



A simple and efficient method for global sensitivity analysis based on cumulative distribution functions



Francesca Pianosi*, Thorsten Wagener

Department of Civil Engineering, University of Bristol, University Walk, BS81TR, Bristol, UK

ARTICLE INFO

Article history:

Received 29 July 2014

Received in revised form

30 December 2014

Accepted 8 January 2015

Available online 27 January 2015

Keywords:

Global sensitivity analysis

Variance-based sensitivity indices

Density-based sensitivity indices

Uncertainty analysis

ABSTRACT

Variance-based approaches are widely used for Global Sensitivity Analysis (GSA) of environmental models. However, methods that consider the entire Probability Density Function (PDF) of the model output, rather than its variance only, are preferable in cases where variance is not an adequate proxy of uncertainty, e.g. when the output distribution is highly-skewed or when it is multi-modal. Still, the adoption of density-based methods has been limited so far, possibly because they are relatively more difficult to implement. Here we present a novel GSA method, called PAWN, to efficiently compute density-based sensitivity indices. The key idea is to characterise output distributions by their Cumulative Distribution Functions (CDF), which are easier to derive than PDFs. We discuss and demonstrate the advantages of PAWN through applications to numerical and environmental modelling examples. We expect PAWN to increase the application of density-based approaches and to be a complementary approach to variance-based GSA.

© 2015 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Global Sensitivity Analysis (GSA) is a set of mathematical techniques aimed at assessing the propagation of uncertainty through a numerical model, and specifically at understanding the relative contributions of the different sources of uncertainty to the variability in the model output. Quantitative GSA uses sensitivity indices, which summarise such relative influence into a scalar measure. Sources of uncertainty may include the model parameters, errors in forcing data, or even non-numerical uncertainties like the resolution of a spatially-distributed simulation model grid.

A well-established and widely used method for GSA is the variance-based approach. Here, the output sensitivity to an uncertain input is measured by the contribution to the output variance coming from the uncertainty of that input. Variance-based sensitivity indices have become increasingly popular in GSA applications across different environmental modelling domains (see for example older and more recent applications in Pastres et al. (1999); Pappenberger et al. (2008); van Werkhoven et al. (2008); Nossent et al. (2011); Ziliani et al. (2013); Baroni and Tarantola (2014), and Saltelli et al. (2008) for a general discussion). The main reason for their diffusion is that they possess several desirable

properties, and in particular: they are applicable independently of the characteristics of the input–output response function (e.g. linear or non-linear); they can be used for both input ranking (so called “factor prioritisation”) and screening; and they are easy to implement and to interpret.

By “easy-to-implement” we mean that the computation of the two main variance-based indices (the so called “main effects” and “total effects” indices) from a given output sample is relatively straightforward. This is because several estimators are available to approximate them via closed-form algebraic equations, provided that the output sample has been generated using a tailored sampling strategy. A review of these approximators and relevant sampling strategy is given in Saltelli et al. (2010). The generation of the output sample is thus by far the most computationally demanding step in the calculation of variance-based indices, and a major limitation to their applicability, especially because obtaining accurate estimates may require a large number of output samples. Specifically, when using the efficient estimators reviewed in Saltelli et al. (2010), the total number of model evaluations grows linearly with the number of uncertain inputs according to the formula $N = n \times (M + 2)$ where N is the number of model evaluations, M is the number of inputs M and n is the proportionality factor selected by the user. Depending on the application, the proportionality factor n may vary between 10^2 and 10^4 . For example, Baroni and Tarantola (2014) found that convergence was reached using

* Corresponding author.

E-mail address: francesca.pianosi@bristol.ac.uk (F. Pianosi).

$N = 7168$ model evaluations for $M = 5$ uncertain inputs ($n = 1024$) while in Nossent et al. (2011) convergence was reached only after $N = 336,000$ evaluations for $M = 26$ inputs ($n = 12,000$).

Another major limitation of variance-based sensitivity indices is that they implicitly assume that output variance is a sensible measure of the output uncertainty, which might not always be the case. For instance, if the output distribution is multi-modal or if it is highly skewed, using variance as a proxy of uncertainty may lead to contradictory results. Borgonovo et al. (2011) provide an example of a highly-skewed distribution where the unconditional output variance is lower than the conditional variance at specific conditioning values for the inputs. In such a case, representing uncertainty by variance would lead to the contradictory result that uncertainty about the output increases when removing uncertainty about one of the inputs. Liu et al. (2006) provide another example where variance-based sensitivity indices fail to properly rank the inputs of a model whose output has a highly skewed distribution.

These limitations have stimulated a number of studies on “moment-independent” sensitivity indices, that is, indices that do not use a specific moment of the output distribution to characterise uncertainty and therefore are applicable independently of the shape of the distribution. These methods are sometimes referred to as “density-based” methods because they investigate the Probability Density Function (PDF) of the model output, rather than its variance only. Here, sensitivity is related to the variations in the output PDF that are induced when removing the uncertainty about one input. Entropy-based sensitivity measures (Park and Ahn, 1994; Krykacz-Hausmann, 2001; Liu et al., 2006) and the δ -sensitivity measure (Borgonovo, 2007; Plischke et al., 2013) follow this line of reasoning.

However, the adoption of density-based methods in the environmental modelling domain has been limited so far. To the authors' knowledge, the few examples are Pappenberger et al. (2008) for the entropy-based indices; and Castaings et al. (2012), Anderson et al. (2014) and Peeters et al. (2014) for the δ -sensitivity measure.

One possible reason for this limited diffusion is that density-based indices are relatively less easy to implement than variance-based ones, mainly because their computation requires the knowledge of many conditional PDFs. As PDFs are generally unknown, empirical PDFs must be used. The simplest approach to derive an empirical PDF is to use a histogram of the data sample, however the resulting shape can be significantly affected by the position of the first bin and the bin width, whose appropriate values may be difficult to determine. PDFs are better approximated using kernel density estimation (KDE) methods, since they only require to specify a single parameter, the bandwidth. A more complex approach is to first approximate the CDF and then derive the PDF as its derivative (see discussion and references in Liu et al. (2006)). However, the approximation procedure cannot be overly complex as the computation of density-based sensitivity indices may require derivation of a large number of empirical PDFs. At a minimum, one conditional PDF per uncertain input is needed, and much more if one wants to consider multiple conditioning values for each input. Furthermore, the number of PDFs to be estimated quickly becomes excessive if one wants to analyse the accuracy or convergence of the sensitivity indices, as this requires the repeated computation of the indices using different bootstrap resamples of the original dataset, or subsamples of different size.

In this paper we present a novel method, called PAWN (derived from the authors names), to efficiently derive a density-based sensitivity index. The key idea of PAWN is to characterise the output distribution by its Cumulative Distribution Function (CDF) rather than its PDF. The advantage is that the approximation of empirical CDFs from a data sample comes at no computing costs and does not require any tuning parameter. This makes PAWN very

easy to implement and facilitates the application of bootstrapping and convergence analysis. We also show how intermediate results generated in the PAWN implementation procedure can be effectively visualised to gather further insights regarding the output behaviour and map it back into the input space (so called Factor Mapping in the GSA literature (Saltelli et al., 2008)). Another advantage of PAWN is that sensitivity indices can be easily computed either considering the entire range of variation of the output or just a sub-range, which can be very useful in applications where one is mainly interested in a specific region of the output distribution, for instance the tail. Thanks to its relative simplicity and the above advantages, we expect PAWN to simplify the application of density-based approaches and to be a valuable complementary approach to variance-based GSA.

