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~~ could be applied to the recent PESHMELBA pesticide transfer model to quantify uncertainties on transfer simulations. We set up a virtual catchment based on a real one and we ~~combined~~ compared 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. After a preliminary screening step, we calculated Sobol' indices obtained from Polynomial Chaos Expansion, HSIC dependence measures and feature importance measures obtained from Random Forest surrogate model~~to get a comprehensive overview of PESHMELBA sensitivity~~. Results from the different methods were compared regarding both the information they provide and their computational cost. 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 hydrodynamic behaviour. ~~The methodology proposed in this paper may be extended to other~~ Conclusions about the advantages and disadvantages of each method may help modellers to conduct global sensitivity analysis on other such modular and distributed hydrological models as there has been a growing interest in these ~~methods~~ approaches in recent years.

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 ecosystems 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~~ To do so, numerical models that simulate pesticide transfers and fate are necessary tools to support risk management~~as they make simulation of contamination and transfers possible. They also~~ . Among others things, such models 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, scenarios of exposure and to assess mitigation measures. For this purpose and to support decision-making, it is important to use physically based models 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, are particularly valuable. When they are distributed at the scale of the catchment, these models also make it possible to take into account the landscape configuration. This is of particular interest as the landscape configuration is of big influence on transfers and regulation and corrective actions can relevantly be thus be relevantly set up at this scale. Building PESHMELBA (Rouzies et al., 2019) is such a modular, process-based and distributed models is highly appropriate to address this scale and to take model that simulates pesticide transfers and fate at the scale of small agricultural catchments and that takes into account the landscape composition, (Buytaert et al., 2008; Kraft et al., 2012). This recent but promising approach consists in coupling different blocks representing one or several processes, often based on a modeling framework (Tortrat, 2005; Moussa et al., 2010; Branger et al., 2010; Rouzies et al., 2019)impact of the landscape composition. However, simulating at such scale also implies additional difficulties. First, catchment-scale models 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 PESHMELBA is characterized by a huge number of parameters that makes the parameter setting complex . Using such models thus raises complex structure, numerous couplings and interactions that raise additional challenges about diagnosing model behaviour and uncertainty quantification (Gupta et al., 2008; van Griensven et al., 2006) . (Gupta et al., 2008; van Griensven et al., 2006). To address the issue of uncertainty quantification, 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, formulating realistic assumptions for its use and paves the way for potential improvements (Faivre et al., 2013).is being increasingly used in environmental models (Hamby, 1994; Tang et al., 2007; Nossent et al., 2011; Garcia et al., 2019; Alipour et al., 2022). Among its objectives, sensitivity analysis contributes to get greater insight about the model behavior. It also contributes to identify which input factors should be best characterized so as to significantly reduce the total uncertainty.

In the field of pesticide transfer modelling, Dubus et al. (2003) first performed a local, and Holvoet et al. (2005) first performed one-at-a-time (OAT) sensitivity analysison 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 varies all input factors simultaneously, on their entire ranges of definition, are is 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 application characteristics (Song et al., 2015; Pianosi et al., 2016; Sarrazin et al., 2016). In the case field of pesticide transfer modelling, variance decomposition and the studies mainly use variance decomposition methods. This approach is particularly appreciated as it allows to characterize not only the contribution of each input factor individually but also

Reference	Model name	Nb. of input factors for ranking	Method	Nb. of model evaluations
Fox et al. (2010)	VFSMOD	18	eFAST	14,977
Lauvernet and Muñoz-Carpena (2018)	VFSMOD	24	eFAST	75,544
Hong and Purucker (2018)	PRZM	11	Sobol	195,000
Gatel et al. (2019)	CATHY	15	Sobol/Sobol with ROA-LHS sampling (Tissot and Prieur, 2015)	1,922/17,000
D'Andrea et al. (2020)	PWC	24	Sobol	11,776
Faúndez Urbina et al. (2020)	SWAP-PEARL	<14	Sobol	>18,000

Table 1. Examples of global sensitivity analysis performed on pesticide transfer models.

the contributions of interactive effects between input factors. In addition, sensitivity indices calculated from variance-based method are easy to interpret because they represent the portion of output variance that can be apportioned to an input factor (Saltelli, 2002). The Sobol' method (Sobol, 1993) ~~have~~ is the most popular variance-based method and it has been widely used ~~(e.g. Hong and Purucker, 2018; Gatel et al., 2019). Although very popular, such methods have several limitations. First, they~~ in pesticide studies (e.g. Hong and Purucker, 2018; Gatel et al., 2019; D'Andrea et al., 2020; Faúndez Urbina et al., 2020). In some other studies (Fox et al., 2010; Lauvernet and Muñoz-Carpena, 2018) variance decomposition is also performed based on Fourier Transform methods (FAST and eFAST methods). However, such variance decomposition methods are characterized by a high computational cost ~~although it can be alleviated using metamodelling techniques~~(meaning that they require a lot of model evaluations to compute sensitivity indices) that cannot be always afforded. As the computational cost quickly increases with the number of input factors, a classical approach consists in first applying a *screening* step to identify (if there are) parameters that have a negligible influence on the output variability. Non-influential parameters are then removed or set to constant values. Once the input space dimension has been reduced, variance decomposition can be applied on the reduced set of input factors to classify remaining input factors according to their relative contribution to output variability (*ranking* step).

Table 1 summarizes the characteristics of the GSA application for the previously cited pesticide studies (such table is meant to provide some typical examples rather than an exhaustive list of studies on the field of pesticide modelling). It shows that in these studies, the input space scrutinized for ranking contains at most 24 parameters (in some cases after a screening step)

and that the ranking step always uses more than 10,000 simulations. However, in the case of the PESHMELBA model (and probably of other distributed, physically based, hydrological and pesticide models) a classical application involves around 150 parameters. In addition, 10,000 simulations may be very hard to reach in realistic applications due to computation time. Such characteristics make classical approaches for variance-based GSA quite hard to apply on PESHMELBA. Then, there is a need for exploring other, low computational cost GSA methods to perform a ranking task in the PESHMELBA model. Recently, new GSA approaches that require much less model evaluations have emerged. For instance, computing a metamodel based on polynomial chaos expansion on the model (Ghanem and Spanos, 1991) allows to compute Sobol' sensitivity indices directly, for a very limited computational cost (Sudret, 2008; Fajraoui et al., 2011; Wang et al., 2015). ~~Second, from a methodological~~ From a totally different 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~~ the Hilbert-Schmidt Independence Criterion (HSIC) proposed by Gretton et al. (2005b) is used as a sensitivity measure in Da Veiga (2015). It describes the global 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 pesticide transfer modelling where the variance decomposition may not be relevant. Additionally~~also from a very limited computational budget. Finally, 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~~ 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 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; Gatel et al., 2018; Garcia~~ ~~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, 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. 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* (or *block*) sensitivity indices to make reference to sensitivity indices synthetized on the whole domain. Site sensitivity indices result in sensitivity maps that detail spatial contributions 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. However, such innovative methods have never been applied to complex, distributed pesticide transfer model before.

In this paper, we explore how the sensitivity of risk-assessment models can be relevantly tackled based on the example of

Then, the objective of this paper is to evaluate and to compare three new, low computational cost GSA methods on an application of the PESHMELBA model. The methods will be especially compared in terms of interpretability of the sensitivity indices they provide and reliability (based on the study of their convergence rate). We also investigate whether such approaches suit to the process-oriented, modular PESHMELBAmodel (PESticide and Hydrology: ModELling at the catchment scale). We perform a GSA of the PESHMELBA model 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. 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 modelspatialized aspect of PESHMELBA. To do so, we investigate the relevancy of computing both local and aggregated indices following the recommandations from Saint-Geours (2012). 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. The input sampling is described in Section 2.3, then we introduce the different GSA approaches and the methodology used for landscape analysis from Section 2.4 to Section 2.6. Results are presented in Section 3focusing . They focus successively on screening (Section 3.1), comparison of GSA methods (Section 3.2) and spatial analysis (Section 3.3). Finally, Section 4 gathers some points of discussion.

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 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 is set from a portion of la Morcille catchment (France) in order to explore the different GSA methods and to ease interpretation of spatialized results. The chosen portion 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 is composed of 10 vineyard plots, 4 vegetative filter strips and 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 use the classification from Frésard (2010) that groups soil types into ~~3~~ three 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 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). Spatial arrangement is set in order to be as realistic as possible in terms of possible interfaces between SUs. Each SU is set at least on one vineyard plot and one VFS on the virtual scenario.

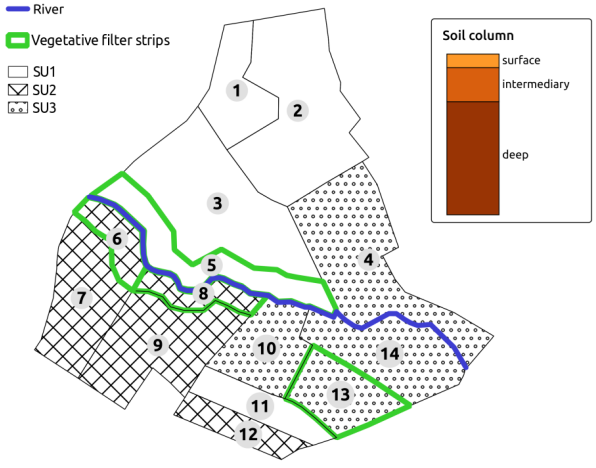

Figure 2. Soil type locations for the case study. Green contours show the vegetative filter strips.

