight discrepancies between Sobol total indices and RF feature importance measures could be interpreted in terms of differences in sensitivity definitions. However, RF feature importance indices have been proven to relate to Sobol total indices (Gregorutti et al., 2017) but ~~our results exhibit large discrepancies between RF feature importance measures and Sobol total indices. Table from Appendix ?? provides~~ this relation (see Eq. 13) is not respected in this study. The quality scores for ~~both~~ PCE and RF metamodels ~~. Results~~ show that the RF metamodel performs much more poorly than the PCE for all variables probably due to the limited sample size (see Table in Appendix ??). Such poor performances may explain the discrepancies between RF feature importance measures and Sobol' total indices. Despite this limited performances, error bounds on RF indices are quite small contrarily to error bounds on Sobol' indices estimates. We incriminate the resampling strategy that differ between the two methods. While the bootstrap technique has been proven to assess the quality of the PCE (Marelli and Sudret, 2018), the subsampling technique set on RF only targets the precision of feature importance measures. A more adapted subsampling approach, for example based on **?** should probably be further investigated on bigger samples to better compute RF sensitivity indices and to accurately assess their quality.

To conclude, ~~Sobol' variance decomposition~~ although Sobol' indices cannot capture influential parameters if they do not predominantly affect the output variance, the limited differences in rankings show that they carry sufficient information for robust sensitivity analysis in this case study. If combining several methods was not possible, the variance decomposition approach may then be sufficient. In addition, it is the only method that gives insight into main and interactive effects, bringing precious knowledge for model validation. ~~Sobol' indices then seem more appropriate to analyse complexe pesticide-related variables, dominated by physical interactions.~~ These indices also have the merit of being easy to interpret ~~. In addition,~~ and a generalized formulation for aggregated indices is provided by Gamboa et al. (2013) and has already been deeply explored. However, the precision of the Sobol' indices obtained from PCE remains quite low and results should be thus interpreted with caution. Considering a high dimension problem and a very limited computational budget (inferior to $1,000$ simulations), HSIC indices may not be discarded to compute accurate sensitivity indices as soon as detailed knowledge about interactive effects is not needed. One should also note that a prior knowledge about input parameter pdf is not needed to compute HSIC indices, contrarily to Sobol' indices computed from PCE.

3.3 Landscape analysis

The previous section showed that Sobol' indices (first and total order) provide valuable information about interactions between parameters. As a result, we focus on this method ~~in what follows~~to compute the indices at the catchment scale. Site rankings such as presented in Figure 7 are gathered for all HUs in the form of sensitivity maps in Figure 8 for water surface runoff, and in Figure 9 for pesticide surface runoff. Broadly speaking, both maps show strong spatial heterogeneities regarding influential parameters and a contrasted behaviour between right and left banks can be identified. For both output variables, ~~hydrodynamical~~ hydrodynamic parameters ($thetas$, $thetar$ and mn) of deep horizon from soil 1 (resp. 2 and 3) are mainly influential only on HUs characterized by soil 1 (resp. 2 and 3) (see Figure 2 for a reminder on soil types).

Local ~~hydrodynamical~~ hydrodynamic parameters are found to be dominating to explain the output variable variance. A particular case is HU4 (indicated by an array on both Figure 8 and 9) which is characterized by soil 3 while parameters from soil 1 explained most of the variance of both water and pesticide surface runoff variables. The location of HU4, near the outlet, downstream several soil-1 HUs, may explain such spatial interactions. In addition to specific soil parameters, other parameters such as the manning roughness on vineyard plots ($manning\_plot$) or the coefficient of adsorption (Koc_pest) have a greater influence on HUs from the ~~left bank (top~~ right bank (bottom part of the catchment in the Figure).

Finally, comparison of first-order and total-order maps shows quite similar results for water surface runoff on the one hand. It indicates that direct effects are significant for all influential parameters. On the other hand, direct effects are far from dominant on pesticide surface runoff. Once again, most parameters are influential nearly only in interaction with other parameters since the fist order indices are very low compared to the total order indices.
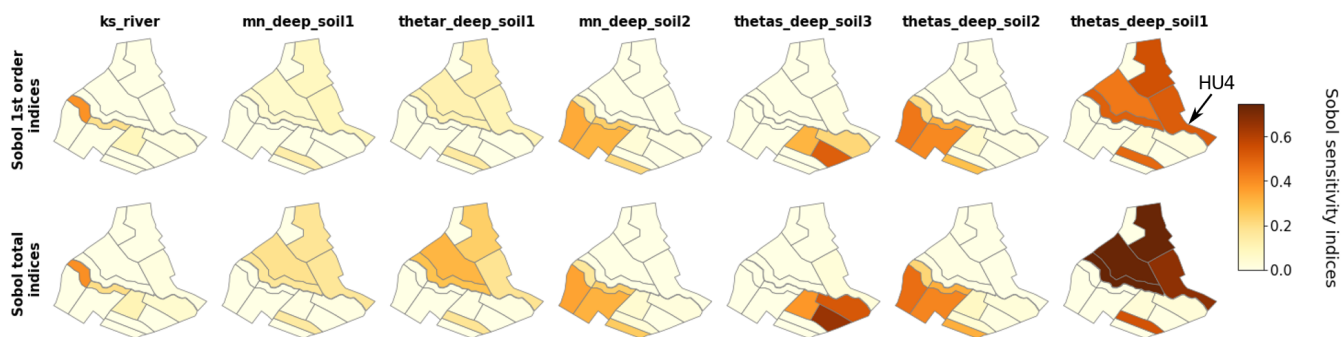
Figure 8. Maps of Sobol site sensitivity indices for water surface runoff for the most significantly influential parameters.
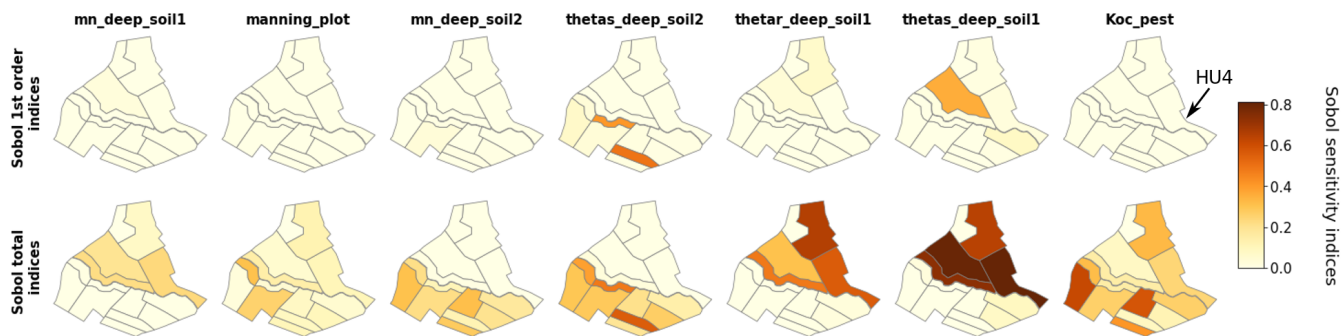
Figure 9. Maps of Sobol site sensitivity indices for pesticide surface runoff for the most significantly influential parameters.

