Available online at www.sciencedirect.com
ScienceDirect
Advances in Water Resources 31 (2008) 1–14
Advances in Water Resources
www.elsevier.com/locate/advwatres

Multi-method global sensitivity analysis of flood inundation models

Florian Pappenberger [a,*], Keith J. Beven [a,d], Marco Ratto [b], Patrick Matgen [c]

[a] *Hydrology and Fluid Dynamics Group, Environmental Science, Lancaster University, Lancaster LA1 4YQ, UK*
[b] *Joint Research Centre, The European Commission, Ispra, Italy*
[c] *Department of Environment and Agro-Biotechnologies, Public Research Centre Gabriel Lippmann, Luxembourg*
[d] *Geocentrum, Uppsala University, Uppsala, Sweden*

Received 26 July 2006; received in revised form 25 April 2007; accepted 26 April 2007
Available online 24 May 2007

Abstract

Global sensitivity analysis is a valuable tool in understanding flood inundation models and deriving decisions on strategies to reduce model uncertainty. In this paper, a sensitivity analysis of a one-dimensional flood inundation model (HEC-RAS) on the River Alzette, Luxembourg, is presented. It is impossible to define sensitivity in a unique way and different methods can lead to a difference in ranking of importance of model factors. In this paper five different methods (Sobol, Kullback–Leibler entropy, Morris, regionalised sensitivity analysis and regression) are applied and the outcomes on selected examples compared. It is demonstrated that the different methods lead to completely different ranking of importance of the parameter factors and that it is impossible to draw firm conclusions about the relative sensitivity of different factors. Moreover, the uncertainty inherent in the sensitivity methods is highlighted.

Keywords: Global sensitivity analysis; Flood modelling; Uncertainty; Inundation model; Decision support; Flood risk

1. Introduction

Quantitative uncertainty analysis is currently a popular topic in hydrological and hydraulic modelling as well as for many other complex systems [22]. At the same time, understanding, rather than simply evaluation of the influence of different uncertainties of model factors on the modelling outcome, has become a key question [49]. A factor is any source of uncertainty in the modelling process including parameters, boundary conditions, model structure, etc. Such understanding is especially important if the goal is to reduce the model uncertainties when resources for making model runs and collecting additional data are limited [16]. Pappenberger et al. [46,47] and Hall et al. [21] present cases that try to achieve such an understanding for flood inundation models with the help of sensitivity analysis (SA). The study of Pappenberger et al. [47] highlights the need for an iterative modelling process based on sensitivity analysis to refine, develop and understand flood inundation models (see also [15,18,30,31]). Hall et al. [21] analyse the sensitivity of the Manning channel roughness parameter changes along the river and highlight areas in which improved sampling or modelling strategies could improve flood inundation predictions. The sensitivity analysis used in both studies provides insight into the robustness of model results, which will be important in any decision making process that depends on the results of a simulation model [20,23,48,59], for example in risk management [2] or mitigation [29].

In past studies, SA has been categorized in multiple ways [14,20,23,27] and in this paper we adopt the division into two broad categories: local SA and global SA (see discussion of this topic in [17]). Local sensitivity analysis concentrates on the local impact of factors on the model and is therefore suited for investigations around a specific point of interest (e.g. height of a flood defence) or optimisation techniques. Beven and Binley [9] have rejected the concept of one optimum model result and argue that multiple

* Corresponding author. Present address: European Centre for Medium Range Weather Forecasts, Reading, UK.
E-mail address: florian.pappenberger@ecmwf.int (F. Pappenberger).

0309-1708/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.advwatres.2007.04.009

combinations of the values of different factors will usually be found that are equally acceptable in comparison to evaluation data (the *equifinality thesis* – for a discussion see [7]). In such a framework, local sensitivities in the area of individual models considered to be acceptable might be quite different, and global sensitivity analysis (GSA) is more appropriate as it apportions the output uncertainty to the uncertainty in the input that cover the potential range for each factor [49,53]. The phenomenon of equifinality has been observed in the uncertainty analysis of many flood inundation models of different types [1,3,25,41,51,60]. Therefore, the methodology in this paper will be based on GSA techniques.

The first part of this paper describes the concept of sensitivity and explores the effect that different definitions of 'sensitivity' have in the context of flood inundation modelling. The ambiguity in those definitions can have a significant impact on any further modelling exercise and makes it necessary to embed the analysis into a multi-method global sensitivity analysis (MMGSA). This will be demonstrated on the example of a flood inundation model of the River Alzette (Grand Duchy of Luxembourg).