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 2 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 moc are available from Van den Bogaert (2011) and Randriambololohasinirina (2012). They are, as well as hydrodynamic parameters for each soil type, described in Table 3. Retention values measured by Van den Bogaert (2011) are used to fit retention curve using SWRCfit tool (Seki, 2007). A Schaap-Van Genuchten conductivity curve is used (Schaap and van Genuchten, 2006; Ross, 2006) whose matching point at saturation Ko and empirical pore-connectivity L are derived from conductivity data and retention parameters from retention curve fitting. Surface organic carbon content 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 are characterized by a 2.8%-organic carbon content (Randriambololohasinirina, 2012) and a saturated hydraulic conductivity of 150 $mm.h^{-1}$ (or $4.31 \cdot 10^{-5}$ $m.s^{-1}$) 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 is considered, thus the ratio Ks_h/Ks_v is set to 1 on the catchment.

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 use a Freundlich isotherm to describe its adsorption equilibrium. Adsorption parameters are obtained from Lewis et al. (2016) (Koc = 769 $mL.g^{-1}$, Freundlich isotherm exponent = 0.84). Surface degradation coefficient is also taken from Lewis et al. (2016) ($DT50$ = 47.1 days) and a decreasing degradation rate in function of depth is set as in FOCUS (2001). A 500g-application is considered at the beginning of the simulation on plots 2 and 7 (see Figure 1). Most of the transformation and adsorption of tebuconazole is supposed to happen on plots and VFSs at this modelling scale. Therefore, no adsorption or degradation is simulated in the river.

A no-flux boundary condition is applied on all sides except on surface where rain and potential evapotranspiration are considered. Rain data are extracted from BDOH database (Gouy et al., 2015). A 3-month simulation is performed on a time period characterized by long and intense rain events (670 mm cumulated) allowing for significant water and pesticide transfers both by surface runoff and subsurface saturated transfers. Potential evapotranspiration (PET) data are obtained from MeteoFrance for the neighbouring site of Liergues (MeteoFrance, 2008). Data are averaged over 10-day periods and corrected by a factor -11 % to match La Morcille site as recommended in Durand (2014) and Caisson (2019). Rain and PET data for the simulation are summarized in Figure 3.

Input factor [units]	Description	Spatial level definition	Nominal value
	Soil parameters		
$thetas$ [m^3m^{-3}]	Saturated water content	Soil horizon	see Table 3
$thetar$ [m^3m^{-3}]	Residual water content	Soil horizon	see Table 3
Ks [ms^{-1}]	Saturated hydraulic conductivity	Soil horizon	
hg [m]	Air-entry pressure in Van Genuchten retention characteristic curve	Soil horizon	see Table 3
mn	Deduced parameter from Van Genuchten retention characteristic curve n: mn = n-1	Soil horizon	see Table 3
Ko [ms^{-1}]	Matching point at saturation in modified Mualem Van Genuchten conductivity curve (Schaap and van Genuchten, 2006)	Soil horizon	see Table 3
L	Empirical pore-connectivity parameter	Soil horizon	see Table 3
bd [gcm^{-3}]	Bulk density	Soil horizon	see Table 3
moc [gg^{-1}]	Organic carbon content	Soil horizon	see Table 3
	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 zone	River reach	1.5
Ks [ms^{-1}]	Saturated conductivity of the river bed	River reach	$2.38 \cdot 10^{-5}$
$manning$ [$sm^{-1/3}$]	River bed Manning's roughness	River reach	0.079
	Plot and VFS parameters		
$hpond$ [m]	Ponding height	Plot-VFS	0.01-0.05
$adsorpthick$ [m]	Mixing layer thickness	All landscape elements (except river)	0.01

Table 2. Input factor description, corresponding spatial level definition and value for the nominal scenario. Nominal values are explicitly distinguished between vineyard plots and VFSs with the character '-' when needed.

	Horizon	depth	$thetas$	$thetar$	hg	n	Ks	Ko	L	bd	moc
		[m]	[m^3m^{-3}]	[m^3m^{-3}]	[m]	[-]	[ms^{-1}]	[ms^{-1}]	[-]	[gcm^{-3}]	[%]
SU1	11-14	0.05	0.34	0.04	$-9.69\cdot10^{-2}$	1.27	$3.93\cdot10^{-5}$-$4.31\cdot10^{-5}$	$2.86\cdot10^{-7}$	-8.43	1.34	0.91-2.80
	2	0.5	0.34	0.05	$-3.29\cdot10^{-2}$	1.20	$8.64\cdot10^{-5}$	$2.28\cdot10^{-7}$	-6.52	1.47	0.39
	3	0.65	0.32	0.08	$-2.09\cdot10^{-2}$	1.20	$5.39\cdot10^{-5}$	$7.47\cdot10^{-7}$	-4.24	1.57	0.10
	4	4	0.28	0.07	$-5.99\cdot10^{-2}$	1.23	$3.11\cdot10^{-5}$	$1.47\cdot10^{-6}$	-0.14	1.53	0.07
SU2	12-15	0.035	0.34	0.04	$-9.69\cdot10^{-2}$	1.27	$3.93\cdot10^{-5}$-$4.31\cdot10^{-5}$	$2.86\cdot10^{-7}$	-8.43	1.34	1.15-2.80
	6	0.4	0.35	0	$-6.60\cdot10^{-2}$	1.13	$2.16\cdot10^{-5}$	$3.19\cdot10^{-7}$	9.66	1.59	0.68
	7	0.55	0.32	0	$-7.18\cdot10^{-2}$	1.08	$9.60\cdot10^{-6}$	$1.67\cdot10^{-7}$	-10	1.66	0.35
	8	4	0.42	0	$-3.02\cdot10^{-1}$	1.08	$3.98\cdot10^{-6}$	$9.72\cdot10^{-8}$	10	1.54	0.28
SU3	13-16	0.06	0.34	0.04	$-9.69\cdot10^{-2}$	1.27	$3.93\cdot10^{-5}$-$4.31\cdot10^{-5}$	$2.86\cdot10^{-7}$	-8.43	1.34	0.75-2.80
	9	0.45	0.33	0.08	$-6.72\cdot10^{-2}$	1.26	$3.05\cdot10^{-5}$	$3.36\cdot10^{-7}$	0.42	1.46	0.37
	10	4	0.32	0.06	$-3.56\cdot10^{-2}$	1.18	$2.38\cdot10^{-5}$	$3\cdot10^{-7}$	1.05	1.62	0.40

Table 3. Soil parameters for SU1, 2 and 3. Hydrodynamic parameters are based on Van Genuchten retention curve and on Schaap-Van Genuchten conductivity curve fitting. Parameters are described in Table 2. 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.

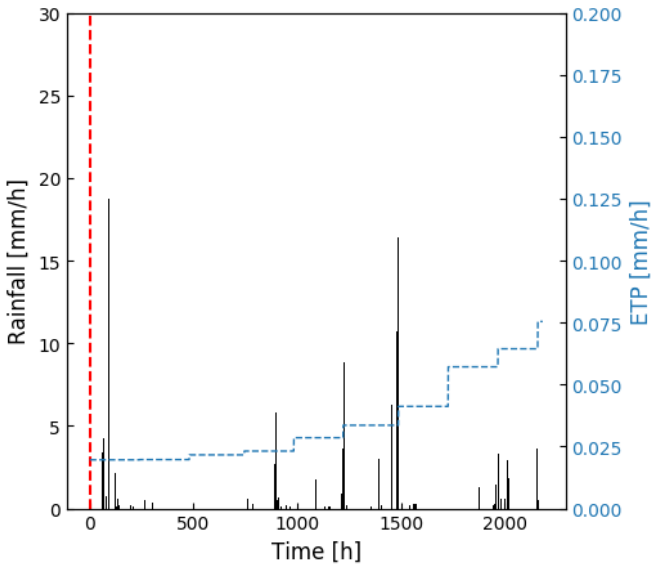

Figure 3. Climatic forcing (rain and potential evapotranspiration) for the simulation. The dotted red line stands for the one-shot pesticide application.

Although virtual, we aim at setting initial conditions as plausible as possible for this scenario. As running the model on a warm-up period is not possible due to data ~~availabilitiy~~ availability and computational cost limitation, initial water table levels 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 are available on a transect, perpendicular to the river. Data are extrapolated over the virtual hillslope width on both sides of the river. An upstream 0.177 m^3s^{-1}-flow is considered in the river based on local measurements (Gouy et al., 2015).

Two types of vegetation are represented in this scenario. Vineyard cover is considered on plots while permanent grassland 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 %) are considered for vineyards following values reported in Smart et al. (2006) and confirmed by expert knowledge in the area. The root depth (Zr) is set to 0.9 m and the root density in the first 10 % of the root depth ($F10$) is set to 33.5 % for grassland (Brown et al., 2007). For vineyards, the Leaf Area Index (LAI) 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 are taken from Brown et al. (2007) and they are detailed in Appendix **??**. 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 2 the reader should note that the LAI parameter relates to the constant LAI value for grassland while $LAImin$, $LAImax$ and $LAIharv$ relate to vineyard. The remaining parameters for root extraction model are also fixed to nominal values proposed by Varado et al. (2006); Li et al. (2001). Manning coefficients are set from data reported in Arcement and Schneider (1989). A mature row crop value (0.033 $s.m^{-1/3}$) is chosen for vineyard while a high grass pasture value (0.2 $s.m^{-1/3}$) is set for VFS cover.