Finally, Figure 10 shows aggregated sensitivity indices for water and pesticide surface runoff variables following Gamboa et al. (2013). Since the two banks have contrasted behaviors, aggregated indices are first calculated at the intermediary scale of the bank, then at the catchment scale. For both output variables, rankings strongly differ on each slope. As proposed at a local scale in Section 3.2, aggregated indices at this scale may constitute a summarized information about the physical pro-

725   cesses dominating in PESHMELBA to explain the output variable. For water surface runoff, ~~hydrodynamical~~ hydrodynamic soil parameters related to vertical infiltration ($thetas$, $thetar$ and mn) dominate. The influence of ks_river is only significant in the right bank. The difference of altitude between right and left bank may explain this contrast in the activation of saturated exchanges between water tables and the river. For pesticide surface runoff, deep horizon parameters from soil 1 and pesticide adsorption coefficient (Koc) explain a major portion of the output variance on the left bank while pesticide half-

730   life time ($DT50$) and surface runoff parameter ($hpond\_plot$) have lower or no impact. On the contrary, surface parameters ($manning\_plot$ and $hpond\_plot$) have a higher impact on the right slope. In that bank, soil horizons are characterized by lower permeabilities that may result in stronger surface runoff generation than on left bank. In addition pesticide parameters (Koc and $DT50$) are also more influential. More broadly, these results show that pesticide surface runoff may result from the activation and interactions of more physical processes on the right bank than on the left bank.
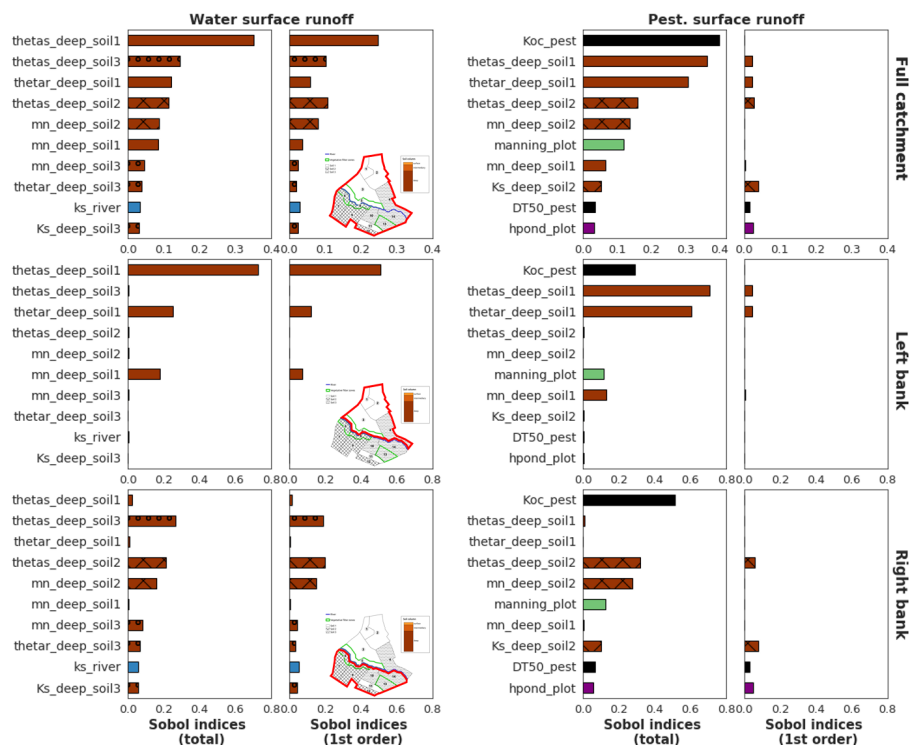
Figure 10. Sobol' first-order and total-order aggregated sensitivity indices for water surface runoff (left) and pesticide surface runoff (right) calculated at the scale of the catchment (top), left bank (middle) and right bank (bottom). Displayed parameters are the 11 most influential parameters regarding Sobol indices at the catchment scale for each output variable. The bar colours are related to physical processes: brown is related to soil parameters and the darker the brown, the deeper the parameter is, blue is related to river parameters and green is related to vegetation parameters. Filling in brown bars refers to the soil type of the parameter: soil 1 is not filled, soil 2 is cross hatched while soil 3 is filled with circles.