2. Settings for sensitivity analysis

A more detailed discussion is necessary in order to understand the differences in SA methods. Therefore, this section first discusses the definition of sensitivity in respect to their factors and the definition of sensitivity itself.

2.1. First, higher and total order of factor sensitivity

A model factor contributes to the sensitivity of a model outcome through variations of this factor alone and by interaction with other factors. The influence a factor exhibits on its own is called first order sensitivity and the effects due to interactions are called higher order effects. The total sensitivity of a model factor is defined through its first order effects plus all higher order effects. For correlated inputs it can occur that the first order effects are larger than the total effects exclusively due to the dependence structure in the inputs. Therefore, a small total sensitivity can still indicate sensitivity. However, this depends largely on the way the sensitivity method is formulated (e.g. the Sobol method assumes additive variances of individual effects and thus should not exhibit this behaviour). Roughly speaking, total sensitivity is derived through varying the value of one factor at a time on a number of MC randomly selected base points in the input factors space. Higher order sensitivity is derived through structured sampling patterns.

It is very desirable to quantify higher order effects, such as interaction between parameters in order to understand the full model complexity. However, it can be debated whether higher order effects, however estimated or computed, make '*sense*'. Saltelli et al. [53] point out that it is not possible to compute higher order effects for non-orthogonal input factors without significant additional assumptions (see also [28]). Therefore, for many physically-based flood inundation models, or indeed for most environmental models, it is impossible to derive a 'correct' covariation/interaction matrix. Channel and floodplain friction can be used as an example to illustrate this problem: Knight and Shiono [32] have shown that the complex interactions between a floodplain and a channel result in a complex covariance/interaction structure between channel and floodplain friction. However, most inundation models are calibrated on inundation extent or on single point hydrographs. Neither of these sources of data can give enough information about the complex interaction structure driving the model behaviour, be it in terms of covariance or other more sophisticated mapping techniques. For example, Hunter et al. [26] illustrate that if one calibrates inundation models against single extent images there is a trade off between values of channel and floodplain friction. Even if additional information on velocity structures within the flow domain could be made available, many inundation models would struggle to incorporate that structural information due to, for example, model scale issues. Measured and predicted velocities will not necessary be directly commensurable. In fact, if it would be possible to unambiguously constrain an inundation model on such local data, model equifinality [5,7] would vanish, at least in the case of a perfect model (for a discussion see [6]).

2.2. Defining sensitivity

Many reviews of sensitivity analysis methods have been conducted (see for example [14,16,20,23,31]). Some of them explicitly highlight the advantages and disadvantages of various methods and provide very good summaries of this topic (the reader is especially referred to [23,22]). Although none of these deal with flood inundation models directly, they contain conclusions which are applicable for the flood modelling community. For example, Frey and Patil [20], amongst others, state that different SA methods can lead to a difference in ranking of importance of model factors. Saltelli et al. [53] have argued that 'importance' or 'importance measures' (the way sensitivity is measured) should be defined at the stage of framing the analysis (called different SA settings). Such settings can also be linked to Type I and Type II errors. If one is particularly interested in avoiding Type I errors, then main effects and factors prioritisation setting will be the target analysis. Alternatively, if Type II errors are to be avoided, total effects and factor fixing need to be considered.

Mokhtari and Frey [38] have introduced a decision tree to choose a sensitivity analysis, similar to the one independently derived for uncertainty analysis ([45] on www.flood-risk.net). Kleijnen and Helton [31] as well as Frey and Patil [20] have proposed that multiple methods should be used to confirm the individual results. We will return to the issue of choice and application of different SA methods as part of the discussion of the results presented in this paper.

3. Methods of global sensitivity analysis

All sensitivity methods which are applied in this paper are based on the same sample of factors. The sampling algorithm used is the replicated latin hypercube design [53], which has been constructed using a replicated latin hypercube design (two replicas and the base sample). It has to be pointed out that the accuracy of the methodologies will depend on the efficiency of the sample to represent the entire response surface. Areas which are not sampled or are under-sampled will introduce errors in our analysis. As most flood inundation models are known to be highly nonlinear such an effect can be expected. A subset of all available methods has been chosen, which represent methodologies on a largely different theoretical basis. The selection is by no means complete and the reader is advised to visit the quoted references for more detailed information. All measures have been computed on the same sample to ensure comparability.