Finally, ponding height is set to 0.01 m on vineyard plots while an increased value is set on VFSs (0.05 m). According to Gao et al. (2004) and Walter et al. (2007), the surface mixing layer thickness 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) is set to 1.5 m (ERT field measure, personal communication) and the saturated hydraulic conductivity (Ks) is set to $2.38 \cdot 10^{-5}$ ms^{-1} accordingly to local saturated conductivities in the neighbouring soil. The ponding height is 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 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; Dubus and Brown, 2002), 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 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 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 (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 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 Methodology for global sensitivity analysis in the PESHMELBA model

Although the PESHMELBA model is dynamic, model outputs considered in this paper are scalar quantities rather than temporal series to keep the problem simple. In order to investigate PESHMELBA abilities to properly represent transfers in a heterogeneous landscape, sensitivity analysis is performed on four hydrological and quality variables: 1/ cumulated water volume transferred in the subsurface (saturated lateral transfers), 2/ pesticide mass transferred in the subsurface (saturated lateral transfers) 3/ cumulated water volume transferred on surface (surface runoff), 4/ cumulated pesticide mass transferred on surface (surface runoff). However, these quantities are spatialized leading to multidimensional outputs. To deal with the spatialized aspect, GSA is first performed on scalar quantities, on each landscape element (see Sections 2.4.1 to 2.4.3), then sensitivity indices are aggregated in a second time providing catchment-scale sensitivity indices (see Section 2.6).

The full workflow used to perform GSA in the PESHMELBA model 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~~2,000-point LHS obtained from the reduced set of input parameters

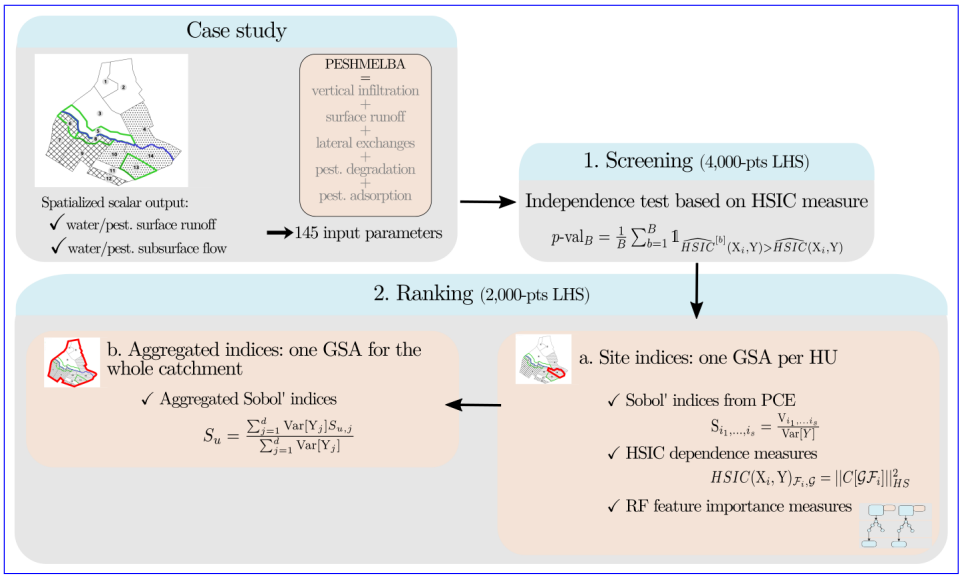

Figure 4. Full workflow used to perform GSA on the PESHMELBA model in 3 steps: 1/ Screening, 2.a/ Ranking at a local scale, 2.b/ Ranking at the catchment scale.

is computed to perform ranking. Sensitivity indices are computed from the ~~1,000~~2,000-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 (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~~In addition, they all belong to the category of 'given-data' methods which means that the input sample they consider does not require a specific structure. This is of particular interest for models that are computationally costly as it allows to take advantage of preexisting simulations. That is why such methods have known a growing interest in the recent years (Saltelli et al., 2021; Sheikholeslami et al., 2021). However, ~~as implementing several~~ such methods may not be ~~possible in every case studies, we also compare these methods~~ equally costly to set and they define the notion of sensitivity in contrasted ways. We compare them 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~~they provide, their accuracy and robustness.

~~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 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).~~

~~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.~~

In what follows, we denote $\mathrm{Y} \in \mathbb{R}$ a given scalar output from PESHMELBA. Y is function of a multivariate input random vector $\mathrm{Y} = \mathcal{M}(\underline{\mathbf{X}})$ where $\underline{\mathbf{X}} = (\mathrm{X}_1, ..., \mathrm{X}_\mathrm{M}) \in \mathbb{R}^M$ contains the 145 input parameters considered in this case study and where $\mathcal{M}$ is the PESHMELBA model.

2.4.1 Variance decomposition

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). 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 based on the decomposition of the total variance of the output:

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

where $\mathrm{V}_{i_1,...i_s}$ indicates the portion of variance that can be attributed to interactions between input parameters $X_i, i \in i_1, ...i_s$. 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, $0 \leq \mathrm{S}_{i_1,...,i_s} \leq 1$. In particular, first order sensitivity indices $\mathrm{S}_i = \frac{\mathrm{V}_i}{\mathrm{Var}[\mathrm{Y}]}$ 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:

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

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 direct computation requires a large sample size that can not be afforded in this case study. As a result, we compute Sobol' indices from a limited sample size, based on Polynomial Chaos Expansion (PCE, Sudret 2008) in order to ~~circumvect~~ circumvent 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, its polynomial chaos expansion is expressed as follows:

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

where the $\Psi_{\boldsymbol{\alpha}}$'s are multivariate orthonormal polynomials 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, 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 Sobol' indices). In our study, the UQLab Matlab software (Marelli and Sudret, 2014) is used to ~~computed~~ compute Sobol' indices from the ~~1,000~~2,000-point LHS (step 2.a, Figure 4). We use 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.

2.4.2 HSIC dependence measure

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. 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 that defines such scalar product. The HSIC measure corresponds to the square of the Hilbert-Schmidt norm of the cross-covariance operator $C[\mathcal{G}\mathcal{F}_i] : \mathcal{G} \to \mathcal{F}_i$, which is:

$$HSIC(\mathrm{X}_i, \mathrm{Y})_{\mathcal{F}_i, \mathcal{G}} = ||C[\mathcal{G}\mathcal{F}_i]||^2_{HS} = \sum_{j,k} \langle u^i_j, C[\mathcal{G}\mathcal{F}_i](v_k) \rangle_{\mathcal{F}_i} = \sum_{j,k} \mathrm{cov}(u^i_j(\mathrm{X}_i), v_k(\mathrm{Y})), \tag{5}$$

where $(u^i_j)_{j \geq 0}$ 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 $\mathrm{X}_i, i \in \{1, ..., M\}$ as:

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

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 $(\mathrm{X}_i, \mathrm{Y})$:

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

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 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 that can fully characterize the independence of variables 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.

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 $(x_i^1, ...x_i^N)$ and the associated output points ($y^1, ..., y^N$), 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 $(y_1, ..., y_N)$ 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 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

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 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 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.

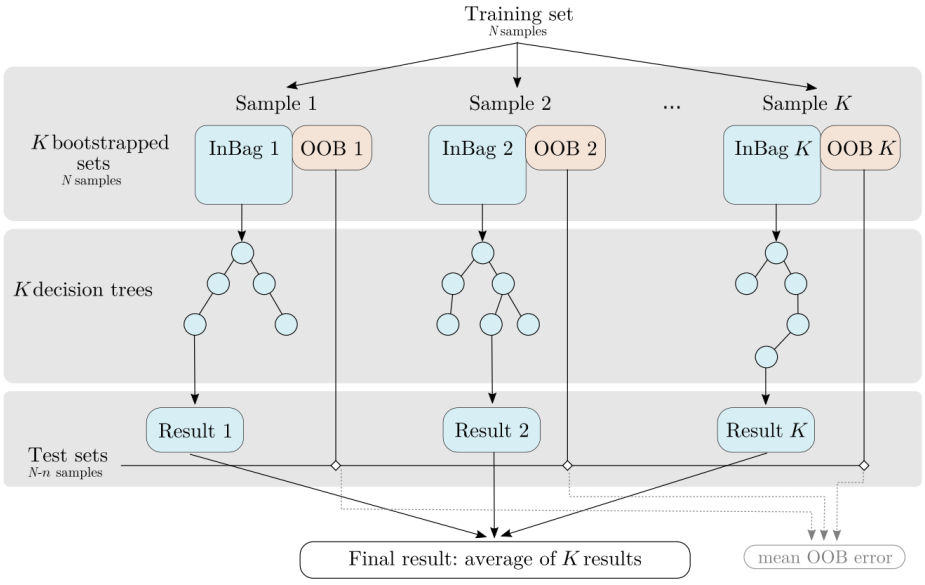

Figure 5. RF workflow (adapted from Rodriguez-Galiano et al. 2014). K bootstrapped sets are extracted from the original training set. Part of each set "InBag" is used to grow an independant decision tree and the final regression value is the average of all trees. The remaining portion of each tree "out-of-bag" (OOB) is used as a test set.