735   Sensitivity maps provide local, detailed information about influential parameters on each location of the catchment. However they are computationally costly as one GSA per HU must be performed. This approach may be hard or even impossible to transpose to real catchment scale composed of several hundreds elements. Catchment scale aggregated indices thus provide a synthetic information at a lower computational ~~price~~cost. In this study, Sobol' aggregated indices were directly computed from site sensitivity indices as they were available. However, if the size of the problem does not allow for direct computation,

740   a ~~pick-and-freeze~~ pick-freeze estimator (Gamboa et al., 2013; De Lozzo and Marrel, 2016) can be used for Sobol' generalized indices. In our case, such overview of sensitivity analysis allows us to focus calibration efforts on deep soil ~~hydrodynamical~~ hydrodynamic parameters and pesticide adsorption coefficient to improve the quality of the simulation of both water and pesticide surface flows. As pointed out in **?**, both approaches are complementary and provide precious knowledge about the model functioning. The scale of the bank, or more generally of the hillslope, may also constitute an adapted intermediary scale

745   to meet both requirements of detailed results for physical interpretation and computational efficiency.

4 Discussion on the methodology

The methodology we followed to perform GSA in the PESHMELBA model is a classical approach: first, a screening step and second, a ranking step applied on the reduced set of input parameters. Various previous GSAs on complex environmental models or pesticide models applied this screening/ranking strategy to cope with either a computationally costly model or a large number of input parameters. However, in this case study, we accumulate both difficulties and classical methods for each step could not thus be implemented. Indeed, most screening steps for environmental models are performed with the Morris method but a rigorous visual clustering was made difficult in such application because of the high dimension input space. Garcia et al. (2019) proposes a rigorous selection method for a fishery model with 133 input parameters but its methodology needs 134,000 model evaluations. Broadly speaking, most applications of Morris method in complex environmental models (Vanuytrecht et al., 2014; Garcia et al., 2019) advocate for a high number of trajectories resulting in an elevated number of model evaluations. A large sample size could not be afforded in our case and it goes against the benefits of the method that aims at remaining computationally cheap. That is why we performed screening based on a statistical test for independence instead. This test is based on the HSIC dependence measure that is recommended for small sample sizes (Da Veiga, 2015; De Lozzo and Marrel, 201 . To the best of our knowledge, this kind of screening method has never been applied in environmental modelling context before. In our case study, it allowed to significantly decrease the number of input parameters based on only 4,000 model evaluations. Model evaluations took 72h to run on the high performance computing cluster HIICS. The p-values were then calculated from 100 bootstraps leading to an additional calculation time of 2 hours on the HIICS cluster for one output variable, which makes the method computationally affordable. In addition, such statistical test is based on an arbitrary sample (LHS in this case but a classical Monte Carlo sample would have also been suitable) contrarily to Morris screening. It means that it may allow to reuse a pre-exisiting simulation set to limit additional computational efforts.

Once the input space dimension was reduced with screening, ranking was performed by combining variance decomposition, dependence measure and feature importance measure from random forest surrogate. Variance decomposition methods are classical for sensitivity analysis in pesticide transfer models. Various studies compute Sobol' indices (Hong and Purucker, 2018; Gatel et al. and eFAST method has also been increasingly used (Lauvernet and Muñoz-Carpena, 2018; Fox et al., 2010; Metta, 2007; Vanuytrecht et al . However, none of these studies combine the difficulties of a large input space dimension and computationally costly model runs. These studies then use a classical Sobol sample. As such could not be done in our case, we also used an alternative approach based on polynomial chaos expansion. PCE allowed to estimate Sobol' indices from a reduced sample size (1,000 points). In addition, innovative methods such as HSIC or RF never used before in pesticide transfer modelling have been tested and have proven their relevancy giving confidence in the results. Combining the different methods also provided additional insights into the sensitivity of each variables. All methods lead to robust sensitivity indices for nearly equivalent computational cost (<5 min-CPU for PCE and RF, <10 min-CPU for HSIC) and nearly equivalent ease of implementation. Again, all ranking methods use an arbitrary sample allowing reusing existing simulations (provided that they were performed with the previously screeened set of input parameters).

5 Conclusion

In this paper, we have described the first global sensitivity analysis of the modular and coupled PESHMELBA model. For this first experiment, a virtual, simplified catchment was set to explore different approaches for GSA and to propose a methodology for future real applications. Even if the scenario was simplified, a particular attention was paid to reproduce and to deal with the challenges that would be faced in real catchment applications: high number of input parameters with spatial heterogeneities, limited size of samples due to the high computational cost of a simulation and spatialized ouputs. We first ~~proposed~~ used a screening step using an independence test based on the HSIC dependence measure. Then, we ~~compared~~ combined several methods to compute sensitivity measures: Sobol' indices computed from a PCE surrogate model, HSIC dependence measures and feature importance measures got from RF surrogate model. ~~Although the results obtained from the three methods were mainly consistent, we noticed that Sobol' analysis was the only method that was able to capture some interactions between parameters, which may be of particular interest in such coupled models.~~