3.1. Sobol – analysis

If the uncertainty of input factors can be approximated by independent probability distributions, then sensitivity indices can be related to the decomposition of the total unconditional variance [56]. The decomposition can be shown in an ANOVA like way:

$$V = \sum_i V_i + \sum_{i<j} V_{ij} + \cdots + V_{1,2,3,\ldots,n} \tag{1a}$$

$$V_i = V[E(Y|X_i)] \tag{1b}$$

$$V_{T_i} = E[V(Y|X_{-i})] \tag{1c}$$

V is the total variance and E the expected value, V_i is the variance contribution due to effects of the random factor X_i. Y is the model response. Higher order terms show the variance contribution between two or more random variables. Each partial variance is normalised with respect to the total unconditional variance and allows sensitivity indices to be obtained:

$$\begin{aligned} S_i &= \frac{V_i}{V}, \quad 1 \leqslant i \leqslant n \\ S_{ij} &= \frac{V_{ij}}{V}, \quad 1 \leqslant i < n \\ S_{i,i+1,\ldots,n} &= \frac{V_{i,i+1,\ldots,n}}{V} \end{aligned} \tag{2}$$

S_i is called the main effect and S_{ij} the interaction effect between i and j. If the input factors are orthogonal, it is further possible to interpret the total effect of i (total sensitivity) as the sum of all terms of any order that include X_i.

The total effect measure will not be considered in this paper; instead, other measures that can rely on a cheaper sampling scheme than Sobol' total effects will be considered, with the aim of avoiding Type II errors.

3.2. Kullback–Leibler entropy

It has been argued that variance based methods rely too heavily on the assumption that the second moment is sufficient to describe the uncertainties and sensitivities encountered. These assumptions may be invalid if the distribution is highly skewed due to nonlinear functions or inputs [35]. Krykacz-Hausmann [33] argued that sensitivity should be measured as deviation of the factor distribution from being uniform (if the factor has been sampled from a uniform prior distribution). This is also exploited in methodologies such as the dynamic identifiability analysis [58] and its numerical expression by Horritt [24] in the form of an entropy measure. Based on entropy, Liu et al. [35] presented an alternative approach, which is using the Kullback–Leibler entropy [34]:

$$D_{\mathrm{KL}}(\mathrm{pdf}_1|\mathrm{pdf}_0) = \int_{-\infty}^{\infty} \mathrm{pdf}_1(y) \cdot \log \frac{\mathrm{pdf}_1(y)}{\mathrm{pdf}_0(y)}\, \mathrm{d}y \tag{3}$$

where D_{KL} is the Kullback–Leibler entropy, y the model response, pdf_0 the pdf of the response $y = h(\mathbf{X})$ and pdf_1 is the pdf of the response $y = h(\mathbf{X}_1)$ due to changes in X_1.

The further derivation of main and total effects is similar to the Sobol methodology and is presented in Liu et al. [35].

3.3. Morris

Additionally, the method by Morris [39] is applied in this paper. In this methodology the impact of changing one factor ($X_{1,\ldots,n}$) at a time is evaluated in turn. This impact (so called elementary effect of $x_{1,\ldots,n}$) is expressed as the gradient of the response surface, between the factor variations. The standard deviation (σ_{M}) of multiple elementary effects and the mean of the absolute of multiple elementary effects $\left(\mu^*_{\mathrm{M}}\right)$ are measures for factor sensitivity (an extension of the base Morris approach first proposed by Campolongo et al. [13]). A high value of σ_{M} means that the elementary effects relative to this factor are significantly different from each other. In contrast, a low σ_{M} indicates very similar values of the elementary effect implying that the effect of the factor is almost independent of the values taken by the other factors. In this paper, μ^*_{M} is used as indicator of the sensitivity. Campolongo et al. [12] show that μ^*_{M} is an excellent and cheap proxi of the variance based total effect, in that it tends to provide similar factor rankings. In particular it is very effective for screening purposes, being robust to Type II errors in highlighting irrelevant input factors.

3.4. Regionalised sensitivity analysis

The regionalised sensitivity analysis has been originally developed in the context of environmental models by Spear and Hornberger [57] and further developed by Beven and Binley [9]. In later application the cumulative distribution functions of sub-sets of the behavioural model space (all models which reproduce the evaluation data acceptably) are compared and the spread is an indication of the sensitivity. Spear and Hornberger [57] combine this with a Smirnov test, which allows a numerical specification into three sensitivity classes. In this original form, the method has many global as well as local properties (for more details follow the arguments of Saltelli et al. [53, p. 155]). This paper uses the total area of the spread as indicator of sensitivity.