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

1. For each tree k:

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

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

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

   - For each input factor X_i:

     - Randomly permute x_i in $\{\underline{\mathbf{x}}^j \in \mathcal{L}_k\}$ to generate a new input set $\{\underline{\mathbf{x}}^{j*} \in \mathcal{L}_k\}$.
     - Estimate $\hat{\epsilon}^*_{OOB_k}(i)$ using the permuted input set:

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

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 establish a link between Mean Decrease in Accuracy and Sobol Total Indices when input parameters are assumed independent. Indeed, Gregorutti et al. (2017) establish that for all input parameter X_i:

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

In this study, the randomForestSRC R package (Ishwaran and Kogalur, 2020) is used to obtain feature importance measures ~~once again from the 1,000-point LHS. The~~ and the number of trees used to train the RF is set to 500.

2.5 Robustness and reliability assessment of the sensitivity indices

~~95% confidence intervals are calculated for all methods and~~ In order to assess the robustness of the calculated sensitivity indices, an additional 200-point test ~~is~~ set is first used to assess PCE and RF metamodel performances. ~~The~~ In a second time, 95% confidence intervals on sensitivity indices are computed for all methods, but by different means : the bootstrap resampling procedure provided in UQLab is used to calculate confidence intervals on PCE-based Sobol indices (see Marelli and Sudret 2018 for justification). ~~As RF building~~ In the case of random forest, building extra bootstraps could affect the consistency of the OOB sample as the RF structure already includes a bootstrap step to build each tree~~, applying another bootstrap could thus affect the consistency of the OOB sample. We then~~ . Then, we use a subsampling approach without replacement with

subsample size set to ~~800~~ 80% of the initial sample size to get error bounds on feature importance measures. ~~As deducing the RF structure from a reduced dataset may lead to limited performances, we first checked that RF performances were reasonably decreased when shifting the training dataset size from 1,000 to 800.~~ The same procedure is applied to estimate error bounds on HSIC indices~~as 800 points is supposed to be far enough to compute consistent estimators. 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~~. As typically found in similar studies (Archer et al., 1997; Yang, 2011), 1,000 bootstrap resamples are used for all methods.

In addition, reliability of each ranking method is assessed by monitoring its convergence properties. To do so, sensitivity indices and associated confidence intervals are calculated on samples of growing sizes. A new sample is generated for each size in $[50, 100, 250, 750, 1000, 2000]$ where the maximum size (2,000) points corresponds to the maximum computational budget available for ranking we guess for a catchment-scale application. To be noted that we create a new sample for each sample size to keep all samples as independent as possible but such approach may be too costly in a catchment scale application. In such case, another strategy could be to generate the sample for ranking from the initial LHS used for screening dropping the dimensions corresponding the non-influential inputs.

2.6 Aggregated sensitivity indices

After computing local sensitivity indices for each landscape unit on a scalar quantity, aggregated indices are computed at the catchment scale (step 2.b, Figure 4). Catchment-scale sensitivity indices are computed considering a multidimensional output $\underline{\mathbf{Y}} \in \mathbb{R}^d$ that gathers scalar outputs ~~on~~ for each landscape unit. Sobol' indices are aggregated at the catchment scale following the formulation by Gamboa et al. (2013) for generalized Sobol' indices :

$$S_u = \frac{\sum_{j=1}^{d} \mathrm{Var}[\mathrm{Y}_j] S_{u,j}}{\sum_{j=1}^{d} \mathrm{Var}[\mathrm{Y}_j]}, \tag{14}$$

where u is a subset of $\{1, ..., M\}$, $\mathrm{Var}[\mathrm{Y}_j]$ is the variance of the scalar j^{th} component of $\underline{\mathbf{Y}}$ and $S_{u,j}$ is the Sobol' indice of subset u on Y_j. Eq. (14) then 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 for scalar output remain valid for vectorial output. Catchment-scale indices ~~based on both HSIC and RF~~ can thus be directly computed on $\underline{\mathbf{Y}}$ when moving to multidimensional case.

3 Results

3.1 Screening

Screening is performed on the 4,000-point LHS. Simulations are run on the HIICS cluster (26 nodes, 692 cores, 64 to 256 GO of RAM per server) available at INRAE, France, for a simulation time of ~~4 days~~12,000 CPU-hours. Screening is performed at a

site scale, on each HU individually, to remain as conservative as possible. Influential parameters at the catchment scale are then deduced 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 **??**. ~~The number of input parameters retained after screening remains quite high proving that performing screening on PESHMELBA variables is a challenging task. 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 with its own set of characteristic parameters. Furthermore, spatial heterogeneities in terms of number of influential parameters are noticed depending on the studied output state variable, as shown in Figure **??**. 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.Screening on pesticide variables results on a higher number of influential parameters on HUs situated on the right bank, close to the outlet. Once again, such spatial heterogeneities in the influential parameters 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). One should also remain aware that all conclusions about influential parameters and heterogeneities are specific to the context of this case study (soils, climate,...) and cannot be generalized.~~

~~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~~

3.2 Ranking on a single HU

~~The~~

For the ranking task, the three methods (Sobol' indices from PCE, HSIC and RF) are applied on each HU based on ~~the 1,000~~a 2,000-point LHS generated for each scrutinized variable 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 6 shows the top parameters ranked from Sobol' total-order indices for all output variables on one HU close to the outlet,~~ 6,000 CPU-hours per variable. Columns 1 and 2 in Figure 6 show Sobol total and first-order indices while columns 3 and 4 show HSIC and RF sensitivity indices for the most influential parameters according to Sobol' total order indices for the four variables on HU14. ~~The identified influential parameters highly depend on the output variable considered~~

3.2.1 Physical interpretation

Considering Sobol' indices, influential parameters that are identified highly differ from one output variable to the other. They are linked to distinct physical processes that may interact with ~~each others and they bring~~ the other ones. This way, sensitivity

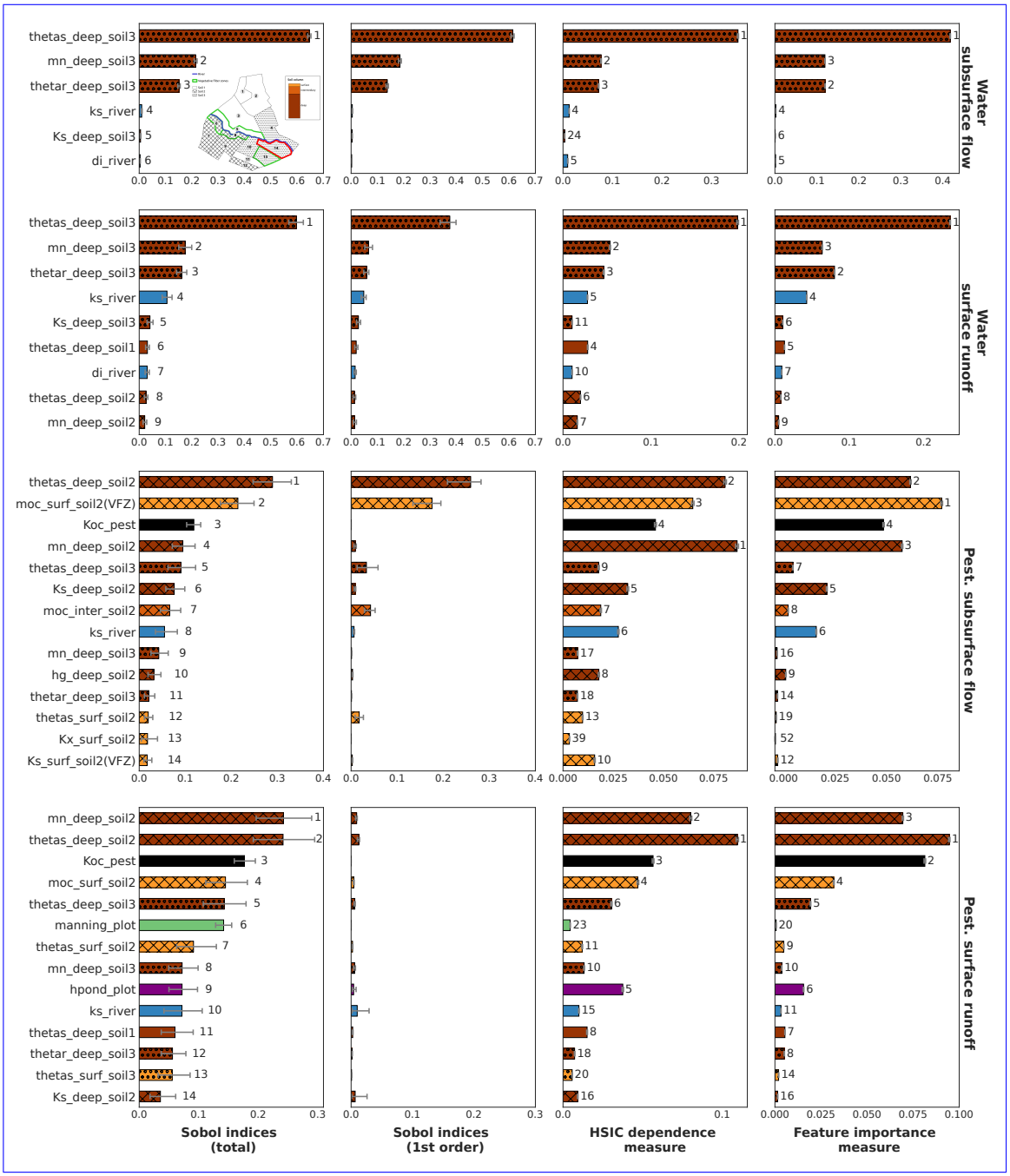

Figure 6. Sobol' total and first order indices computed using PCE, HSIC and RF site sensitivity indices for all output variables on HU14 with associated 95% confidence intervals. RF feature importance measures are normalized by 2Var[Y] following Eq. (13). HU14 is displayed with a red contour on top left figure. For all methods, displayed parameters are the 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. For each method, numbers next to bars stand for ranking position of parameters.