The paper is structured as follows. In the next Section, we introduce the main concepts that are used throughout the paper and list good properties that a sensitivity index should satisfy. In Section 3, we present our PAWN method, discuss its properties, and compare it to other density-based approaches that can be found in the literature. We also discuss the radical difference between PAWN and the widely used approach of Regional Sensitivity Analysis (Spear and Hornberger, 1980). In Section 4 we demonstrate PAWN by application to a set of numerical and environmental modelling examples. Current limitations of our work and directions for further research are given in the concluding section.

2. Background and motivation

2.1. Conceptualisation and definitions

In this paper, we consider a numerical model in the form

$$y = f(\mathbf{x}) \quad (1)$$

where $\mathbf{x} = [x_1, x_2, \dots, x_M] \in \mathcal{X} \subseteq \mathbb{R}^M$ is a vector of model inputs, i.e. any numerical variable that can be changed before model execution, and $y \in \mathbb{R}$ is the model output, i.e. a variable that is obtained after model execution.

The function f can be available in closed form or it may be given only in the form of a numerical procedure to compute y given \mathbf{x} , as in the case of simulation models where a set of differential equations is integrated over a spatial-temporal domain. In this case, the output y is a scalar variable that “summarises” the wide range of variables (often time series, possibly spatially-distributed) provided by the simulation procedure. Typically, y will either be a performance measure obtained by comparison with observations, for instance the root mean squared error, or a statistic of the simulated time series that is of interest per se, for instance the value of a variable at given time in a given location, or its average over a given spatial and temporal domain. In this case, the input–output relation of Eq. (1) is often referred to as *response surface* rather than “model” to avoid confusion with the underlying simulation model which might have more inputs and outputs than \mathbf{x} and y .

Among the inputs x_i we consider numerical and scalar variables, for instance the parameters appearing in the model equations. However, the definition can be extended to non-scalar quantities, such as the time series of input forcing of a simulation model, or even non-numerical quantities, e.g. the spatial resolution grid for numerical integration (Baroni and Tarantola, 2014).

2.2. Purposes of sensitivity analysis

SA investigates the relative contribution of the uncertainty (variability) of the inputs \mathbf{x} on the uncertainty (variability) in the output y . While local SA considers uncertainty stemming from

input variations around a specific point \bar{x} , global SA considers variations of the inputs within their entire feasibility space. In this paper we will focus on the latter case. Often, three specific purposes (settings) for global SA are defined (Saltelli et al., 2008):

- *Factor Priorization* (FP) aims at ranking the inputs x_i in terms of their relative contribution to output uncertainty.
- *Factor Fixing* (FF), or *screening*, aims at determining the inputs, if any, that do not give any contribution to output uncertainty.
- *Factor Mapping* (FM) aims at determining the regions in the inputs space that produce specific output values, for instance above a prescribed threshold.

Global SA aimed at FP and FF often employs sensitivity indices, or importance measures. These are synthetic indices that quantify the relative contribution to output uncertainty from each input. A sensitivity index of zero means that the associated input is non-influential (which is useful for FF) while the higher the index the more influential the input (FP).

2.3. Good properties of a sensitivity index

Liu and Homma (2009), building on Saltelli (2002b) and Iman and Hora (1990), discuss key properties that a “good” global sensitivity index should satisfy. These include:

- To be *global*, i.e. to consider inputs variations in the entire feasible space \mathcal{X} .
- To be *quantitative*, i.e. computable through a numerical, reproducible procedure.
- To be *model independent*, i.e. applicable independently of the form of the input–output relationship $f(\cdot)$ in Eq. (1), e.g. linear or non-linear, additive or non-additive, etc.
- To be *unconditional* on any assumed input value. An example of an approach that does not satisfy the above property is the entropy-based approach further described in Section 3.4, which computes sensitivity to the i -th input by comparing the output distributions that are obtained when all inputs vary and when they all vary but x_i . This approach is global, as inputs are let to vary in their entire feasibility space, however, it is not unconditional, as the results depend on the conditioning value of x_i .
- To be *easy to interpret*. For instance, variance-based sensitivity indices are “easy to interpret” in that they represent the contribution to the output variance due to variations of an input. They thus take values between zero and one, regardless of the range of variation of the output y . This is very helpful for cross-comparing SA results across different case studies or different definitions of the output.
- To be *easy to compute*. In this paper, when saying that a sensitivity index is “easy to compute” we mean that the computational procedure for its approximation is easy to implement, although it can still be time consuming to execute. For instance, the estimation of variance-based sensitivity indices can be time consuming, as it requires many evaluations of the model of Eq. (1), and still easy to implement because, once the sample $\langle y_j \rangle_{j=1, \dots, N}$ of output evaluations has been generated, the sensitivity indices can be approximated by simply applying a closed form, algebraic equation over the output samples (Saltelli, 2002a).
- To be *stable*, i.e. to provide consistent results from sample to sample or simulation to simulation. As this property might be difficult to demonstrate, a weak version might sound like: it

should be possible to easily assess the robustness of the estimated sensitivity index to different samples or different sample sizes. In this weak version, the property is strongly linked to the easy-to-compute property. In fact, if a sensitivity index is straightforward to compute than it is also very easy to assess its robustness and convergence by repeating computations over different bootstrapping resamples and for various sample sizes.

- To be *moment independent*, i.e. not assuming any specific moment of the output distribution to fully characterise the output uncertainty.

Variance-based sensitivity indices satisfy all the above properties but the last one. In fact, they rely on the assumption that the second-order moment, i.e. the output variance, is sufficient to fully characterise the output uncertainty, which might not be the case if the output distribution is multi-modal or if it is highly skewed, as discussed in the Introduction.

3. The PAWN sensitivity index

Density-based sensitivity indices are moment-independent indices because, by definition, they consider the entire probability distribution of the output rather than one of its moments only. They measure sensitivity by estimating the variations that are induced in the output distribution when removing the uncertainty about one (or more) inputs. More specifically, the sensitivity to input x_i is measured by the distance between the unconditional probability distribution of y that is obtained when all inputs vary simultaneously, and the conditional distributions that are obtained when varying all inputs but x_i (i.e. x_i is fixed at a nominal value \bar{x}_i).

In our approach, and in contrast to other density-based approaches, we characterise the conditional and unconditional distributions by their Cumulative Distribution Functions (CDFs) rather than their Probability Distribution Functions (PDFs). The reason for preferring CDFs is that they are much easier to approximate, as discussed in the Introduction. As a measure of distance between unconditional and conditional CDFs, we use the Kolmogorov–Smirnov statistic. With respect to other distance measures, this has several advantages. First, it varies between 0 and 1 regardless of the range of variation of the model output y , which ensures that our sensitivity index is an absolute measure. Secondly, when using our approach for Factor Fixing we can build on the statistical results of the two-sample Kolmogorov–Smirnov test (see for instance Wall (1996a)) to determine non-influential inputs at a given confidence level.

In the following, we will denote the unconditional cumulative distribution function of the output y by $F_y(y)$, and the conditional cumulative distribution function when x_i is fixed by $F_{y|x_i}(y)$. Since $F_{y|x_i}(y)$ accounts for what happens when the variability due to x_i is removed, its distance from $F_y(y)$ provides a measure of the effects of x_i on y . The limit case is when $F_{y|x_i}(y)$ coincides with $F_y(y)$ (case (a) in Fig. 1): it means that removing the uncertainty about x_i does not affect the output distribution, and one can conclude that x_i has no influence on y . If instead the distance between $F_{y|x_i}(y)$ and $F_y(y)$ increases, the influence of x_i increases as well (case (b) in Fig. 1). As a measure of distance between unconditional and conditional CDFs, we use the Kolmogorov–Smirnov statistic (Kolmogorov, 1933; Smirnov, 1939):

$$KS(x_i) = \max_y |F_y(y) - F_{y|x_i}(y)| \quad (2)$$

As KS depends on the value at which we fix x_i , the PAWN index T_i considers a statistic (e.g. the maximum or the median) over all possible values of x_i , i.e.