3.5. Regression

Regression can be used to identify linear and monotonic patterns between distributions. The usage of regression in sensitivity analysis has been discussed by for example Kleijnen and Helton [31] and indirectly by Pappenberger et al. [46]. In this paper, we compute the Spearman rank correlation coefficient between the model factor and the model response.

3.6. Uncertainty in sensitivity analysis

All sensitivity analysis methods are subject to uncertainty as they are estimated on a limited sample. Several of the methodologies, which are used in this paper have methods to estimate the uncertainties (for Sobol see for example [52]). However, this paper tries to compare these

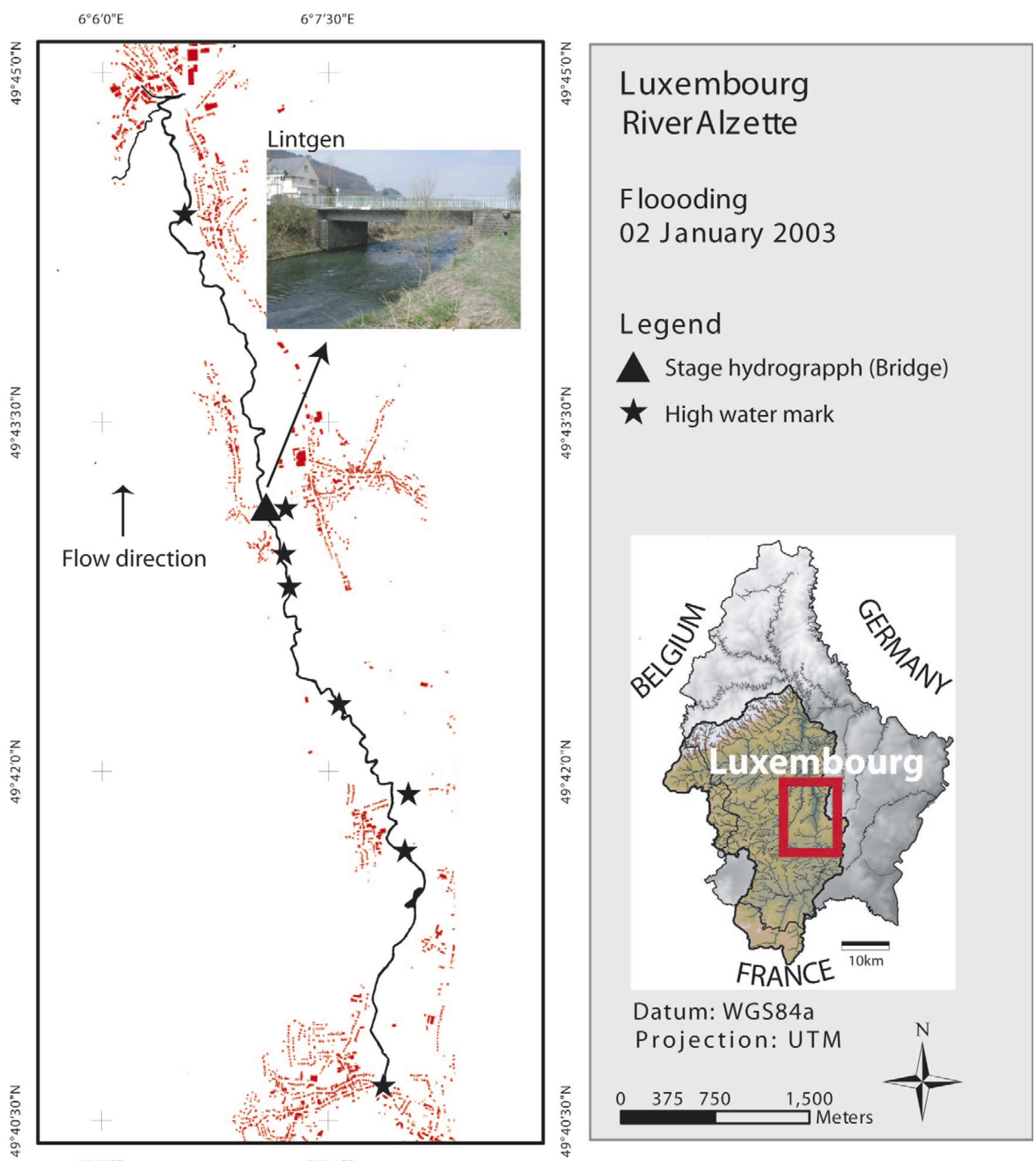

Fig. 1. Description of the Alzette catchment in Luxembourg. The location of the stage recorder (bridge Lintgen) and the location of the high water are marked.