analysis brings knowledge about the way PESHMELBA represents the hydrological functioning of the virtual catchment. Water subsurface flow (top line) is driven by deep soil 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~~ Indeed, 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~~ deep soil parameters such as the saturated water content ($thetas$). Parameters linked to pesticide adsorption such as the pesticide adsorption coefficient and the organic carbon content (moc) ~~and saturated water content ($thetas$) are involved in the calculation of adsorption equilibrium~~of surface and intermediary soil horizons also rank among the top influential parameters. 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 6, 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 6 show~~

3.2.2 Comparison of methods

The rankings that are provided by the three methods are broadly consistent giving confidence in their robustness. Looking in more detail, HSIC and RF ~~sensitivity indices for Sobol' top-ten parameters. On the whole, the rankings obtained from Sobol ' indices are more similar to Sobol~~ total-order ~~, HSIC and RF sensitivity indices are consistent giving confidence in the robustness of these methods~~indices than to Sobol first-order indices contrarily to the conclusions of De Lozzo and Marrel (2016) on comparison of Sobol and HSIC indices. The most influential parameters ~~identified from Sobol'~~ that are identified by Sobol' total-order indices are also captured by the other methods. ~~Additional results (not shown here) also~~ Indeed, the different rankings show that the top-ten rankings based on Sobol' ~~total~~ total-order indices at least contain the five most influential parameters based on HSIC and RF rankings. Then, Figure 6 does not miss any preponderant parameters for HSIC and RF. ~~Rankings match perfectly well for water subsurface flow while there are~~ In addition, while rankings from the three methods only exhibit slight differences for water ~~surface runoff. Differences~~ variables, differences between rankings are more pro-

nounced for pesticide variables. In that case, most parameters have zero or very low first order Sobol' indices characterizing nearly purely interacting effects.

~~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 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.~~

~~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 indices but our findings reach the opposite conclusion. 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, not all differences between rankings come from interactive effects. As an example, *manning_plot* and *hpond_plot* that appear in pesticide surface runoff top-ten ranking according to Sobol' indices are both characterized by nearly fully interactive effects. However, *hpond_plot* also ranks influential according to HSIC and RF ~~feature importance indices have been proven to relate to Soboltotal indices (Gregorutti et al., 2017) but this relation (see Eq. 13) is not respected in this study. The quality scores for PCE and RF metamodels 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 4). 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 boundson Sobol' indices estimates. We~~ while *manning_plot* is missed by both of them.

Regarding error bounds, they are very small for HSIC and RF contrarily to Sobol' indices estimates. HSIC is expected to be very accurate from very small sample sizes (Da Veiga et al., 2021) which may explain such differences of magnitude for error of Sobol and HSIC indices. Regarding RF error bounds, we incriminate the resampling strategy that differ ~~between the two methods. While~~ from Sobol method. Indeed, 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 Ishwaran and Lu (2019) should probably be further investigated

on bigger samples to better compute RF sensitivity indices and to accurately assess their quality. In addition to Figure 6, Table 4 gathers the quality scores Q^2 for PCE and RF metamodels on the test set. Results show that the RF metamodel performs much more poorly than the PCE for all variables. Such poor performances may explain the discrepancies between RF feature importance measures and Sobol' total indices. Indeed, RF feature importance indices have been proven to relate to Sobol total indices (Gregorutti et al., 2017) but this relation (see Eq. 13) is not respected in this study. These results thus underline how crucial it is to analyse both the quality of the metamodel and the quality of the indices it calculates and not just one of the two.

~~To conclude, although Sobol' indices cannot capture influential parameters if they do not predominantly affect the output variance, the li~~

Table 4. Q^2 score for all variables on HU14 calculated from the 200-point test set. The Q^2 score is calculated as $Q^2 = 1 - \frac{\sum_{i=1}^{N}(\mathcal{M}(\underline{\mathbf{X}}^i) - Y^i)^2}{\sum_{i=1}^{N}(Y^i - \overline{Y})^2}$, where $\overline{Y} = \frac{1}{N}\sum_{i=1} Y^i$ is the empirical mean of the sample.

3.2.3 Convergence rate

In order to assess each method convergence rate, sensitivity indices are calculated for growing sample sizes, from 50 to 2000 points. Results for water surface runoff and pesticide surface runoff are presented respectively on Figure 7 and Figure 8. Results for water (resp. pesticide) subsurface flow are not presented as they come to the same conclusions than for water (resp. pesticide) surface runoff.

Figure 7 shows that Sobol indices are close from stability for sample sizes from 1,000 points as ranking and Sobol total and first-order indices are rather stable. 95% confidence intervals associated to Sobol indices (first and total-order) also seem to be rather stable for sample sizes from 1,000 points. Considering HSIC and RF indices, ranking differences for the five most sensitive input factors are less than one position for sample sizes from 1,000 points, which roughly suits to the convergence criteria for parameter ranking defined in Sarrazin et al. (2016). Identically, for both methods sensitivity indices values and 95% confidence intervals show few variability. As already noticed on Figure 6, the 95% confidence intervals for HSIC and RF are much smaller than for PCE. Again, such results should be ~~thus~~ interpreted with caution ~~. Considering a high dimension problem and a very limited computational budget (inferior to~~ as confidence intervals are not calculated in the same way for the different methods. On the whole, the HSIC method reaches convergence first as ranking and sensitivity indice values stabilize for sample sizes from 750 while 1,000 ~~simulations) , HSIC indices~~ points are necessary for Sobol and RF methods.

Results about pesticide surface runoff (Figure 8) are more contrasted. Ranking and sensitivity indice values from PCE show more variability than for water variable. This is especially the case for first order Sobol' indices which shows that 2,000 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~~

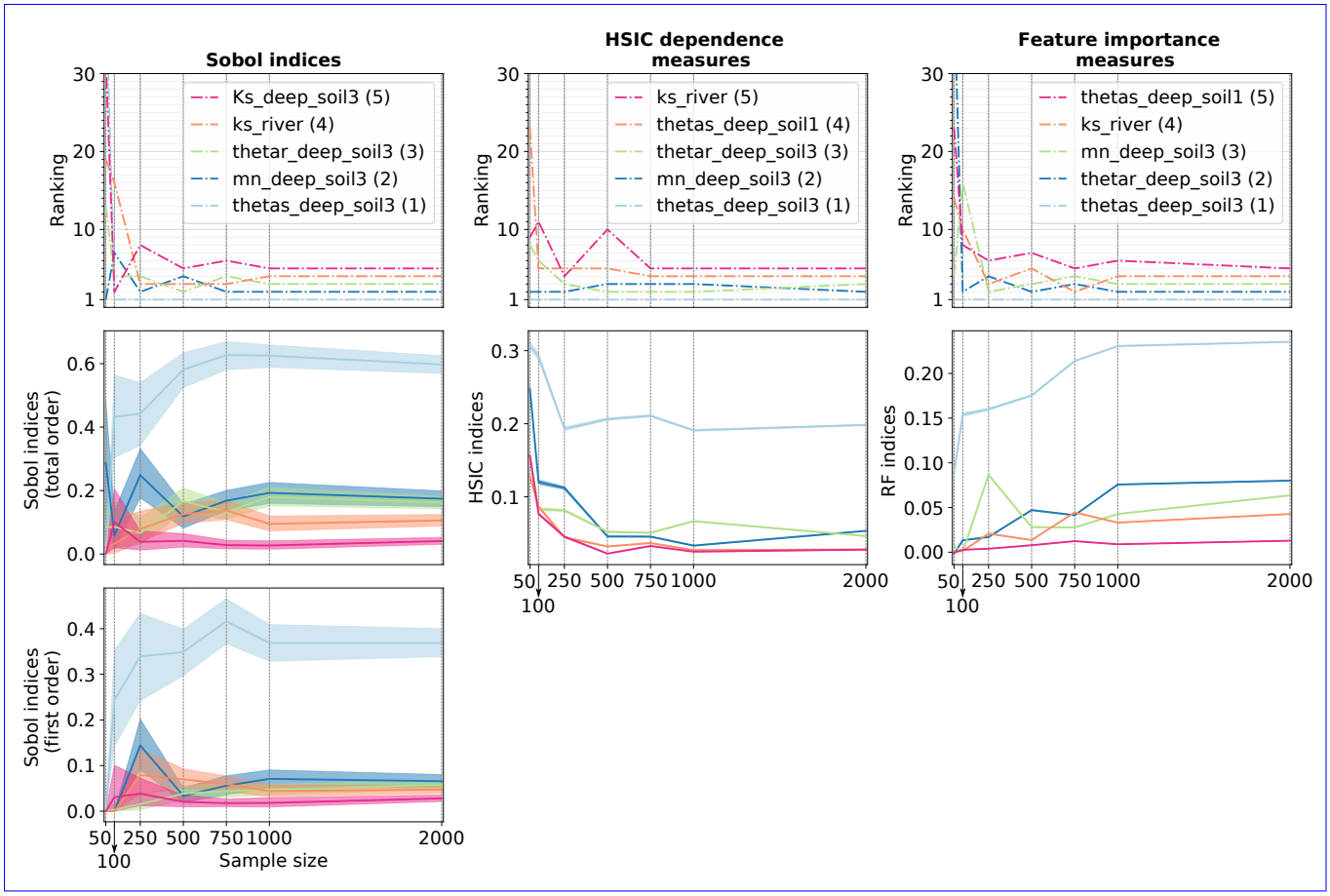

Figure 7. Convergence plot for solute water runoff variable: the figure shows the ranking for the 5 most influential parameters identified with a sample of 2,000 points so as the corresponding sensitivity indices values with associated 95% confidence intervals for growing sample sizes (even if they are hardly visible for HSIC and RF, confidence intervals are displayed on all plots from line 2). In the legend, the number in brackets is the ranking position of the parameter in the case of a sample of 2,000 points.