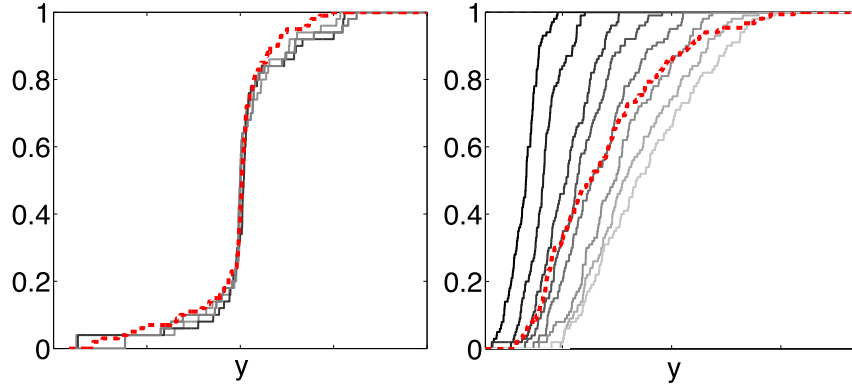


Fig. 1. Two illustrative examples of Cumulative Distribution Functions (CDFs) of the model output y . The red dashed line is the unconditional distribution function $F_y(\cdot)$ and the grey lines are the conditional distribution function $F_{y|x_i}(\cdot)$ at different fixed value of input x_i . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$T_i = \text{stat}_{x_i}[\text{KS}(x_i)] \quad (3)$$

By definition:

- (1) T_i varies between 0 and 1.
- (2) The lower the value of T_i , the less influential x_i .
- (3) If $T_i = 0$, then x_i has no influence on y .

The PAWN index T_i can be used for both Factor Prioritisation and Factor Fixing. It satisfies all the properties discussed in Section 2.3. In fact, it is global, quantitative and model independent. It is unconditional on any assumed input value because the dependency of the KS statistic on the value of x_i is removed by the statistic over x_i that appears in Eq. (3). It is easy to interpret because it is an absolute measure and thus its numerical value does not depend on the units of measurements of y . It is easy to compute because it can be easily approximated also when CDFs are not known, as further explained below. It is therefore also easy to analyse in terms of robustness and convergence.

3.1. Use for regional response sensitivity analysis

One more property of index T_i is that it can be easily tailored to focus on a particular sub-range of the output distribution, rather than considering the entire range. To this end, it is sufficient that the maximum appearing in Eq. (2) be taken with respect to y values in the sub-range of interest. A practical example is given in Section 4.3. Liu et al. (2006) use the term *Regional Response Probabilistic Sensitivity Analysis* (RRPSA) to refer to this type of global (or “probabilistic” in their terminology) SA that can focus on specific regions of the response surface. RRPSA is valuable when one is particularly interested in the model behaviour in a specific range of the output distribution, for instance extreme values as required in natural hazard assessment studies.

3.2. Numerical implementation

Since the analytical computation of the index T_i will be impossible in the majority of cases, we suggest the following approximate numerical procedure. First of all, the Kolmogorov–Smirnov statistic of Equation (2) is approximated by

$$\hat{KS}(x_i) = \max_y |\hat{F}_y(y) - \hat{F}_{y|x_i}(y)| \quad (4)$$

where $\hat{F}_y(\cdot)$ and $\hat{F}_{y|x_i}(\cdot)$ are the empirical unconditional and conditional CDFs. The unconditional CDF $\hat{F}_y(\cdot)$ is approximated using N_u output evaluations obtained by sampling the entire input feasibility space. The conditional CDF $\hat{F}_{y|x_i}(\cdot)$ instead is approximated using N_c output evaluations obtained by sampling the non-fixed inputs only, while keeping the value of x_i fixed.

Secondly, the statistic with respect to the conditioning value of x_i is replaced by its sample version, i.e. Eq. (3) is approximated by

$$\hat{T}_i = \text{stat}_{x_i = \bar{x}_i^{(1)}, \dots, \bar{x}_i^{(n)}} [\hat{KS}(x_i)] \quad (5)$$

where $\bar{x}_i^{(1)}, \bar{x}_i^{(2)}, \dots, \bar{x}_i^{(n)}$ are n randomly sampled values for the fixed input x_i . The steps in the numerical implementation of the PAWN index are summarised in Fig. 2.

The total number of model evaluations necessary to compute the sensitivity indices of Eq. (5) for all the M inputs is $N_u + n \times N_c \times M$. The values for n , N_u and N_c can be chosen by trial-and-error. As n sets the number of conditioning values sampled from the one-dimensional space of variation of x_i , its value might reasonably be in the order of few dozens (we used n between 10 and 50 in the examples of the next section), while N_u and N_c , which set the number of samples taken in the M -dimensional and $(M - 1)$ -dimensional space of all inputs and all-inputs-but- x_i , should be significantly higher. Still, given the regularity properties of CDFs (continuity, monotonicity, relative smoothness) our approximation strategy is quite effective even when using small values for N_c and N_u (in the order of few hundreds in the examples of the next section), to limit the total number of model evaluations. Furthermore, given the simplicity of the computation of the PAWN index, its robustness to the selected values of N_c and N_u can be quickly estimated by repeating computations using bootstrap resamples of different sizes from the same dataset of input/output samples. For instance, Fig. 6 shows the PAWN sensitivity indices with confidence intervals obtained by bootstrapping for the numerical example of Section 4.2, with fixed n and increasing values of N_c and N_u . Here, n is set to 50, N_c is first set to 20 and then increased to 30 and 50 (these numbers are quite low as in this example there are 2 inputs only, and therefore the conditional CDFs require sampling in a space of dimension $M - 1 = 1$); and N_u is first set to 200 and then increased to 300 and 500 (these numbers are higher since unconditional CDFs require sampling in a 2-dimensional space). Figure 6 shows that sensitivity indices have reached convergence and the associated confidence interval are rather small.

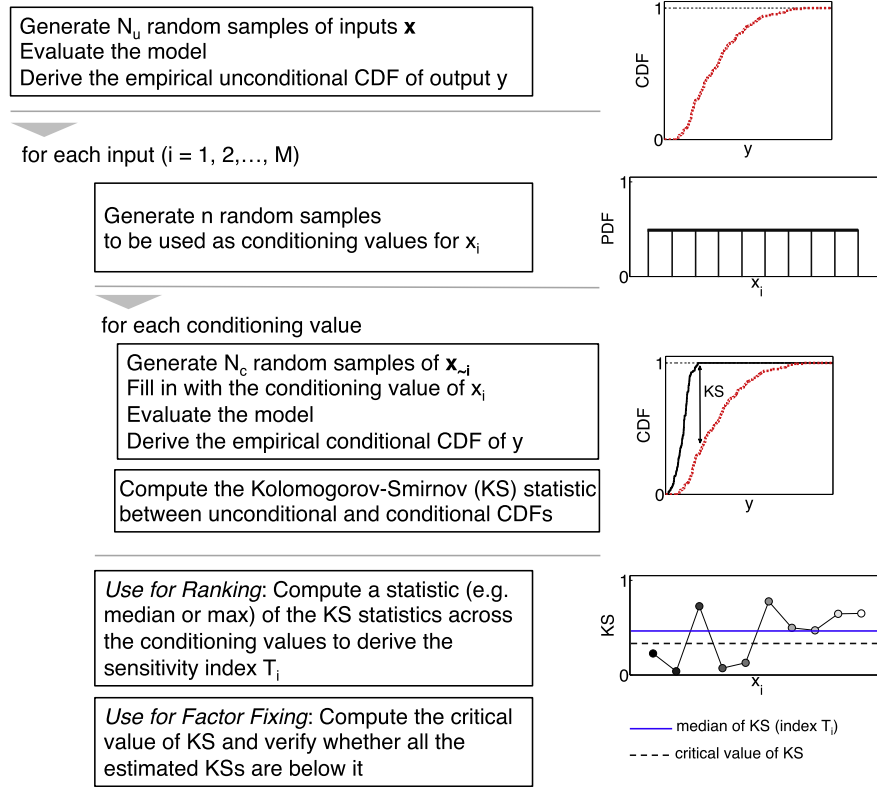


Fig. 2. The steps in the numerical implementation of the PAWN index. Here, $\mathbf{x} = [x_1, \dots, x_M]$ is the vector including all the uncertain inputs and $\mathbf{x}_{-i} = [x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_M]$ is the vector of all the inputs but the i -th.