All methods were first applied on each landscape element individually. They provided us with local measures of sensitivity called site sensitivity indices. Such site sensitivity indices can be gathered into sensitivity maps and they highlight local contributions of parameters. Although very informative about the ~~hydrodynamical~~ hydrodynamic functioning of the scenario, these maps were computationally demanding to produce. Then, we used an extension of the previous methods to multidimensional outputs in order to get an overview of sensitivity over the whole physical domain. These aggregated indices were first computed at the catchment scale to characterize the whole output uncertainty. They may allow the users to focus calibration efforts. Additionally, they were computed on each *bank* hillslope and we propose to use this scale as an intermediary scale to get an aggregated information about the catchment functioning as it still reflects spatial heterogeneities of ~~hydrodynamical~~ hydrodynamic processes. When extending this methodology to other case studies, the intermediary spatial scales to be focused on should be defined depending on the characteristics and the goals of the study to make the most of the analysis.

Further research could extend global sensitivity analysis of PESHMELBA to contrasted climatic scenarios in order to encompass different environmental conditions as sensitivity analysis results highly depend on climatic and site condition (**?**Lauvernet and Muñoz-Carpena, 2018; **?**). Additionally, parameters were assumed to be independent in this study but this assumption may be too simplistic, especially for ~~hydrodynamical~~ hydrodynamic parameters. Further research should adapt the tested methods to dependent parameters. Dealing with dependent parameters has already been explored in the case of Sobol' indices (**?**) but it needs to be further explored in the case of HSIC and RF-based methods. It would also be necessary to investigate sensitivity of some time series to get a more comprehensive vision of the model functioning. To do so, the temporal series can be analyzed as a multivariate output for example with clustering-based GSA (**?**) or using the principal components of the model's functional outputs. The definition and the use of adequate hydrological signatures such as proposed in **?** and **?** may also be of interest to understand space-time variability and to capture a broader range of physical processes.

Global sensitivity analysis is a necessary but not yet systematic step to model evaluation, especially in the case of spatialized, risk assessment models that can be complex to deal with. This study proposes a comprehensive method based on complementary indices and thus paves the way for systematic analysis of such environmental exposure models.

References

815   Abily, M., Bertrand, N., Delestre, O., Gourbesville, P., and Duluc, C.-M.: Spatial global sensitivity analysis of high reso­lution classified topographic data use in 2D urban flood modelling, Environmental Modelling & Software, 77, 183–195, https://doi.org/10.1016/j.envsoft.2015.12.002, 2016.

Adriaanse, P.: Exposure assessment of pesticides in field ditches: The TOXSWA model, Pesticide Science, 49, 210–212, https://doi.org/10.1002/(SICI)1096-9063(199702)49:2<210::AID-PS496>3.0.CO;2-1, 1997.

820   Alletto, L., Pot, V., Giuliano, S., Costes, M., Perdrieux, F., and Justes, E.: Temporal variation in soil physical properties improves the water dynamics modeling in a conventionally-tilled soil, Geoderma, 243-244, 18–28, https://doi.org/10.1016/j.geoderma.2014.12.006, 2015.

Antoniadis, A., Lambert-Lacroix, S., and Poggi, J.-M.: Random forests for global sensitivity analysis: A selective review, Reliability Engi­neering and System Safety, 206, 107 312, https://doi.org/10.1016/j.ress.2020.107312, 2021.

Arcement, G. and Schneider, V.: Guide for selecting Manning's roughness coefficients for natural channels and flood plains, Tech. rep., U.S.
825   G.P.O, 1989.

Archer, G. E. B., Saltelli, A., and Sobol, I. M.: Sensitivity measures, ANOVA-like Techniques and the use of bootstrap, Journal of Statistical Computation and Simulation, 58, 99–120, https://doi.org/10.1080/00949659708811825, 1997.

Aulia, A., Jeong, D., Mohd Saaid, I., Kania, D., Taleb Shuker, M., and El-Khatib, N. A.: A Random Forests-based sen­sitivity analysis framework for assisted history matching, Journal of Petroleum Science and Engineering, 181, 106 237,
830   https://doi.org/10.1016/j.petrol.2019.106237, 2019.

Bénard, C., Da Veiga, S., and Scornet, E.: MDA for random forests: inconsistency, and a practical solution via the Sobol-MDA, working paper or preprint, 2021.

Blatman, G. and Sudret, B.: Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics, 230, 2345–2367, https://doi.org/10.1016/j.jcp.2010.12.021, 2011.

835   Borgonovo, E., Castaings, W., and Tarantola, S.: Moment independent importance measures: new results and analytical test cases, Risk Analysis, 31, 404–428, https://doi.org/10.1111/j.1539-6924.2010.01519.x, 2011.

Branger, F., Braud, I., Debionne, S., Viallet, P., Dehotin, J., Henine, H., Nedelec, Y., and Anquetin, S.: Towards multi-scale integrated hydrological models using the LIQUID® framework. Overview of the concepts and first application examples, Environmental Modelling and Software, 25, 1672–1681, https://doi.org/10.1016/j.envsoft.2010.06.005, 2010.

840   Breiman, L.: Bagging predictors, Machine learning, 24, 123–140, 1996.

Breiman, L.: Random forests, Machine learning, 45, 5–32, 2001.

Brown, C., Alix, A., Alonso-Prados, J.-L., Auteri, D., Gril, J.-J., Hiederer, R., Holmes, C., Huber, A., de Jong, F., M. Liess, S., Loutseti, Mackay, N., Maier, W.-M., Maund, S., Pais, C., Reinert, W., Russell, M., Schad, T., Stadler, R., Streloke, M., Styczen, M., and van de Zande, J.: Landscape and mitigation factors in aquatic risk assessment. Volume 2: detailed technic, Tech. rep., European Commission,
845   2007.