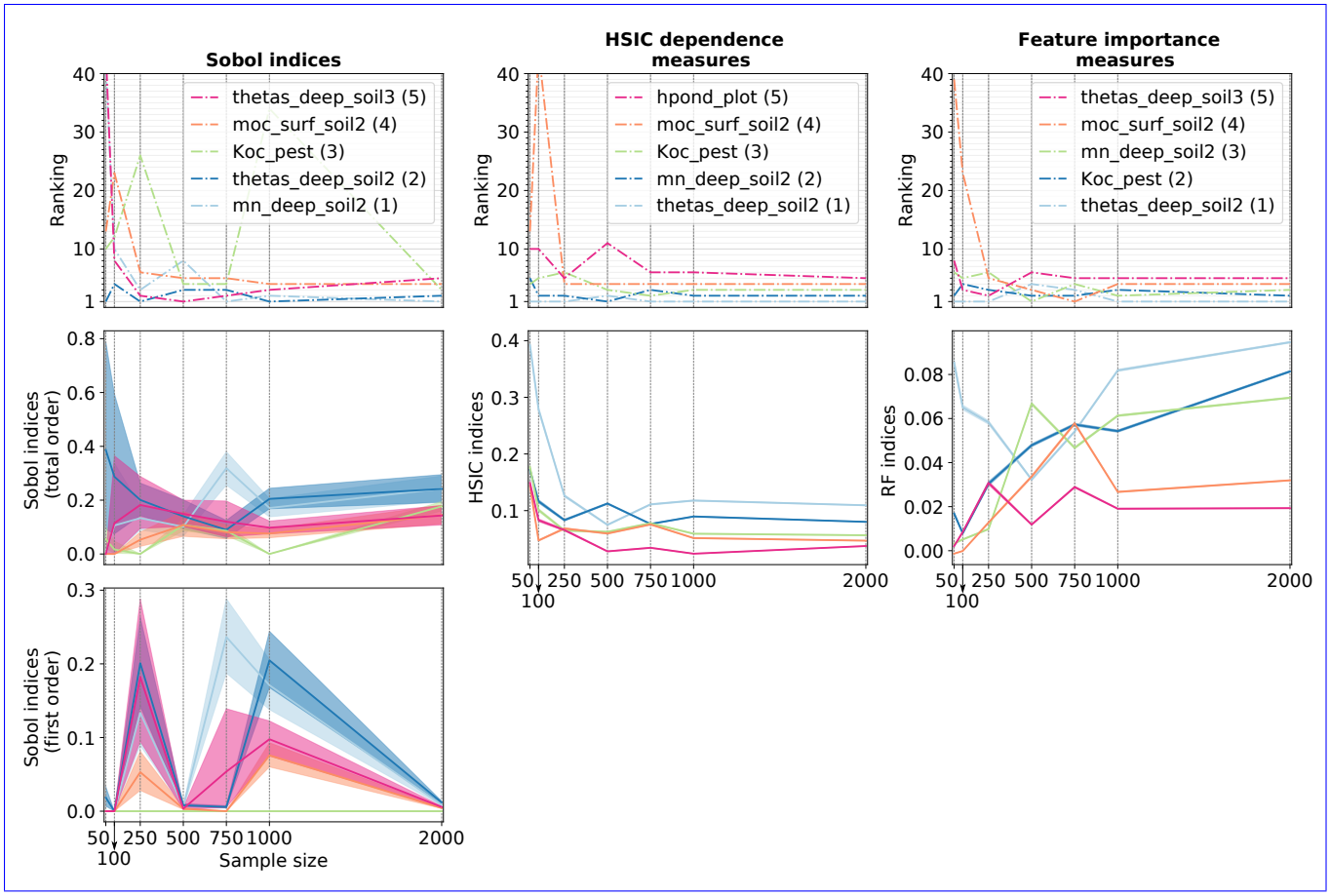

Figure 8. Convergence plot for solute surface runoff variable: the figure shows the ranking for the 5 most influential parameters identified with a sample of 2,000 points so as the corresponding sensitivity indices values with associated 95% confidence intervals for growing sample sizes (even if they are hardly visible for HSIC and RF, confidence intervals are displayed on all plots from line 2). In the legend, the number in brackets is the ranking position of the parameter in the case of a sample of 2,000 points.

~~to Sobol' indices computed from PCE~~a sufficient sample size to get robust results. Ranking and total order indices are more stable except for the parameter *Koc_pest*. This parameter is characterized by fully interactive effects which may explain why calculating robust Sobol' indices is a more complex task. Ranking and sensitivity measures from HSIC stabilize from 750 points while the error bounds remain very limited for all sample sizes. Finally, RF ranking and indices are almost stable from 1,000 points except for *Koc_pest* whose RF value still shows variability for the biggest sample sizes.

3.3 Landscape analysis

~~The previous sectionshowed that~~ In this section, we focus on Sobol' indices (first and total order) ~~provide valuable information about interactions between parameters. As a result, we focus on this method to compute the indices at the catchment scale.~~ despite larger error bounds as it is the only method used in this study that allows to get separate information on interactive effects.

Site rankings such as presented in Figure 6 are gathered for all HUs in the form of sensitivity maps in Figure 9 for water surface runoff, and in Figure 10 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, 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~~ The, local 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 9 and 10) 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 right bank (bottom part of the catchment in the Figure). ~~Finally~~In addition, 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.

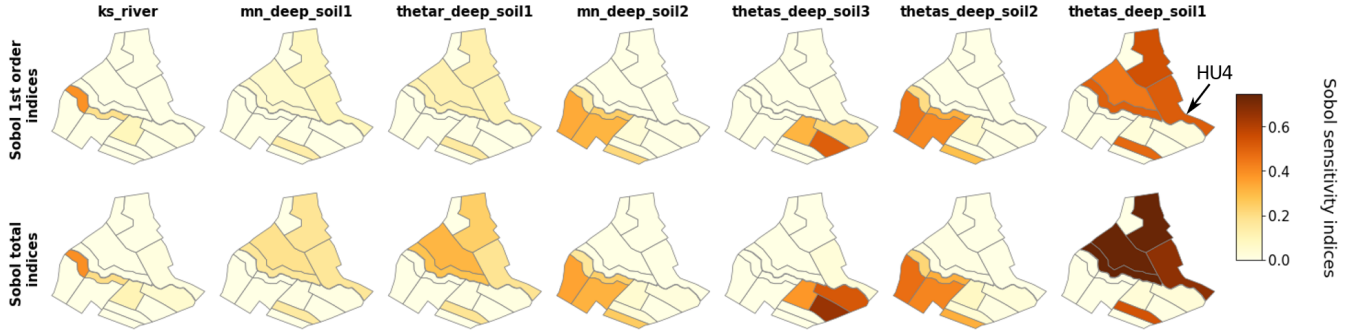

Figure 9. Maps of Sobol site sensitivity indices for water surface runoff for the most significantly influential parameters.

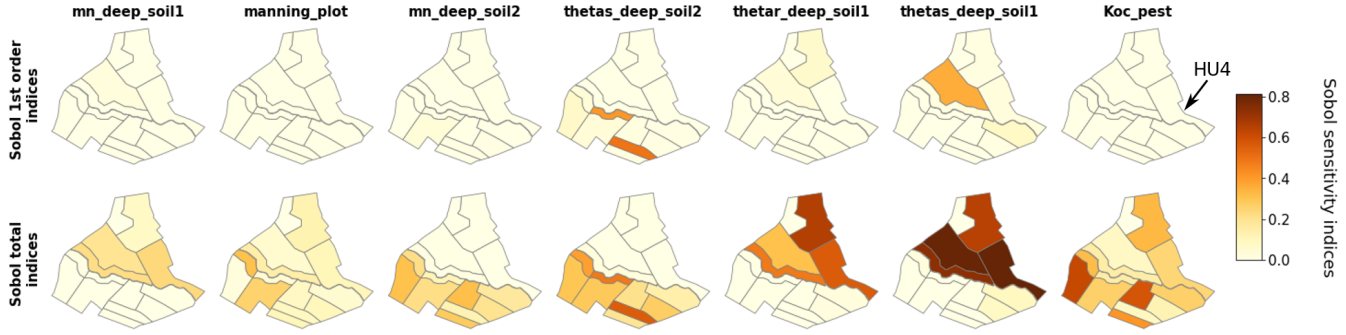

Figure 10. Maps of Sobol site sensitivity indices for pesticide surface runoff for the most significantly influential parameters.