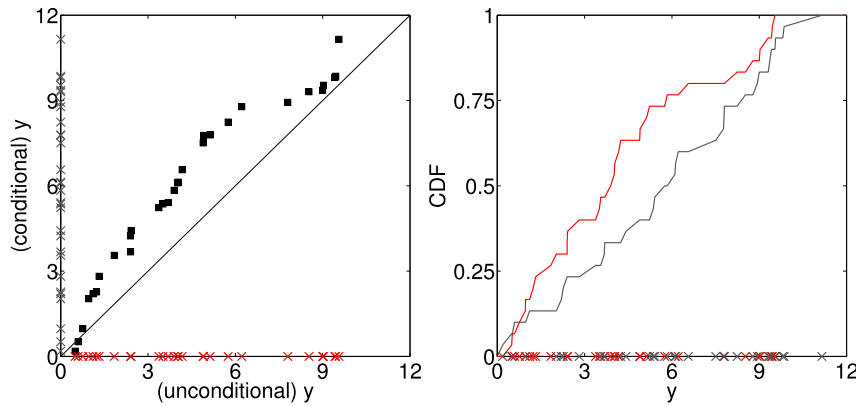


Fig. 3. Left: the visual test by Andres (1997) where a sample of “unconditional” output is plotted against a sample of “conditional” outputs (i.e. obtained when fixing input x_i to a prescribed value). If datapoints align around the bisector, then x_i is non influential. Right: in PAWN the similarity of the two output samples is assessed by the divergence between the associated unconditional (red) and conditional (grey) CDFs. Furthermore, in PAWN multiple conditional CDFs obtained at different conditioning values are compared (see Fig. 1). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3. Using the two-sample Kolmogorov–Smirnov test for Factor Fixing

If the specific purpose of SA is to determine non-influential inputs (Factor Fixing), the PAWN approach can be used in combination with the two-sample Kolmogorov–Smirnov test. The test rejects the hypothesis that two distributions (e.g. $F_y(\cdot)$ and $F_{y|x_i}(\cdot)$) are the same if

$$\widehat{KS} > c(\alpha) \sqrt{\frac{N_u + N_c}{N_u N_c}} \quad (6)$$

where \widehat{KS} is the Kolmogorov–Smirnov distance between the two empirical CDFs, N_u and N_c are the number of samples used to build the empirical CDFs, α is the confidence level, and the critical value $c(\alpha)$ can be found in the literature (see for instance Tables A VII, A VIII and A IX in Wall (1996a) or Wall (1996b)). In our context, rejecting the hypothesis implies that input x_i is influential. In fact, if x_i was non-influential, then the distributions $F_y(\cdot)$ and $F_{y|x_i}(\cdot)$ should coincide at all the conditioning values. Because in practice one cannot apply the test at all possible conditioning values, we suggest to use the frequency with which the test is passed over the n conditioning values $\bar{x}_i^{(1)}, \bar{x}_i^{(2)}, \dots, \bar{x}_i^{(n)}$ used to compute the approximate index \widehat{T}_i (see Eq. (5)). If the hypothesis that the two

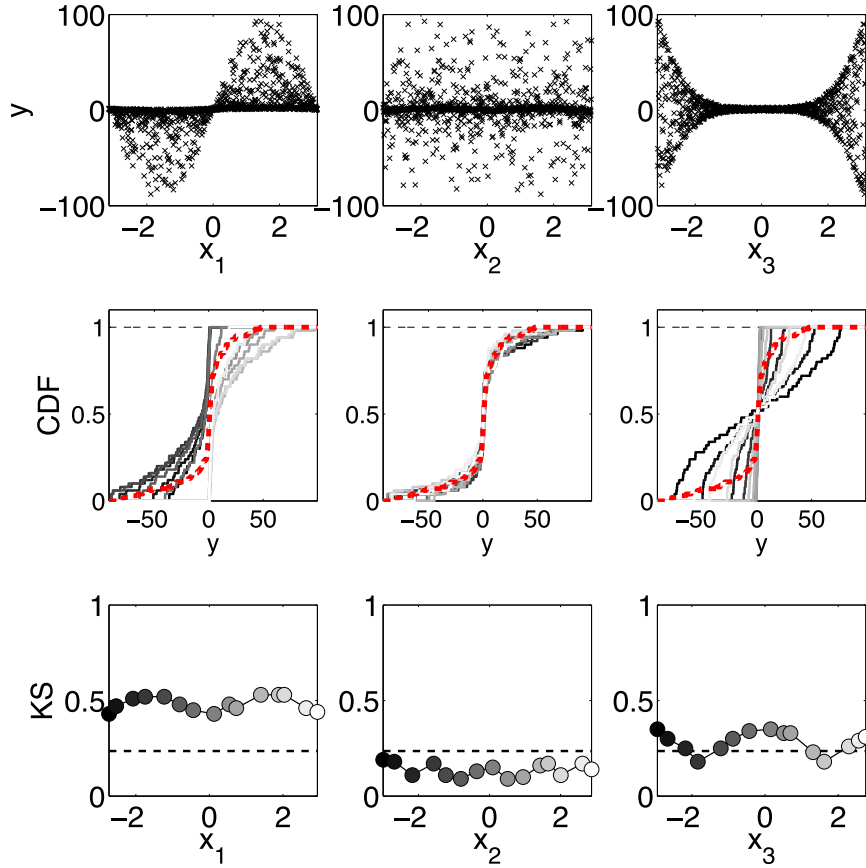


Fig. 4. Top panels: scatter plots of the Ishigami–Homma function of Eq. (9). Middle panels: empirical unconditional output distribution $\hat{F}_y(\cdot)$ (red dashed lines) and conditional ones $\hat{F}_{y|x_i}(\cdot)$ (grey solid lines). Bottom panels: Kolmogorov–Smirnov statistic $\hat{KS}(x_i)$ at different conditioning values of x_i . The dashed horizontal line is the critical value of the KS statistic at confidence level of 0.05. [Experimental setup: $n = 15$; $N_u = 100$; $N_c = 50$]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

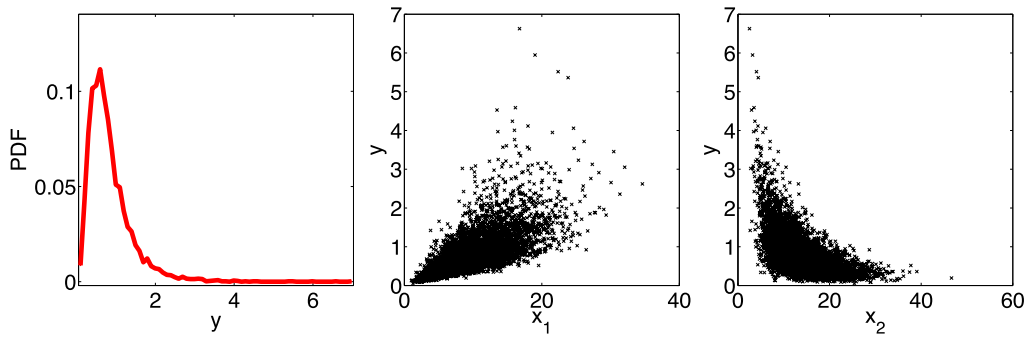


Fig. 5. Empirical unconditional PDF $\hat{f}(y)$ of the highly-skewed model of Eq. (10) and associated scatter plots.