Buis, S., Piacentini, A., and Déclat, D.: PALM: a computational framework for assembling high-performance computing applications, Con­currency and Computation: Practice and Experience, 18, 231–245, https://doi.org/10.1002/cpe.914, 2006.

Buytaert, W., Reusser, D., Krause, S., and J.-P., R.: Why can't we do better than Topmodel?, Hydrological Processes, 22, 4175–4179, https://doi.org/10.1002/hyp.7125, 2008.

850 Caisson, A.: Prise en main et application d'un modèle spatialisé à base physique (CATHY) sur un versant expérimental pour la mise en place d'un système d'assimilation de données, Master's thesis, ENGEES, 2019.

Campbell, N., D'Arcy, B., Frost, A., Novotny, V., and Sampson, A.: Diffuse pollution: an introduction to the problems and solutions, IWA Publishing, 2004.

Carsel, E., J. and Baldwin, J. E.: PRZM-3, A model for predicting pesticide and nitrogen fate in the crop root and unsaturated soil zones,
855 Tech. rep., Environmental Protection Agency, 2000.

Catalogne, C., Lauvernet, C., and Carluer, N.: Guide d'utilisation de l'outil BUVARD pour le dimensionnement des bandes tampons végétalisées destinées à limiter les transferts de pesticides par ruissellement, Tech. rep., Agence française pour la biodiversité, 2018.

Coutadeur, C., Coquet, Y., and Roger-Estrade, J.: Variation of hydraulic conductivity in a tilled soil, European Journal of Soil Science, 53, 619–628, https://doi.org/10.1046/j.1365-2389.2002.00473.x, 2002.

860 Da Veiga, S.: Global sensitivity analysis with dependence measures, Journal of Statistical Computation and Simulation, 85, 1283–1305, https://doi.org/10.1080/00949655.2014.945932, 2015.

Dairon, R.: Identification des processus dominants de transfert des produits phytosanitaires dans le sol et évaluation de modèles numériques pour des contextes agro-pédo-climatiques variés, Ph.D. thesis, Université Claude Bernard - Lyon 1, 2015.

Darcy, H.: Recherches expérimentales relatives au mouvement de l'eau dans les tuyaux, Impr. Impériale, 1857.

865 De Lozzo, M. and Marrel, A.: New improvements in the use of dependence measures for sensitivity analysis and screening, Journal of Statistical Computation and Simulation, https://doi.org/10.1080/00949655.2016.1149854, 2014.

De Lozzo, M. and Marrel, A.: Sensitivity analysis with dependence and variance-based measures for spatio-temporal numerical simulators, Stochastic Environmental Research and Risk Assessment, 31, https://doi.org/10.1007/s00477-016-1245-3, 2016.

Dehotin, J., Braud, I., Vazquez, R., Debionne, S., and Viallet, P.: Prise en compte de l'hétérogénéité des surfaces continentales dans la
870 modélisation couplées zone non saturé-zone saturée, Bulletin du GFHN, 54, 57–62, 2008.

Dosskey, M. G., Helmers, M. J., and Eisenhauer, D. E.: A design aid for sizing filter strips using buffer area ratio, Journal of Soil and Water Conservation, 66, 29–39, https://doi.org/10.2489/jswc.66.1.29, 2011.

Dubus, I. G. and Brown, C. D.: Sensitivity and First-Step Uncertainty Analyses for the Preferential Flow Model MACRO, Journal of Environmental Quality, 31, 227–240, https://doi.org/10.2134/jeq2002.2270, 2002.

875 Dubus, I. G., Brown, C. D., and Beulke, S.: Sensitivity analyses for four pesticide leaching models, Pest Management Science, 59, 962–982, https://doi.org/10.1002/ps.723, 2003.

Durand, C.: Modélisation du transfert de pesticides à l'échelle de la parcelle. Application au bassin versant de la Morcille (Nord Beaujolais, 69) et analyse de sensibilité du modèle, Master's thesis, ENGEES, 2014.

Faivre, R., Iooss, B., Mahévas, S., Makowski, D., and Monod, H.: Analyse de sensibilité et exploration de modèles, Collection Savoir-Faire,
880 Editions Quae, 2013.

Fajraoui, N., Ramasomanana, F., Younes, A., Mara, T., Ackerer, P., and Guadagnini, A.: Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media, Water Resources Research, 47, https://doi.org/10.1029/2010WR009639, 2011.

FOCUS: FOCUS surface water scenarios in the EU evaluation process under 91/414/EEC, European commission, report of the FOCUS
885 Working Group on Surface Water Scenarios, EC Document Reference SANCO/4802/2001, 2001.

Fouilloux, A. and Piacentini, A.: The PALM Project: MPMD paradigm for an oceanic data assimilation software, in: Euro-Par'99 Parallel Processing: 5th International Euro-Par Conference Toulouse, France, August 31 - September 3, 1999 Proceedings, pp. 1423–1430, Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.

Fox, G. A., Muñoz-Carpena, R., and Sabbagh, G. J.: Influence of flow concentration on parameter importance and prediction uncertainty of pesticide trapping by vegetative filter strips, Journal of Hydrology, 384, 164 – 173, https://doi.org/10.1016/j.jhydrol.2010.01.020, 2010.