Finally, Figure 11 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 processes dominating in PESHMELBA to explain the output variable. For water surface runoff, 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-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.

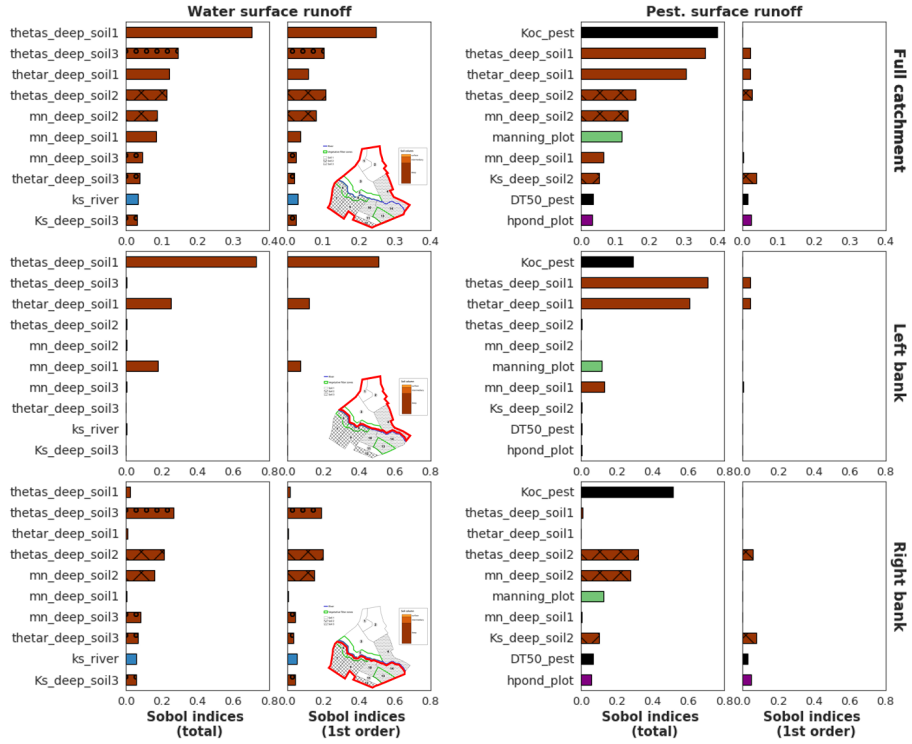

Figure 11. Sobol' first-order and total-order aggregated sensitivity indices for water surface runoff (left) and pesticide surface runoff (right) calculated at the scale of the catchment (top), left bank (middle) and right bank (bottom). Displayed parameters are the 11 most influential parameters regarding 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.

~~Sensitivity maps provide local, detailed information about influential parameters on each location of the catchment. Howeverthey 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 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, a 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 hydrodynamic parameters and pesticide adsorption coefficient to improve the quality of the simulation of both water and pesticide surface flows. As pointed out in Marrel et al. (2015), both approaches are complementary and provide precious~~

4 Discussion

4.1 On screening

For all variables considered, the number of input parameters retained after screening remained quite high proving that performing screening on PESHMELBA variables is a challenging task. We can first incriminate the many physical processes interacting in PESHMELBA in a spatially-distributed way, each of them with its own set of characteristic parameters. However, it can also be explained by the methodology that may not be discriminating enough. Many previous studies developed efficient screening techniques for complex environmental models (e.g. Tang et al., 2007; Nossent et al., 2011; Touzani and Busby, 2014; Becker et al., 2018; C
. Exploring other screening approaches is beyond the scope of this study but future work may focus on applying and critically comparing those techniques, especially with very limited sample size (inferior to 2,000 points for example).

4.2 On the choice of a ranking method

By exploring several methods for ranking, the aim was to analyze their specificities and the interest of each one for the sensitivity analysis of a complex environmental model, characterized by many parameters and a high computational cost such as PESHMELBA. Considering the results of this study, we believe that the choice of the method depends on the properties of the model, the objective of the sensitivity analysis and the sample size available:

1. Sobol method remains attractive when sensitivity analysis is used to gain knowledge about the model ~~functioning. The scale of the bank, or more generally of the hillslope, may also constitute an adapted intermediary scale to meet both requirements of detailed results for physical interpretation and computational efficiency.~~

5 ~~Discussion on the methodology~~

by finely analyzing its behavior. Indeed, Sobol' indices provide a clear interpretation of the calculated indices (percentages of variance explained) and explicit information about the interactions between parameters. These elements are particularly valuable when one wishes to use sensitivity analysis to understand the functioning of the model and this is why this approach is still widely used in the hydrological community. In the case of variables that are reasonably complex and that are not characterized by too much interactions of physical processes (such as water variables in our case), using chaos polynomials to estimate Sobol' indices is particularly interesting and efficient since it allows the use of a pre-existing sample, of very limited size. Conclusions are much contrasted for complex variables such as pesticide variables as convergence results showed that 2,000 points may not be enough to get fully robust Sobol' indices. This is particularly the case for parameters that are mainly characterized by interactive effects.

~~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 casestudy, 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~~

2. If the sensitivity analysis aims at simplifying the model or focusing the calibration efforts, if the physical interpretation of the results is not a priority and if one has a preexisting sample of very limited size (inferior to 750 points 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, 2016). 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 HHCS. The p-values were then calculated from 100 bootstraps leading to an additional calculation time of 2 hours on the HHCS 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; Gate and eFAST method has also been increasingly used (Lauvernet and Muñoz-Carpena, 2018; Fox et al., 2010; Metta, 2007; Vanuytrech~~), the use of HSIC indices is a good option as it provides robust sensitivity indices. However, it is important to note that using HSIC dependence measures for sensitivity analysis is a recent idea and that there is still little knowledge available about identifying and differentiating the types of dependency that are captured. In addition, the choice of the kernel may affect the ranking results because each specific kernel is likely to give more or less importance to the infinite number of dependency forms that are captured by HSIC. The question of the choice of the kernel is delicate and it is still not very much addressed in the literature. While a few papers propose to choose the type and parameterization of the kernels in a way that maximizes the possible dependence between X_i and Y (Fukumizu et al., 2009; Balasubramanian et al., 2013), the interpretation of the results seems to be less clear. On the other hand, there are still relatively few works that apply this method for GSA and the limitations are not necessarily all identified yet. Using HSIC for a classification exercise will therefore remain delicate as long as there is no consensus on the choice of the kernel and the interpretation of the results. 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~~ these problems do not arise when using HSIC for screening and such method is therefore perhaps to be preferred for this type of exercise.

3. The RF indices is also of interest in sensitivity analysis task as it is supposed to provide an estimator of total Sobol indices. Those indices can thus be easily interpreted and as for the other methods, they can be estimated from a pre-existing 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)~~However, PCE is still to be preferred since it provides more complete information including not only the total Sobol indices but also the indices at all orders. 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~~ our results showed that the metamodel constructed by RF is of lower quality than the one constructed by PCE at equivalent sample size giving less confidence in the ~~results. Combining~~ resulting sensitivity indices.

Beyond the comparison of 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)~~, we also tried to evaluate if it was possible and useful to combine several of these methods. However, considering the results obtained, we believe that combining the tested methods is still of little interest for hydrologists to better understand the model functioning. Indeed, the differences we found in rankings remain difficult to interpret. This is particularly the case when combining Sobol and HSIC indices, due to the fact that the results from HSIC dependence measures remain fuzzy to interpret.

4.1 On choosing local or aggregated sensitivity analysis

Results about landscape analysis showed that, on the one hand, 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. On the other hand, catchment scale aggregated indices provide a synthetic information at a lower computational cost but the spatialized aspect of the GSA is lost. As pointed out in Marrel et al. (2015), both approaches are complementary and provide precious knowledge about the model functioning but they cannot always be performed together. As a compromise, we propose to use the scale of the bank, or more generally of the hillslope, as it may constitute an adapted intermediary scale to meet both requirements of detailed results for physical interpretation and computational efficiency. Indeed, in this study Sobol' aggregated indices were directly computed from site sensitivity indices as they were available but a pick-freeze estimator (Gamboa et al., 2013; De Lozzo and Marrel, 2016) can be used for a direct computation of Sobol' generalized indices. In our case, such overview of sensitivity analysis allows us to focus calibration efforts on deep soil hydrodynamic parameters and pesticide adsorption coefficient to improve the quality of the simulation of both water and pesticide surface flows.