CDFs be the same is never rejected, than input i can be considered non-influential.

The test can be also used as a means to verify the results of other SA methods and specifically their ability to correctly identify non-influential inputs. With this regard, it can be linked to the visual approach first proposed by Andres (1997) and used for instance in Tang et al. (2007). In this approach, exemplified in the left panel of Fig. 3, two output samples are compared in a scatter plot: the “unconditional” output samples that are obtained when letting all inputs vary, and the “conditional” samples obtained when fixing input x_i to a prescribed value. If datapoints align around the bisector of the first quadrant, it means that the output variability is not affected by removing the uncertainty about x_i and therefore the

hypothesis that x_i be non-influential is confirmed. The limitations of this validation approach are that it is qualitative, and that its results are conditional on the fixed value chosen for the presumed non-influential inputs. The two-sample Kolmogorov–Smirnov test for Factor Fixing can thus be seen as a quantitative, unconditional version of this approach.

3.4. Links to other density-based methods

As anticipated in the introduction, several other density-based methods have been proposed in previous studies, including entropy- and δ -sensitivity-based approaches. Just as our PAWN method, they are also based on the comparison of the

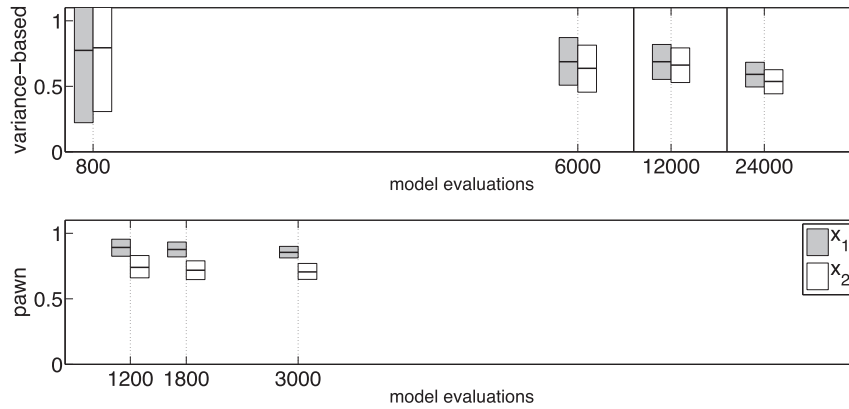


Fig. 6. Variance-based total effects (top) and PAWN (bottom) sensitivity indices for the highly-skewed model of Eq. (10) estimated using an increasing number of model evaluations. Boxes represent confidence intervals obtained by bootstrapping. Black lines indicate the mean index estimate for each input and each sample size.

unconditional output distribution and the conditional ones, however, they usually employ the output PDFs rather than the corresponding CDFs.

In entropy-based sensitivity indices, the divergence between the unconditional and the conditional PDFs is measured by either the Shannon entropy (Krykacz-Hausmann, 2001) or the Kullback–Leibler entropy (Park and Ahn, 1994; Liu et al., 2006). In order to reduce the computing burden of this estimation, Liu et al. (2006) compute the divergence at one conditioning value of x_i only. The drawback is that results are highly dependent on the choice of the conditioning value. In other words, according to the terminology introduced in Section 2.3, entropy-based indices are not “unconditional”. Another disadvantage of entropy-based sensitivity indices is that they do not possess an absolute meaning, given that entropy is a relative measure. Therefore they can only be used to make pairwise-comparisons between inputs within a specific experimental setup, and do not allow for comparison across different definitions of the model output, application site, etc.

The δ -sensitivity measure, first proposed by Borgonovo (2007), considers different conditioning values of x_i , and measures sensitivity using the average area enclosed between the conditional PDFs $f_{y|x_i}(\cdot)$ and the unconditional PDF $f_y(\cdot)$, i.e.

$$\delta_i = \frac{1}{2} E_{x_i} \int_{-\infty}^{+\infty} |f_y(y) - f_{y|x_i}(y)| dy \quad (7)$$

The δ -sensitivity measure does not require specifying a conditioning value of x_i and therefore is “unconditional”. Furthermore, it is “easy-to-interpret” because it is an absolute measure. In fact, by definition its value ranges between 0 and 1, and can be further bounded within this range as explained in Borgonovo (2007).

However, the computation of the δ -sensitivity measure still suffers from the disadvantage that it requires approximating several PDFs. To overcome this last issue, Liu and Homma (2009) propose to replace the area in between $f_y(\cdot)$ and $f_{y|x_i}(\cdot)$ by a distance metric between the corresponding CDFs. They show that, if the PDFs intersect at one point only, the area enclosed by the two PDFs is equal to twice the absolute vertical distance between the two CDFs, evaluated at the intersection point. Because the intersection point is also the one where the absolute vertical distance between the two CDFs is maximum, one can redefine the δ -sensitivity measure as

$$\delta_i = E_{x_i} \max_y |F_y(y) - F_{y|x_i}(y)| \quad (8)$$

which coincides with our sensitivity measure of Eq. (3) when $\text{stat} = E$. However, because the expected value can be very sensitive to extreme values of KS that might be obtained for some specific conditioning value of x_i , in our approach we rather suggest to use the median as a summary statistic, possibly complemented by the maximum, which could help spotting any of such extreme behaviour.

Moreover, as also highlighted by Liu and Homma (2009), the main computational difficulty with this approach is that to obtain the number and position of the intersection points one has to solve the differential equation $d(F_y(y) - F_{y|x_i}(y))/dy = 0$. In some cases, the computational complexity of solving this differential equation might cancel out the numerical advantage of using CDFs rather than PDFs.

3.5. Difference between density-based methods and Regional Sensitivity Analysis

Another well established GSA method that is somewhat related to PAWN is Regional Sensitivity Analysis (RSA), which was first proposed and investigated in Young et al. (1978) and Spear and Hornberger (1980). RSA-like approaches are also referred to as “Monte Carlo filtering” (Saltelli et al., 2008).

In RSA, the input samples are divided into two (or more) groups depending on whether the associated model output satisfies a given condition, e.g. is above/below a given threshold. Then, the empirical CDFs of each input in each group are computed and compared. Visual inspection of input CDFs across groups provides information useful for mapping, for instance by highlighting a reduction in the variability range of that input within a specific group. Application of formal statistics, for instance the Kolmogorov–Smirnov statistic, can be used to quantify the divergence between the CDFs and thus serve as a criterion for input ranking. The underlying assumption is that if the input CDF varies significantly from one group to another then sensitivity to that input is high.

Our PAWN approach thus has the use of CDFs in common with RSA, however, it is very different in the way in and aims for which CDFs are applied. RSA considers variations in the CDFs of the inputs while PAWN considers variations in the CDF of the output. This is not only a technical difference but also reflects a different philosophy and purpose. In RSA, the focus is on the input space and how input distributions vary when conditioning the output, which is mainly interesting for factor mapping. Input CDFs are used as a means to visualise the variations induced by such mapping. In PAWN the focus is on the output, and how the output distribution varies when conditioning an input. Variations of the output CDF

provide the quantitative basis for factor prioritization and fixing, though we have also shown that additional insights about factor mapping can be gathered when analysing disaggregated PAWN results.