Frésard, F.: Cartographie des sols d'un petit bassin versant en Beaujolais viticole, en appui à l'évaluation du risque de contamination des eaux par les pesticides, Master's thesis, Université de Franche Comté, 2010.

Gamboa, F., Janon, A., Klein, T., and Lagnoux, T.: Sensitivity indices for multivariate outputs, Comptes Rendus Mathématiques, 351, 307 – 310, https://doi.org/10.1016/j.crma.2013.04.016, 2013.

Gao, B., Walter, M., Steenhuis, T., Hogarth, W., and Parlange, J.: Rainfall induced chemical transport from soil to runoff: theory and experiments, Journal of Hydrology, 295, 291 – 304, https://doi.org/10.1016/j.jhydrol.2004.03.026, 2004.

Garcia, D., Arostegui, I., and Prellezo, R.: Robust combination of the Morris and Sobol methods in complex multidimensional models, Environmental Modelling and Software, 122, 104 517, https://doi.org/10.1016/j.envsoft.2019.104517, 2019.

Gascuel-Odoux, C., Aurousseau, P., Cordier, M.-O., Durand, P., Garcia, F., Masson, V., Salmon-Monviola, J., Tortrat, F., and Trepos, R.: A decision-oriented model to evaluate the effect of land use and agricultural management on herbicide contamination in stream water, Environmental Modelling & Software, 24, 1433 – 1446, https://doi.org/10.1016/j.envsoft.2009.06.002, special issue on simulation and modelling in the Asia-Pacific region, 2009.

Gatel, L., Lauvernet, C., Carluer, N., Weill, S., Tournebize, J., and Paniconi, C.: Global evaluation and sensitivity analysis of a physically based flow and reactive transport model on a laboratory experiment, Environmental Modelling and Software, 113, 73–83, https://doi.org/10.1016/j.envsoft.2018.12.006, 2018.

Gatel, L., Lauvernet, C., Carluer, N., Weill, S., and Paniconi, C.: Sobol global sensitivity analysis of a coupled surface/subsurface water flow and reactive solute transfer model on a real hillslope, Water, 12, 121, https://doi.org/10.3390/w12010121, 2019.

Ghanem, R. G. and Spanos, P. D.: Spectral stochastic finite-element formulation for reliability analysis, Journal of Engineering Mechanics, 117, 2351–2372, 1991.

Gouy, V., Liger, L., Carluer, N., and Margoum, C.: Site Atelier Ardières Morcille, Irstea, https://doi.org/10.17180/obs.saam, 2015.

Gregorutti, B., Michel, B., and Saint-Pierre, P.: Correlation and variable importance in random forests, Statistics and Computing, 27, 659–678, https://doi.org/10.1007/s11222-016-9646-1, 2017.

Gretton, A., Bousquet, O., Smola, A., and Schölkopf, B.: Measuring statistical dependence with Hilbert-Schmidt norms, in: International conference on algorithmic learning theory, pp. 63–77, Springer, 2005a.

Gretton, A., Herbrich, R., Smola, A., Bousquet, O., and Schölkopf, B.: Kernel Methods for Measuring Independence, J. Mach. Learn. Res., 6, 2075–2129, 2005b.

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrological Processes, 22, 3802–3813, https://doi.org/10.1002/hyp.6989, 2008.

Harper, E. B., Stella, J. C., and Fremier, A. K.: Global sensitivity analysis for complex ecological models: a case study of riparian cottonwood population dynamics, Ecological Applications, 21, 1225–1240, https://doi.org/10.1890/10-0506.1, 2011.

Herman, J. D., Kollat, J. B., Reed, P. M., and Wagener, T.: Technical Note: Method of Morris effectively reduces the computational demands of global sensitivity analysis for distributed watershed models, Hydrology and Earth System Sciences, 17, 2893–2903, https://doi.org/10.5194/hess-17-2893-2013, 2013.

Hong, T. and Purucker, S. T.: Spatiotemporal sensitivity analysis of vertical transport of pesticides in soil, Environmental Modelling and Software, 105, 24–38, https://doi.org/10.1016/j.envsoft.2018.03.018, 2018.

Ishwaran, H. and Kogalur, U.: Fast Unified Random Forests for Survival, Regression,and Classification (RF-SRC). R package version 2.9.3., 2020.

Kraft, P., Vache, K. B., Frede, H.-G., and Breuer, L.: CMF: A Hydrological Programming Language Extension For Integrated Catchment Models, Environmental Modelling and Software, 26, 828–830, https://doi.org/doi.org/10.1016/j.envsoft.2010.12.009, 2012.

Larsbo, M. and Jarvis, N.: MACRO 5.0: A model of water flow and solute transport in macroporous soil : technical description, Emergo (Uppsala), Department of Soil Sciences, Swedish University of Agricultural Sciences, 2003.

Lauvernet, C. and Muñoz-Carpena, R.: Shallow water table effects on water, sediment, and pesticide transport in vegetative filter strips – Part 2: model coupling, application, factor importance, and uncertainty, Hydrology and Earth System Sciences, 22, 71–87, https://doi.org/10.5194/hess-22-71-2018, 2018.

Le Gratiet, L., Marelli, S., and Sudret, B.: Metamodel-based sensitivity analysis: polynomial chaos expansions and Gaussian processes, chap. 38, pp. 1289–1325, Springer International Publishing, cham, Switzerland, 2017.

Lewis, K.-A., Tzilivakis, J., Warner, D., and Green, A.: An international database for pesticide risk assessments and management, Human and Ecological Risk Assessment: An International Journal, 22, 1050–1064, https://doi.org/10.1080/10807039.2015.1133242, 2016.