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 existing samples due to the high computational cost of a simulation and spatialized ouputs. We first ~~used~~ performed a screening step using an independence test based on the HSIC dependence measure. Then, we ~~combined several~~ compared several innovative 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. 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~~ Results showed that Sobol' indices computed from polynomial chaos expansion are particularly attracted as they provide easy-to-interpret sensitivity indices but also that very limited sample sizes may not be sufficient to reach convergence in the case of complex variables with dominant interactive effects. In a second time, we gathered such local sensitivity indices into sensitivity maps ~~and they highlight~~ that highlighted local contributions of parameters. Although very informative about the 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 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~~ This study constitutes the first attempt of 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 (Alves Ferreira et al., 1995; Lauvernet and Muñoz-Carpena, 2018; Saltelli et al., 2019)~~the PESHMELBA model. Future research should go a step further by considering the other sources of uncertainties that can affect the model and interact with parameter uncertainties. The impact of forcings, soil types, quantities and dates of application of pesticides should be addressed as already done in Holvoet et al. (2005) for instance. Additionally, parameters were assumed to be independent in this study but this assumption may be ~~too simplistic~~arguable, especially for hydrodynamic parameters. ~~Further research should adapt the tested methodsto dependent parameters~~For the three methods, sensitivity indices that are produced are meaningless in the case of dependent inputs. Dealing with dependent parameters has already been explored in the case of Sobol' indices ~~(Chastaing et al., 2015) but it needs to be further explored in the case of HSIC and RF-based methods. It~~ and HSIC measures, for example based on Shapley effects (Da Veiga et al., 2021) but again, these formulations should be extensively tested with complex variables and very small sample sizes. Finally, 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 (Roux et al., 2021) or using the principal components of the model's functional outputs. The definition and the use of adequate hydrological signatures such as proposed in Branger and McMillan (2020) and Horner (2020) 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~~ brings additional knowledge on GSA strategies for modellers who deal with such complex models and thus paves the way for systematic analysis of ~~such~~ environmental exposure models.

Code and data availability. The PESHMELBA model is an open-source model coded in Python (Version 2.7.17) and Fortran 90 and embedded in the OpenPALM coupler (Version 4.3.0). The code for the OpenPALM coupler is available from www.cerfacs.fr/globc/PALM_WEB/user.html#download after registration. The exact version of PESHMELBA used to produce the simulations is archived on Zenodo (https://zenodo.org/record/6319769#.YinMV1TjKUk) as are input data and scripts used to produce all the sensitivity indices presented in this paper (https://zenodo.org/record/6319773#.YinMc1TjKUk). These scripts use the open-source Matlab software for uncertainty quantification UQLab Version 2.0.0 (www.uqlab.com). The R package randomForestSRC is freely available for download at https://cran.r-project.org.

Author contributions. All authors contributed to writing the text and to all stages of editing. PCE computation was performed by Bruno Sudret and Emilie Rouzies whereas HSIC and RF indices computation was led by Emilie Rouzies with extensive support from Claire Lauvernet and Arthur Vidard.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors kindly acknowledge Jérémy Verrier for his support on HIICS cluster usage, Stefano Marelli for his support on UQlab and PCE usage for GSA and Bertrand Ioss for his advice on the computation of RF feature importance measures.

References

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.

Alipour, A., Jafarzadegan, K., and Moradkhani, H.: Global sensitivity analysis in hydrodynamic modeling and flood inundation mapping, 152, 105 398, https://doi.org/10.1016/j.envsoft.2022.105398, 2022.

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.

Alves Ferreira, V., Weesies, G., Yoder, D., Foster, G., and Renard, K.: The site and condition specific nature of sensitivity analysis, Journal of Soil and Water Conservation, 50, 493–497, 1995.

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. 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, https://doi.org/10.1016/j.petrol.2019.106237, 2019.

Becker, W. E., Tarantola, S., and Deman, G.: Sensitivity analysis approaches to high-dimensional screening problems at low sample size, Journal of Statistical Computation and Simulation, 88, 2089–2110, https://doi.org/10.1080/00949655.2018.1450876, 2018.

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.

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. and McMillan, H. K.: Deriving hydrological signatures from soil moisture data, Hydrological Processes, 34, 1410–1427, https://doi.org/10.1002/hyp.13645, 2020.

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.

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, 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.

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, 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.

Chastaing, G., Gamboa, F., and Prieur, C.: Generalized Sobol sensitivity indices for dependent variables: numerical methods, Journal of Statistical Computation and Simulation, 85, 1306–1333, https://doi.org/10.1080/00949655.2014.960415, 2015.

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.

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.

Da Veiga, S., Gamboa, F., Iooss, B., and Prieur, C.: Basics and Trends in Sensitivity Analysis, https://doi.org/10.1137/1.9781611976694, 2021.

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.

D'Andrea, M. F., Letourneau, G., Rousseau, A. N., and Brodeur, J. C.: Sensitivity analysis of the Pesticide in Water Calculator model for applications in the Pampa region of Argentina, 698, 134 232, https://doi.org/10.1016/j.scitotenv.2019.134232, 2020.

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

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 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.

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, 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.

Faúndez Urbina, C. A., van den Berg, F., van Dam, J. C., Tang, D. W. S., and Ritsema, C. J.: Parameter sensitivity of SWAP–PEARL models for pesticide leaching in macroporous soils, 19, e20 075, https://doi.org/10.1002/vzj2.20075, 2020.

FOCUS: FOCUS surface water scenarios in the EU evaluation process under 91/414/EEC, European commission, report of the FOCUS 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.

Hamby, D. M.: A review of techniques for parameter sensitivity analysis of environmental models, 32, 135–154, https://doi.org/10.1007/BF00547132, 1994.

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.

Holvoet, K., van Griensven, A., Seuntjens, P., and Vanrolleghem, P. A.: Sensitivity analysis for hydrology and pesticide supply towards the river in SWAT, 30, 518 – 526, https://doi.org/10.1016/j.pce.2005.07.006, assessment of Anthropogenic Impacts on Water Quality, 2005.

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.

Horner, I.: Design and evaluation of hydrological signatures for the diagnostic and improvement of a process-based distributed hydrological model, Ph.D. thesis, Université Grenoble Alpes, http://www.theses.fr/2020GRALU014, thèse de doctorat dirigée par Branger, Flora Océan, Atmosphère, Hydrologie Université Grenoble Alpes 2020, 2020.

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

Ishwaran, H. and Lu, M.: Standard errors and confidence intervals for variable importance in random forest regression, classification, and survival, Statistics in Medicine, 38, 558–582, https://doi.org/10.1002/sim.7803, 2019.

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.

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.

Marrel, A., Marie, N., and De Lozzo, M.: Advanced surrogate model and sensitivity analysis methods for sodium fast reactor accident assessment, Reliability Engineering & System Safety, 138, 232 – 241, https://doi.org/10.1016/j.ress.2015.01.019, 2015.

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.

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: LANDMOD2010, 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.

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.

Nossent, J., Elsen, P., and Bauwens, W.: Sobol' sensitivity analysis of a complex environmental model, 26, 1515–1525, https://doi.org/10.1016/j.envsoft.2011.08.010, 2011.

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 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 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 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.

Roux, S., Buis, S., Lafolie, F., and Lamboni, M.: Cluster-based GSA: Global sensitivity analysis of models with temporal or spatial outputs using clustering, Environmental Modelling and Software, 140, 105 046, https://doi.org/10.1016/j.envsoft.2021.105046, 2021.

Rouzies, E., Lauvernet, C., Barachet, C., Morel, T., Branger, F., Braud, I., and Carluer, N.: From agricultural catchment to management 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 risk management policies, Environmental Modelling & Software, 60, 153–166, https://doi.org/10.1016/j.envsoft.2014.06.012, 2014.

Saltelli, A.: Sensitivity analysis for importance assessment, 22, 579–590, https://doi.org/10.1111/0272-4332.00040, 2002.

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.

Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S., and Wu, Q.: Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices, Environmental Modelling and Software, 114, 29–39, https://doi.org/https://doi.org/10.1016/j.envsoft.2019.01.012%7D, 2019.

Saltelli, A., Jakeman, A., Razavi, S., and Wu, Q.: Sensitivity analysis: A discipline coming of age, Environmental Modelling and Software, 146, 105 226, https://doi.org/10.1016/j.envsoft.2021.105226, 2021.

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.

Sheikholeslami, R., Razavi, S., Gupta, H. V., Becker, W., and Haghnegahdar, A.: Global sensitivity analysis for high-dimensional problems: How to objectively group factors and measure robustness and convergence while reducing computational cost, Environmental Modelling and Software, 111, 282–299, https://doi.org/10.1016/j.envsoft.2018.09.002, 2019.

Sheikholeslami, R., Gharari, S., Papalexiou, S. M., and Clark, M. P.: VISCOUS: A Variance-Based Sensitivity Analysis Using Copulas for Efficient Identification of Dominant Hydrological Processes, Water Resources Research, 57, e2020WR028 435, https://doi.org/10.1029/2020WR028435, 2021.

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.

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.

Tang, Y., Reed, P., Wagener, T., and van Werkhoven, K.: Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, 11, 793–817, https://doi.org/10.5194/hess-11-793-2007, 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.

Tissot, J.-Y. and Prieur, C.: A randomized orthogonal array-based procedure for the estimation of first- and second-order Sobol' indices, Journal of Statistical Computation and Simulation, 85, 1358–1381, https://doi.org/10.1080/00949655.2014.971799, 2015.

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.

Touzani, S. and Busby, D.: Screening Method Using the Derivative-based Global Sensitivity Indices with Application to Reservoir Simulator, Oil and Gas Science and Technology, 69, 619 – 632, https://doi.org/10.2516/ogst/2013195, 2014.

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.

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.

Yang, J.: Convergence and uncertainty analyses in Monte-Carlo based sensitivity analysis, Environmental Modelling & Software, 26, 444 – 457, https://doi.org/https://doi.org/10.1016/j.envsoft.2010.10.007, 2011.

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