4. Examples applications

We now use a set of static and dynamic models to demonstrate our method and discuss its properties. First, we apply PAWN to one of the most frequently used benchmark models in the SA literature, the Ishigami–Homma function (see for instance Saltelli et al. (2008)) with the purpose of illustrating the working principles of our method and verifying that it produces sensible results. Then, we use another numerical case study taken from Liu et al. (2006) to demonstrate the advantage of PAWN over variance-based indices when the output distribution is highly-skewed. Finally, we apply it to a dynamic model from the hydrology literature, the HyMod rainfall-runoff model, to show how PAWN can be used for Regional Response SA and for mapping back information about the output sensitivity into the input space (so called Factor Mapping).

4.1. Ishigami–Homma function

We first consider the widely-used Ishigami–Homma function (see for instance Eq. (4.34) in Saltelli et al. (2008))

$$y = \sin(x_1) + a \sin(x_2)^2 + bx_3^4 \sin(x_1) \quad (9)$$

where all x_i follow a uniform distribution over $[-\pi, +\pi]$, and $a = 2$ and $b = 1$. The top panels of Fig. 4 reports the scatter plots of the output against the three inputs. It can be noticed that: (i) x_1 seems to be the most influential input; (ii) x_2 seems to be non-influential. This is confirmed by variance-based SA: in fact, the total effects index of x_1 is the highest and the total effects of x_2 is almost zero (specifically: $S_{T_1} = 0.9991$; $S_{T_2} = 0.0009$; $S_{T_3} = 0.6161$; these numbers can be derived analytically via the equations given in Saltelli et al. (2008)).

If we apply our PAWN approach we obtain the three sets of conditional CDFs in the middle panels of Fig. 4. In all panels, the red dashed line is the empirical unconditional output CDF $\hat{F}_y(\cdot)$ while the grey lines are the conditional CDFs $\hat{F}_{y|x_i}(\cdot)$ ($i = 1, 2, 3$) obtained at n different conditioning values of x_i . These values can be read on the horizontal axis of the respective bottom panels, which report the Kolmogorov–Smirnov statistics estimated by Eq. (4). The horizontal line is the value $c(\alpha)\sqrt{(N_u + N_c)/(N_u N_c)}$ (with $\alpha = 0.05$) for the two-sample Kolmogorov–Smirnov test. The Figure shows that:

- Visual analysis of the CDFs (middle panel of Fig. 4) immediately shows that x_1 and x_3 are both much more influential than x_2 , as their conditional CDFs are more widespread around the unconditional one (red line), while those of x_2 almost overlap with the unconditional CDF.
- In quantitative terms, x_1 is the most influential, as shown by the analysis of the Kolmogorov–Smirnov statistic (bottom panel). In particular, the PAWN sensitivity indices are equal to $T_1 = 0.48$, $T_2 = 0.14$ and $T_3 = 0.3$ if considering the median (stat = median in Eq. (5)) and $T_1 = 0.53$, $T_2 = 0.19$ and $T_3 = 0.35$ if considering the max (stat = max in Eq. (5)). In other words, when used for Factor Prioritisation, PAWN provides the same ranking as variance-based SA.
- By applying the two-sample Kolmogorov–Smirnov test for Factor Fixing, one would conclude that x_2 is non-influential (at confidence level $\alpha = 0.05$) because its KS statistics are consistently below the threshold value (see central bottom panel in Fig. 4). This is again consistent with a qualitative

judgement that might be formulated based on variance-based SA results where the total effects index of x_2 is very low ($S_{T_2} = 0.0009$).

4.2. Highly-skewed function

We now consider the simple nonlinear model proposed by Liu et al. (2006):

$$y = x_1/x_2 \quad (10)$$

where the two inputs x_1 and x_2 both follow a χ^2 distribution with degrees of freedom set to 10 and 13.978, respectively. The resulting distribution of y is positively-skewed with a heavy right tail (see left panel in Fig. 5). Liu et al. (2006) analyse the propagation of uncertainty through the model and show that the effect of x_1 is higher than that of x_2 . This result is also confirmed by the scatter plots in Fig. 5.

However, variance-based sensitivity analysis fails to reveal the difference between the inputs and gives both inputs the same importance. This is revealed in the top panel in Fig. 6, which shows the variance-based total effects sensitivity indices of the two model inputs, estimated at different sample sizes using the formulae by Saltelli (2002a). For each sample size, confidence intervals (at level $\alpha = 0.05$) are also obtained by bootstrapping. The two inputs are attributed similar total effects with confidence intervals largely overlapping even with a rather large sample size, although we know that x_1 is more influential than x_2 . The reason is that variance is not an adequate proxy for uncertainty when dealing with such a highly-skewed distribution.

The bottom panel of Fig. 6 reports sensitivity estimates by the PAWN method, i.e. via Eq. (5) where stat = max. We derive upper and lower confidence bounds by repeating computations using 1000 bootstrap samples of the output samples to estimate the unconditional and conditional CDFs $F_y(\cdot)$ and $F_{y|x_i}(\cdot)$. The Figure shows that PAWN is able to effectively discriminate the higher influence of input x_1 while using a much lower number of model evaluations.

4.3. Dynamic system example: the HyMod model

Lastly we consider an illustrative SA of the parameters of the HyMod model. HyMod is a lumped conceptual hydrological model that can be used to simulate rainfall-runoff processes at the catchment scale. It was first introduced by Boyle (2001) and is described in Wagener et al. (2001). The application study site is the Leaf River catchment, a 1950 km² catchment located north of Collins, Mississippi, USA. A detailed description of the Leaf catchment can be found in Sorooshian et al. (1983). HyMod is characterised by five storages, the soil moisture reservoir – represented by a Pareto distribution function to describe the rainfall excess model (see Moore (1985)), three linear reservoirs in series mimicking the fast runoff component, and one slow reservoir. HyMod has five parameters: the maximum soil moisture storage capacity (Sm); a coefficient accounting for the spatial variability of soil moisture in the catchment (β); the ratio of effective rainfall that is sent to the fast reservoirs (α), the discharge coefficient of the fast reservoirs (R_F) and that of the slow reservoir (R_S). During simulation, differential equations are solved using the forward explicit Euler method with a daily resolution time series of rainfall (mm/day) and evaporation (mm/day) over a simulation horizon of two years starting from 10/10/1948.

We apply PAWN to investigate the propagation of parameter uncertainty when the model is used for flood analysis. We thus take

Table 1
Parameters of the HyMod model, their units of measurements and feasible ranges.

	Parameter	uom	Range
x_1	Maximum soil moisture storage capacity Sm	(mm)	[0,400]
x_2	Coefficient of spatial variability of soil moisture β	(–)	[0.1,2]
x_3	Ratio of effective rainfall to the fast pathway α	(–)	[0.03,0.6]
x_4	Discharge coefficient of the fast reservoirs R_F	(day ^{–1})	[0.03,0.1]
x_5	Discharge coefficient of the slow reservoirs R_S	(day ^{–1})	[0.6,1]

as a scalar output y the maximum predicted flow over the simulation horizon, and as inputs x_i ($i = 1, \dots, 5$) the model parameters. The feasible ranges of variation of the parameters (inputs) are given in Table 1. They are defined in such a way that any model parameterisations within the range would guarantee a “sensible” output performance (i.e. exhibit a coefficient of determination $R^2 = 1 - \text{VAR}(e)/\text{VAR}(q)$, where $\text{VAR}(e)$ is the variance of simulation errors and $\text{VAR}(q)$ is the variance of observed flows, higher than 0.6).