Li, K., De Jong, R., and Boisvert, J.: An exponential root-water-uptake model with water stress compensation, Journal of Hydrology, 252, 189–204, https://doi.org/10.1016/S0022-1694(01)00456-5, 2001.

Lighthill, M. J. and Whitham, G. B.: On kinematic waves I. Flood movement in long rivers, Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences, 229, 281–316, https://doi.org/10.1098/rspa.1955.0088, 1955.

Liu, H., Chen, W., and Sudjianto, A.: Relative entropy based method for probabilistic sensitivity analysis in engineering design, Journal of Mechanical Design, 128, 326–336, https://doi.org/10.1115/1.2159025, 2005.

Marelli, S. and Sudret, B.: UQLab: A framework for uncertainty quantification in Matlab, in: Proc. 2nd Int. Conf. on Vulnerability, Risk Analysis and Management (ICVRAM2014), 2014.

Marelli, S. and Sudret, B.: An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, Structural Safety, 75, 67–74, https://doi.org/10.1016/j.strusafe.2018.06.003, 2018.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 21, 239–245, https://doi.org/10.2307/1268522, 1979.

MeteoFrance: Evapotranspiration potentielle MONTHEIH, 2008.

Metta, J. W.: Sensitivity analysis of the Catchment Modeling Framework (CMF) and use in evaluating two agricultural management scenarios, Master's thesis, Oregon State University, 2007.

Meynaoui, A.: New developments around dependence measures for sensitivity analysis : application to severe accident studies for generation IV reactors, Ph.D. thesis, INSA, Toulouse, 2019.

Meynaoui, A., Marrel, A., and Laurent-Bonneau, B.: Méthodologie basée sur les mesures de dépendance HSIC pour l'analyse de sensibilité de second niveau, in: 50èmes Journées de Statistique (JdS2018), Palaiseau, France, 2018.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33, 161–174, https://doi.org/10.1080/00401706.1991.10484804, 1991.

960 Moussa, R., Colin, F., Dagès, C., Fabre, J.-C., Lagacherie, P., Louchart, X., Rabotin, M., Raclot, D., and Voltz, M.: Distributed hydrological modelling of farmed catchments (MHYDAS) : assessing the impact of man-made structures on hydrological processes, in: LAND-MOD2010, Symposcience - Éditions Quae, France, Europe, 2010.

Muñoz-Carpena, R., Parsons, J., and Gilliam, J.: Modeling hydrology and sediment transport in vegetative filter strips, Journal of Hydrology, 214, 111–129, https://doi.org/10.1016/S0022-1694(98)00272-8, 1999.

965 Muñoz-Carpena, R., Fox, G., and Sabbagh, G.: Parameter importance and uncertainty in predicting runoff pesticide reduction with filter strips, Journal of environmental quality, 39, 630–41, https://doi.org/10.2134/jeq2009.0300, 2010.

Nossent, J. and Bauwens, W.: Multi-variable sensitivity and identifiability analysis for a complex environmental model in view of integrated water quantity and water quality modeling, Water Science and Technology, 65, 539–549, https://doi.org/10.2166/wst.2012.884, 2012.

Peyrard, X., Liger, L., Guillemain, C., and Gouy, V.: A trench study to assess transfer of pesticides in subsurface lateral flow for a soil with

970 contrasting texture on a sloping vineyard in Beaujolais, Environmental Science and Pollution Research, 2016.

Pianosi, F., Beven, K., Freer, J., Hall, J., Rougier, J., Stephenson, D., and Wagener, T.: Sensitivity analysis of environmental models: A systematic review with practical workflow, Environmental Modelling & Software, 79, 214 – 232, https://doi.org/10.1016/j.envsoft.2016.02.008, 2016.

Randriambololohasinirina, P.: Pesticide dissipation properties in soils of a wine-growing watershed., Master's thesis, Université Pierre et

975 Marie Curie (Paris 6) ; Institut des Sciences et Industries du Vivant et de l'Environnement., 2012.

Reichenberger, S., Bach, M., Skitschak, A., and Frede, H.-G.: Mitigation strategies to reduce pesticide inputs into ground- and surface water and their effectiveness; A review, Science of The Total Environment, 384, 1–35, https://doi.org/10.1016/j.scitotenv.2007.04.046, 2007.

Rodriguez-Galiano, V., Mendes, M., Garcia-Soldado, M., Chica-Olmo, M., and Ribeiro, L.: Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural

980 setting (Southern Spain), Science of The Total Environment, 476-477, 189 – 206, https://doi.org/10.1016/j.scitotenv.2014.01.001, 2014.

Ross, P. J.: Modeling soil water and solute transport - fast, simplified numerical solutions, Agronomy Journal, 95, 1352–1361, https://doi.org/10.2134/agronj2003.1352, 2003.

Ross, P. J.: Fast solution of Richards' equation for flexible soil hydraulic property descriptions, Tech. rep., CSIRO, 2006.

Rouzies, E., Lauvernet, C., Barachet, C., Morel, T., Branger, F., Braud, I., and Carluer, N.: From agricultural catchment to management

985 scenarios: A modular tool to assess effects of landscape features on water and pesticide behavior, Science of The Total Environment, 671, 1144 – 1160, https://doi.org/10.1016/j.scitotenv.2019.03.060, 2019.

Saint-Geours, N.: Analyse de sensibilité de modèles spatialisés : application à l'analyse coût-bénéfice de projets de prévention du risque d'inondation, Ph.D. thesis, Université de Montpellier 2, 2012.