measures and thus a common approach to estimate the uncertainty in sensitivity analysis has been applied via bootstrap sampling. For this, each measure has been computed 1000 times on a random sub-set of 80% of the sample. The size of this bootstrap sample can have an effect on the further analysis but tests with different percentages revealed similar patterns. The percentiles of this Monte Carlo analysis can be used to approximate the uncertainty, although, only a complete sample of the response surface (or of the representative features) will present the true sensitivity for each measure.

3.7. Model and region

The reach is on the River Alzette (Luxembourg), which is approximately 10 km long and in which the river meanders gently across a floodplain that ranges from approximately 250 m to 1 km in width (see Fig. 1 and [47] for more details). Cross-sectional surveys which include information on both channel width and depth were available at 74 locations along the reach. Upstream stage measurements are routinely recorded and allow the estimation of approximate upstream hydrographs via rating curves. A medium scale (1 in 5 years) flood event took place in January 2003. Discharge at Lintgen peaked at around 67 $m^3 s^{-1}$.

A one-dimensional unsteady flood inundation model (HEC-RAS) has been used to model this region, based on 74 cross-sections. The model has been simplified and only two different 'roughnesses' are used to represent the floodplain and the channel. Any structures, such as bridges (see [46,47]), within the flow domain have been ignored for simplicity of argument in this paper.

The uncertainty analysis in this paper is embedded into the generalised likelihood uncertainty estimation framework, which consists of running repeated simulations of a model using a range of values for each, uncertain, input parameter and evaluating each simulation. The models are then classified into behavioural and non-behavioural and a weighting of each behavioural simulation computed by rescaling the performance measure. The sensitivity analysis was then carried out only on the behavioural set, since these are the models of greatest interest. The division into behavioural and non-behavioural reduces the initial sample.

The uncertainty in the upstream boundary (inflow) has been derived from an uncertainty analysis from which 20 representative rating curves have been chosen (see [47]). The uncertainty for the downstream boundary could not be implemented in the same way as the upstream boundary due to a lack of data. Therefore, a looped rating curve which uses the simplified form of the momentum and Manning's equation has been applied. The friction slope is derived at each time step between the two most downstream cross-sections. The uncertainty in this boundary condition has been evaluated by applying additional roughness values to the last two cross-section, which are varied independently from all other cross-sections. Table 1 lists the factors which are used as calibration factors in each model implementation. All factors have to be understood as effective factors which compensate for model deficiencies in the model structure and input data. For example bed roughness is a resistance coefficient that accounts for bed shear, form drag by vegetation and obstacles as well as compensating for the one-dimensional model structure, which approximates the real three-dimensional phenomena, and errors in the representation of upstream inputs and floodplain topography and infrastructure. The factors considered in the SA exclude topography, which can be an important factor in obtaining good predictions of inundation [61] but which has been fixed here for all model runs.

GSA should be always performed on conditioned model simulations (the behavioural set of models in this study). The model is an imperfect representation of reality and needs to be constrained on evaluation data where suitable data are available. Several data sets are available for the Alzette catchment to evaluate model performance from which a sub-set has been chosen here to illustrate the sensitivity issues (Table 2). Most of the evaluation methods

Table 1
Model factors and sampling range of the factors in this study after [47]

Factors	Sampling range
Input rating curve	
Representative	1–20 (integer)
Main model region	
Manning roughness channel	0.01–0.1
Manning roughness floodplain (always higher than channel roughness)	0.01–0.2
Theta (weighting parameter of the numerical scheme)	0.6–1.0
Downstream boundary	
Manning roughness (channel and floodplain)	0.01–0.1
Initial slope at downstream boundary	0.0005–0.05

All factors have been sampled from a uniform distribution and have to be understood as effective factors.