The left panel of Fig. 7 shows the PAWN sensitivity indices for the five model parameters and associated confidence intervals. It indicates that the most influential parameters are α , which controls the amount of effective rainfall that goes into the fast pathway, and R_F , which controls how quickly water moves through the fast pathway. This is consistent with our choice of the output y , in fact, one would expect that the value of the maximum flow be essentially determined by the characteristics of the fast pathway in the flow-routing module, while it should be much less sensitive to the parameterisation of the soil moisture account (Sm, β) and of the slow pathway (R_S).

As discussed in Section 3, one advantage of the PAWN sensitivity index is that it can be very easily focused on specific sub-ranges of the output. As a matter of illustration, the right panel of Fig. 7 reports the PAWN indices estimated over the range $y > 50$, i.e. applying Eq. (5) where the Kolmogorov–Smirnov statistic is computed as

$$\hat{KS}(x_i) = \max_{y > 50} |\hat{F}_y(y) - \hat{F}_{y|x_i}(y)| \quad (11)$$

The Figure shows that when focussing on such sub-range, the ranking of the parameters changes: parameters Sm, β, R_S become even less influential while parameters α and R_F swap their relative positions in the ranking. Here, the sub-range $y > 50$ was chosen arbitrarily, the point being to show that the analysis of ranking reversals at different definitions of the output sub-range in PAWN is extremely simple, since it only requires reapplying Eqs. (11) and (5) to the same collection of CDFs while no new runs of the simulation model are needed.

Another advantage of PAWN is that the intermediate results used to compute the sensitivity index can be effectively visualised to gather more insights into the model behaviour, for instance to find threshold values in the input space corresponding to shifts in the output distribution. For example, Fig. 8 shows the unconditional output distribution (red dashed) and the conditional ones (grey) when fixing parameter α . The right panel plots the Kolmogorov–Smirnov statistics of each conditional distribution against the conditioning value of α . This picture shows that the distance between conditional and unconditional CDFs is minimum around the value of 0.3; when $\alpha < 0.3$ the conditional CDF is shifted to the left of the unconditional one, when $\alpha > 0.3$ it is shifted to the right. This is consistent with the expected model behaviour, in fact, as α increases, more and more effective rainfall is sent to the fast flow-routing pathway and correspondingly the probability of higher value of maximum flow y is increased.

5. Conclusions

In this paper, we have introduced a new approach to define and compute a density-based sensitivity index, called PAWN. Our sensitivity index measures the influence of a given input as the variation in the output CDF when the uncertainty about that input is removed. In practice, this is done by computing the Kolmogorov–Smirnov statistic KS between the empirical unconditional CDF $F_y(\cdot)$ of the model output and the conditional distribution $F_{y|x_i}(\cdot)$ when x_i is fixed. As results may differ depending on the fixed value for x_i , the KS statistic is computed at several conditioning values and the sensitivity index is a statistic, for instance the maximum or the median, of the individual results.

The index so defined can be used for ranking the inputs in terms of their contributions to the output uncertainty, as well as screening non-influential inputs. For the latter purpose, we show how to integrate the two-sample Kolmogorov–Smirnov test into our implementation procedure to formally assess whether to reject or accept the hypothesis that an input is non-influential.

With respect to the widely used variance-based sensitivity indices, the main advantage of PAWN is that it can be applied whatever the type of output distribution, including highly skewed ones. Furthermore, PAWN can be easily tailored to focus on a specific sub-range of the output, for instance extreme values, which may be very interesting for natural hazard models or any other application where the tail of the output distribution is of particular interest. With respect to other density-based sensitivity indices, which also share the two properties discussed above, PAWN has the advantage of being very easy to implement, which also facilitates

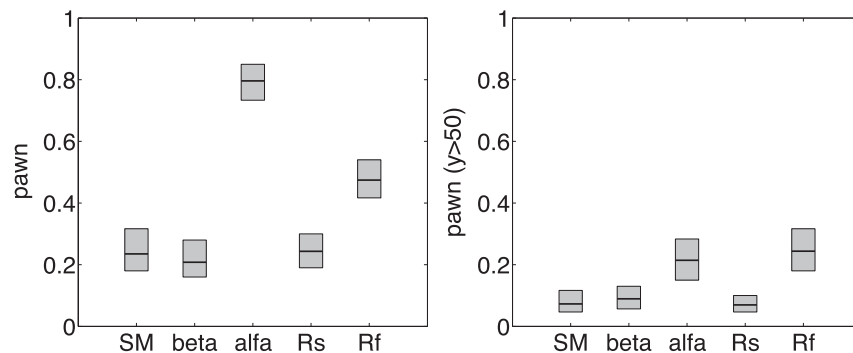


Fig. 7. PAWN sensitivity indices of the HyMod model estimated considering the entire output range (left) or the sub-range where $y > 50$ (right). Boxes represent confidence intervals obtained by bootstrapping, black lines indicate the mean index estimate. [Experimental setup: sampling strategy: Latin-Hypercube; $n = 10$; $N_u = 150$; $N_c = 100$; number of bootstrap resamples: 200].

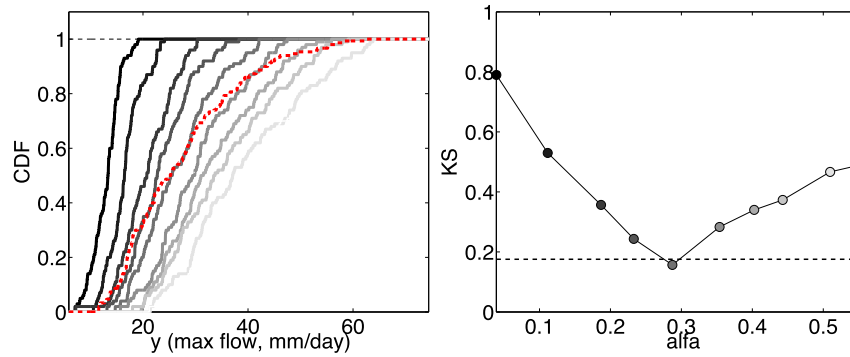


Fig. 8. Left: empirical unconditional output distribution $\hat{F}_y(\cdot)$ (dashed red line) and conditional ones $\hat{F}_{y|x_i}(\cdot)$ (solid grey lines) for input 3 (parameter α) of the HyMod model. Right: Kolmogorov–Smirnov statistic at different conditioning values of α , the dashed horizontal line is the critical value of the KS statistic at confidence level of 0.05. [Experimental setup: sampling strategy: Latin-Hypercube; $n = 10$; $N_u = 150$; $N_c = 100$]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the assessment of convergence and accuracy, and very easy to interpret, thanks to a number of visualisation tools (see for instance Fig. 8) that can be used to investigate intermediate results. These visualisation tools also allow the user to map the changes in the output distribution back into the input space, to gain more insights about the model behaviour.

While the implementation of the PAWN index from a given dataset of output samples is very straightforward, the major computational bottleneck is the generation of the dataset itself. Preliminary results suggest that the PAWN approach should not be more demanding than other GSA methods, as accurate and converging sensitivity estimates were obtained using a relatively low number of output samples in all the three applications reported here. More research is needed to extensively assess the convergence properties of the PAWN index as well as provide more detailed guidelines on the choice of the parameters n , N_u and N_c that determine the estimation accuracy but also the total number of model evaluations.

In this paper, we have shown results of the PAWN method when using an ad hoc sampling strategy, as detailed in Section 3.2. However, PAWN can also be applied to an already available dataset. In this case, the unconditional CDF could be approximated using all the available output samples, while the conditional CDFs $\hat{F}_{y|x_i}(\cdot)$ could be derived using n sub-samples, corresponding to n clusters of the sampled values of x_i . The cluster centres $\bar{x}_i^{(1)}, \bar{x}_i^{(2)}, \dots, \bar{x}_i^{(n)}$ should be selected to cover the range of variability of x_i as uniformly as possible, while samples falling in the k -th clusters should have a value of x_i such that the distance $|x_i - \bar{x}_i^{(k)}|$ does not exceed a prescribed threshold. Further research will investigate the loss in accuracy due to such sub-optimal implementation of the PAWN index.