Saint-Geours, N., Bailly, J.-S., Grelot, F., and Lavergne, C.: Multi-scale spatial sensitivity analysis of a model for economic appraisal of flood

990 risk management policies, Environmental Modelling & Software, 60, 153–166, https://doi.org/10.1016/j.envsoft.2014.06.012, 2014.

Saltelli, A., Tarantola, S., Campolongo, F., and Ratto, M.: Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models, Wiley, https://doi.org/10.1002/0470870958, 2004.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

995 Sarrazin, F., Pianosi, F., and Wagener, T.: Global sensitivity analysis of environmental models: convergence and validation, Environmental Modelling & Software, 79, 135 – 152, https://doi.org/10.1016/j.envsoft.2016.02.005, 2016.

Schaap, M. G. and van Genuchten, M. T.: A modified Mualem-van Genuchten formulation for improved description of the hydraulic conductivity near saturation, Vadose Zone Journal, 5, 27–34, 2006.

Schwen, A., Bodner, G., Scholl, P., Buchan, G., and Loiskandl, W.: Temporal dynamics of soil hydraulic properties and the water-conducting porosity under different tillage, Soil and Tillage Research, 113, 89–98, https://doi.org/10.1016/j.still.2011.02.005, 2011.

Seki, K.: SWRC fit - a nonlinear fitting program with a water retention curve for soils having unimodal and bimodal pore structure, Hydrology and Earth System Sciences Discussions, 4, 407–437, https://doi.org/10.5194/hessd-4-407-2007, 2007.

Smart, D., Schwass, E., Lakso, A., and Morano, L.: Grapevine rooting patterns: A comprehensive analysis and a review, American Journal of Enology and Viticulture, 57, 89–104, 2006.

Smith, J. D., Lin, L., Quinn, J. D., and Band, L. E.: Guidance on evaluating parametric model uncertainty at decision-relevant scales, Hydrology and Earth System Sciences Discussions, 2021, 1–30, https://doi.org/10.5194/hess-2021-324, 2021.

Sobol, I.: Sensitivity estimates for nonlinear mathematical models, Mathematical Modelling and Computational Experiments, 1, 407–414, 1993.

Sobol, I.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation, 55, 271 – 280, https://doi.org/10.1016/S0378-4754(00)00270-6, the Second IMACS Seminar on Monte Carlo Methods, 2001.

Soleimani, F.: Analytical seismic performance and sensitivity evaluation of bridges based on random decision forest framework, Structures, 32, 329–341, https://doi.org/10.1016/j.istruc.2021.02.049, 2021.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., and Xu, C.: Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications, Journal of Hydrology, 523, 739 – 757, https://doi.org/10.1016/j.jhydrol.2015.02.013, 2015.

Sudret, B.: Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety, 93, 964 – 979, https://doi.org/10.1016/j.ress.2007.04.002, 2008.

Székely, G., Rizzo, M., and Bakirov, N.: Measuring and testing dependence by correlation of distances, The annals of statistics, 35, 2769–2794, 2007.

Tarantola, S., Giglioli, N., Jesinghaus, J., and Saltelli, A.: Can global sensitivity analysis steer the implementation of models for environmental assessments and decision-making?, Stochastic Environmental Research and Risk Assessment, 16, 63–76, https://doi.org/10.1007/s00477-001-0085-x, 2002.

Tortrat, F.: Modélisation orientée décision des processus de transfert par ruissellement et subsurface des herbicides dans les bassins versants agricoles., Ph.D. thesis, Ecole nationale supérieure d'agronomie de Rennes, 2005.

Van den Bogaert, R.: Typologie des sols du bassin versant de la Morcille, caractérisation de leur propriétés hydrauliques et test de fonctions de pédotransfert, Master's thesis, Université Pierre et Marie Curie, AgroParisTech, 2011.

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., and Srinivasan, R.: A global sensitivity analysis tool for the parameters of multi-variable catchment models, Journal of Hydrology, 324, 10 – 23, https://doi.org/10.1016/j.jhydrol.2005.09.008, 2006.

Vanuytrecht, E., Raes, D., and Willems, P.: Global sensitivity analysis of yield output from the water productivity model, Environmental Modelling and Software, 51, 323–332, https://doi.org/10.1016/j.envsoft.2013.10.017, 2014.

Varado, N., Braud, I., and Ross, P.: Development and assessment of an efficient vadose zone module solving the 1D Richards' equation and including root extraction by plants, Journal of Hydrology, 323, 258–275, https://doi.org/10.1016/j.jhydrol.2005.09.015, 2006.

Walter, M., Gao, B., and Parlange, J.-Y.: Modeling soil solute release into runoff with infiltration, Journal of Hydrology, 347, 430 – 437, https://doi.org/10.1016/j.jhydrol.2007.09.033, 2007.

1035 Wang, S., Huang, G., Baetz, B., and Huang, W.: A polynomial chaos ensemble hydrologic prediction system for efficient parameter inference and robust uncertainty assessment, Journal of Hydrology, 530, 716 – 733, https://doi.org/10.1016/j.jhydrol.2015.10.021, 2015.

Wiener, N.: The homogeneous chaos, American Journal of Mathematics, 60, 897–936, 1938.

Xiu, D. and Karniadakis, G.: The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing, 24, 619–644, https://doi.org/10.1137/S1064827501387826, 2002.

1040 Zajac, Z., B.: Global sensitivity and uncertainty analysis of spatially distributed watershed models, Ph.D. thesis, University of Florida, 2010.