Table 2
Evaluation data and method of Alzette catchment after [47]

Name	Brief description	Evaluation method
Maximum water level	Measurements of eight maximum water levels	Fuzzy membership function at each of the eight locations
Hydrographs	Stage hydrographs measured at the bridge location in Lintgen	Nash–Sutcliffe efficiency, mean absolute errors (see e.g. [42])
Travel time	Travel time of flood peak from upstream to downstream boundary	Fuzzy membership function

The table shows the name of each performance measure with a brief description. The evaluation method are described in more detail in the text and in [48].

are based on fuzzy membership functions. Fuzzy membership functions can often be described by the same number of parameters as for example variance based methods (assuming one uses a cut-off to restrict values towards infinity). However, fuzzy memberships have the advantage to define areas of constant performance and have more flexibility in shapes. The exact evaluation procedure has been described in detail by Pappenberger et al. [47].

The following description of the measures is taken from Pappenberger et al. [47].

3.7.1. Maximum water level

Eight different measurements of the maximum water level are available for this model region, which again have been evaluated by fuzzy membership functions. Any error below 0.5 m has been given the membership value of one and any error of more than 1 m has been given a value of zero.

3.7.2. Hydrographs (water level)

The hydrographs have been evaluated with two different functions: Nash–Sutcliffe efficiency [40] and mean absolute errors at the location of the stage recorder at Lintgen.

3.7.3. Travel time

The peak time evaluation has been evaluated similar to the maximum water levels. Any errors greater than 1.5 h have been given a membership value of zero and errors smaller than 0.5 h a membership value of 1. These values have been chosen to reflect the incommensurability error and the precision of the measurements.

The shape and range of the fuzzy performance measures has been chosen to reflect the measurement and commensurability error. The reader is referred to Pappenberger et al. [47] for a more detailed discussion. The study by Pappenberger et al. [47] demonstrated that the model is capable of reproducing the performance measures used individually.

4. Results

The results have been separated into three parts. In the first part, the difference between the various sensitivity measures is illustrated and discussed. In the second part, the impact of uncertainty on these sensitivity measures is shown and the last section highlights the management decisions which can be derived from a combined sensitivity index. This type of SA produces a large number of sets of results (no. of sensitivity methods × no. of evaluation methods × no. of uncertainty runs) and the discussion is consequently restricted to the most important examples.

4.1. Influence of the sensitivity method

In Fig. 2, we compare different sensitivity measures to demonstrate a change in ranking of importance in this flood inundation example. The Sobol index and the Kullback–Leibler entropy based index are compared for the hydrograph at Lintgen (mean absolute error measure). The absolute value of each analysis has been converted into ranks for comparison purposes. On the horizontal axis the ranks for the entropy measure are shown and on the vertical the ranks for the Sobol methodology are displayed. The shaded rectangles (in five categories) represent the percentage by which these ranks overlap for each bootstrap sample. Taking the top-left hand sub-plot (input magnitude) as an example >75–100% of all implementations achieved a rank of importance of 2 for both measures. For the outflow roughness (bottom left sub-plot) >25–50% of simulation have achieved an importance rank of 4 for the entropy and an importance rank of 3 for the Sobol measure. Each column and row also displays the total percentage out of the total number of simulations that achieved that rank.

Overall, Fig. 2 suggests that the input magnitude is the second most important factor for this evaluation measure at Lintgen. The 'channel roughess' exhibits a less clear picture with no overlapping percentage over 25%, but a large percentage of runs being ranked as 1 with the Sobol index. Although, 31% are also ranked 1 with the entropy measure, the method also leads to 29% of the bootstrap simulation being ranked as 5. The entropy measure does include some interaction effect (in 'Sobol' definition terms'), which can explain this discrepancy. A similarly incoherent picture is displayed for the floodplain roughness which is ranked at 4 with 50% by the Sobol method but has a less peaked distribution of ranks when the entropy measure is used. 'Theta' has an overall large percentage of simulations in the lower ranks, but does follow the previous two factors in respect of the results of the ranking of the entropy distribution. The outflow roughness factor is ranked 3 and upwards for both methods and the 'initial slope' is at ranks 4 and 5 for both methods.

This type of figure is useful for understanding which factors are less or more important, however, it also clearly shows that simulations can be ranked quite differently when the different methods are applied. This problem is complicated by the fact that it is possible to use different performance measures for the evaluation of this hydrograph.