Although not shown in this paper, the PAWN approach can also be applied to assess the influence of groups of inputs. For instance, if one wants to assess the influence of a pair of inputs (x_i, x_j) , or test the assumptions that both inputs are non-influential, it will be sufficient to compute the PAWN index, or apply the two-sample Kolmogorov–Smirnov test, using the conditional distributions $F_{y|x_i, x_j}(\cdot)$ where both inputs are fixed.

Another topic for further investigation is the assessment of interactions between uncertain inputs. One possible way to do this is by comparing the PAWN indices that are obtained when conditioning input x_i only, and x_i grouped with other inputs. If results are significantly different, it means that the interactions of x_i with the other inputs are very important, because removing the uncertainty of x_i alone has less influence on the output distribution than removing the joint uncertainty of x_i and grouped inputs.

While further investigations will address these topics, we suggest that the PAWN approach can already be used to validate and complement other SA approaches, especially the widely used variance-based approach. Consistent results between PAWN and the variance-based approach will reinforce the conclusions of SA, while contradictory outcomes can reveal where implicit assumptions, e.g. that variance is a sensible indicator of output uncertainty, may not be satisfied, and provide directions for further investigation of the model's response surface.

Acknowledgements

This work was supported by the Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1]. The matlab/octave code to implement the PAWN method is freely available for non commercial purposes as part of the SAFE Toolbox (www.bristol.ac.uk/cabot/resources/safe-toolbox/).

References

- Anderson, B., Borgonovo, E., Galeotti, M., Roson, R., 2014. Uncertainty in climate change modeling: can global sensitivity analysis be of help? *Risk Anal.* 34 (2), 271–293.
- Andres, T., 1997. Sampling methods and sensitivity analysis for large parameter sets. *J. Stat. Comput. Simul.* 57 (1–4), 77–110.
- Baroni, G., Tarantola, S., 2014. A general probabilistic framework for uncertainty and global sensitivity analysis of deterministic models: a hydrological case study. *Environ. Model. Softw.* 51, 26–34.
- Borgonovo, E., 2007. A new uncertainty importance measure. *Reliab. Eng. Syst. Saf.* 92, 771–784.
- Borgonovo, E., Castaings, W., Tarantola, S., 2011. Moment independent importance measures: new results and analytical test cases. *Risk Anal.* 31 (3), 404–428.
- Boyle, D., 2001. Multicriteria Calibration of Hydrological Models (Ph.D. thesis), Dep. of Hydrol. and Water Resour., Univ. of Ariz., Tucson.
- Castaings, W., Borgonovo, E., Morris, M., Tarantola, S., 2012. Sampling strategies in density-based sensitivity analysis. *Environ. Model. Softw.* 38 (0), 13–26.
- Iman, R., Hora, S., 1990. A robust measure of uncertainty importance for use in fault tree system analysis. *Risk Anal.* 10, 401–406.
- Kolmogorov, A., 1933. Sulla determinazione empirica di una legge di distribuzione. *G. Ist. Ital. Attuari* 4, 83–91.
- Krykacz-Hausmann, B., 2001. Epistemic sensitivity analysis based on the concept of entropy. In: Prado, P., Bolado, R. (Eds.), *Proceedings of SAMO2001*, Madrid, pp. 31–35.
- Liu, H., Sudjianto, A., Chen, W., 2006. Relative entropy based method for probabilistic sensitivity analysis in engineering design. *J. Mech. Des.* 128, 326–336.
- Liu, Q., Homma, T., 2009. A new computational method of a moment-independent uncertainty importance measure. *Reliab. Eng. Syst. Saf.* 94, 1205–1211.
- Moore, R., 1985. The probability-distributed principle and runoff production at point and basin scales. *Hydrol. Sci. J.* 30 (2), 273–297.
- Nossent, J., Elsen, P., Bauwens, W., 2011. Sobol sensitivity analysis of a complex environmental model. *Environ. Model. Softw.* 26 (12), 1515–1525.

- Pappenberger, F., Beven, K., Ratto, M., Matgen, P., 2008. Multi-method global sensitivity analysis of flood inundation models. *Adv. Water Resour.* 31 (1), 1–14.
- Park, C., Ahn, K., 1994. A new approach for measuring uncertainty importance and distributional sensitivity in probabilistic safety assessment. *Reliab. Eng. Syst. Saf.* 46, 253–261.
- Pastres, R., Chan, K., Solidoro, C., Dejak, C., 1999. Global sensitivity analysis of a shallow-water 3D eutrophication model. *Comput. Phys. Commun.* 117, 62–74.
- Peeters, L.J.M., Podger, G.M., Smith, T., Pickett, T., Bark, R.H., Cuddy, S.M., 2014. Robust global sensitivity analysis of a river management model to assess nonlinear and interaction effects. *Hydrol. Earth Syst. Sci.* 18 (9), 3777–3785.
- Plischke, E., Borgonovo, E., Smith, C.L., 2013. Global sensitivity measures from given data. *Eur. J. Oper. Res.* 226 (3), 536–550.
- Saltelli, A., 2002a. Making best use of model valuations to compute sensitivity indices. *Comput. Phys. Commun.* 145, 280–297.
- Saltelli, A., 2002b. Sensitivity analysis for importance assessment. *Risk Anal.* 22, 579–590.
- Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., Tarantola, S., 2010. Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Comput. Phys. Commun.* 181 (2), 259–270.
- Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., Tarantola, S., 2008. *Global Sensitivity Analysis, The Primer*. Wiley.
- Smirnov, N., 1939. On the estimation of the discrepancy between empirical curves of distribution for two independent samples. *Bull. Math. Univ. Moscou* 2 (2).
- Sorooshian, S., Gupta, V., Fulton, J., 1983. Evaluation of maximum likelihood parameter estimation techniques for conceptual rainfall-runoff models: influence of calibration data variability and length on model credibility. *Water Resour. Res.* 19, 251–259.
- Spear, R., Hornberger, G., 1980. Eutrophication in peel inlet. II. Identification of critical uncertainties via generalized sensitivity analysis. *Water Res.* 14 (1), 43–49.
- Tang, Y., Reed, P., Wagener, T., van Werkhoven, K., 2007. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. *Hydrol. Earth Syst. Sci.* 11, 793–817.
- van Werkhoven, K., Wagener, T., Reed, P., Tang, Y., 2008. Rainfall characteristics define the value of streamflow observations for distributed watershed model identification. *Geophys. Res. Lett.* 35, L11403.
- Wagener, T., Boyle, D., Lees, M., Wheeler, H., Gupta, H., Sorooshian, S., 2001. A framework for development and application of hydrological models. *Hydrol. Earth Syst. Sci.* 5, 13–26.
- Wall, J.V., 1996a. Practical statistics for astronomers — ii. Correlation, data-modelling and sample comparison. *Q. J. R. Astron. Soc.* 37, 519–563.
- Wall, J.V., 1996b. Tests for Comparison of Two Independent Samples. http://ned.ipac.caltech.edu/level5/Wal12/Wal4_3.html (accessed 20.12.14.).
- Young, P.C., Spear, R.C., Hornberger, G.M., 1978. Modeling badly defined systems: some further thoughts. In: *Proceedings SIMSIG Conference*, Canberra, pp. 24–32.
- Ziliani, L., Surian, N., Coulthard, T., Tarantola, S., 2013. Reduced-complexity modeling of braided rivers: assessing model performance by sensitivity analysis, calibration, and validation. *J. Geophys. Res. Earth Surf.* 118 (4), 2243–2262.