In Fig. 3, the Nash–Sutcliffe measure is used and a ranking matrix displayed. The rank distributions in this plot have much clearer peaks (as expected for a Nash–Sutcliffe measure based on the sum of squared errors in comparison to an absolute error criteria), but do not always agree. The entropy method indicates that the 'input magnitude' is the most important factor in contrast to Fig. 2. An interesting difference from Fig. 2 is seen in the results for the 'channel roughness', which is now considered to be most likely only on scoring rank 5 for the Sobol measure (compared to rank 1 in Fig. 2). The entropy measure clearly indicates a higher importance of

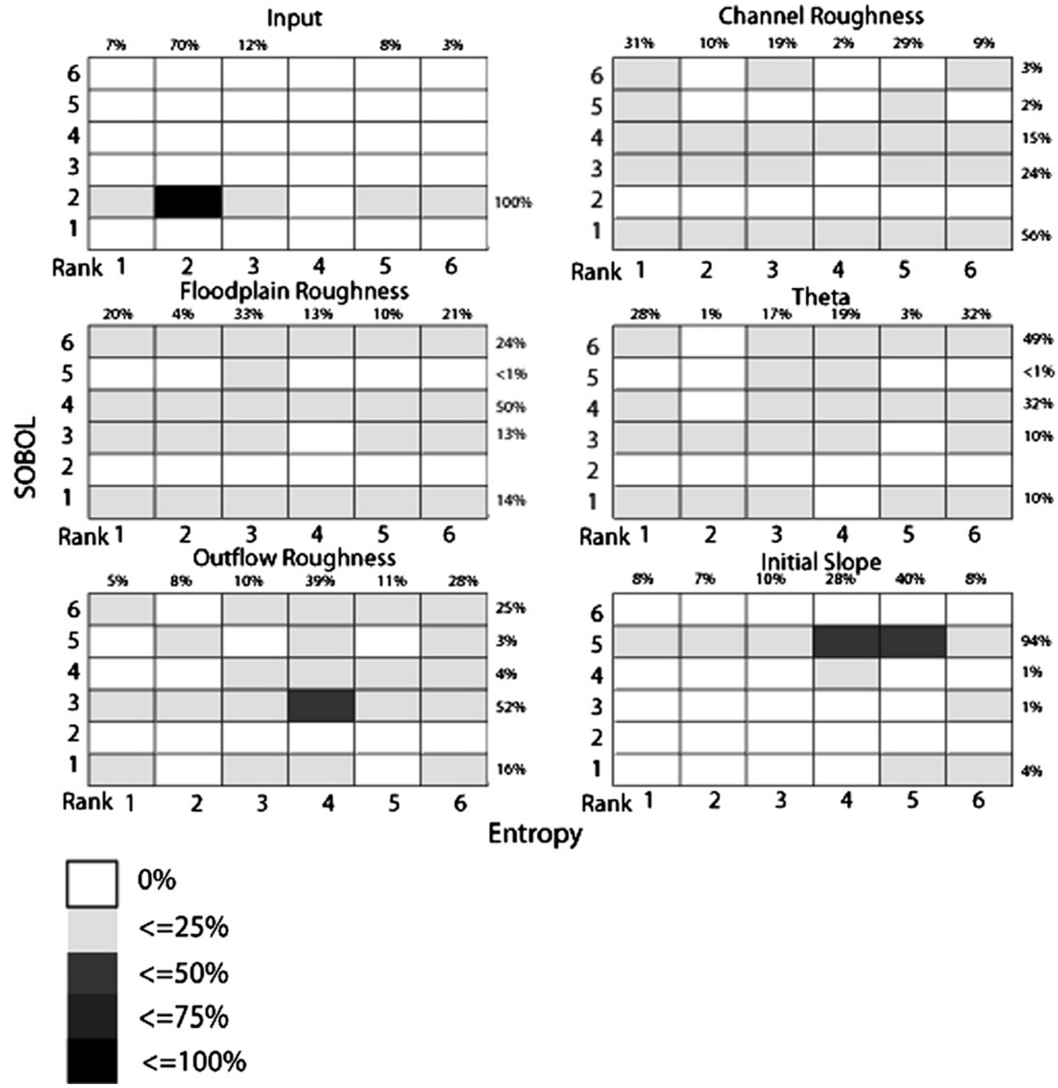

Fig. 2. Comparison of the Sobol index and the entropy based index for the hydrograph at Lintgen (absolute error measure). On the horizontal the ranks for the entropy measure are shown and on the vertical the ranks for the Sobol methodology are displayed. The shaded rectangles (in five categories) represent the percentage by which these ranks overlap for each bootstrap sample. It can be seen that the ranks achieved with both methods not always agree.

this parameter (ranks 2 or 3). The importance of ‘floodplain roughness’ has not changed significantly from the previous graph. However ‘theta’ is seen as a very unimportant factor for the Sobol index as well as the entropy. ‘Outflow roughness’ is still predominantly unimportant but shows an additional peak at the higher ranks. A similar effect has been observed in one-dimensional models elsewhere [44] and is probably explained by a backwater effect. In fact such models should be classified as unbehavioural as they are not representative to the hydraulic understanding of this event (no such large backwater effect has been observed). Similar behaviour can be observed for the ‘initial slope’ factor. This illustrates that performance measures should not be used alone to classify models into behavioural or non-behavioural, but have to be supported by a closer analysis of the model results. In summary, the figure demonstrates the large dependency of the sensitivity analysis on the chosen performance measures.

4.2. *Effect of uncertainty in the sensitivity measure*

The previous example already highlights the considerable uncertainty in the estimation of any sensitivity. The uncertainty of the sensitivity measures evaluated in this paper is purely based on an incomplete representation of the response surface (bootstrap sample) and thus does not present the full uncertainty. The full sample obviously represents the best estimate of the response surface, however, the bootstrap approach chosen in this paper takes account of that by randomly choosing a sub-set of the samples defining the response surface. This is further illustrated

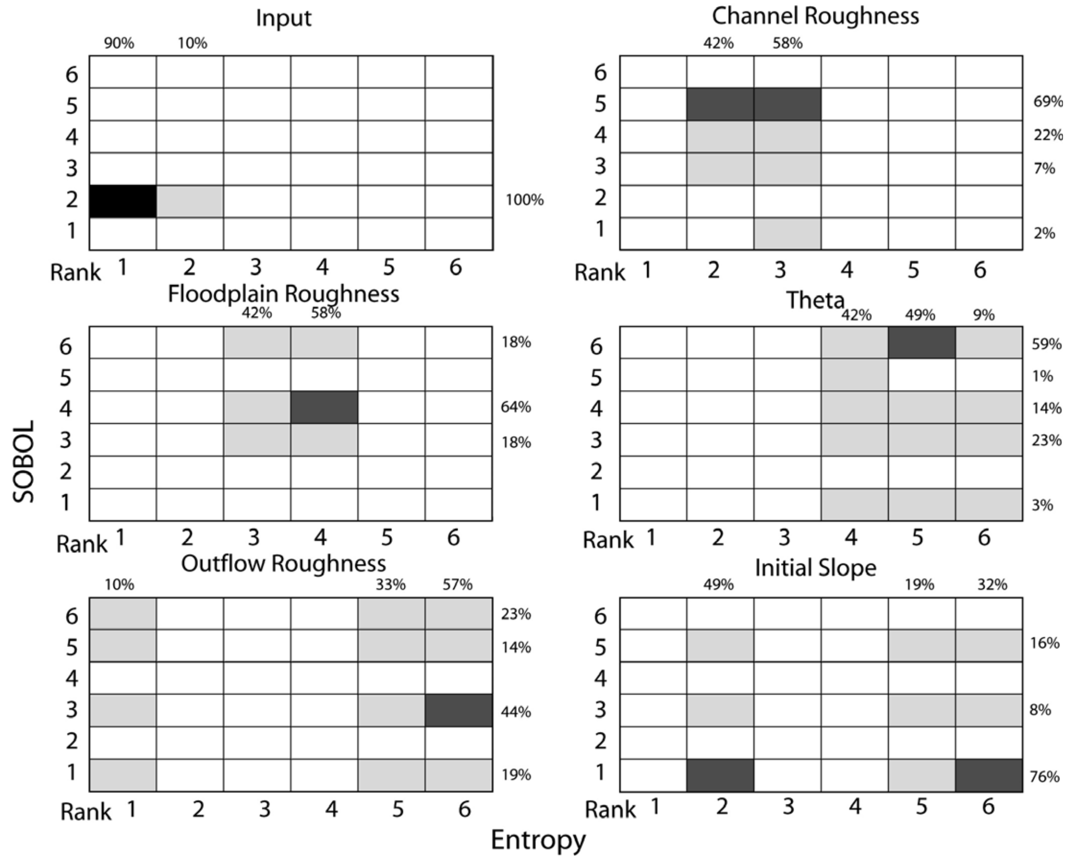

Fig. 3. Comparison of the Sobol index and the entropy based index for the hydrograph at Lintgen (Nash–Sutcliffe measure). On the horizontal the ranks for the entropy measure are shown and on the vertical the ranks for the Sobol methodology are displayed. The shaded rectangles (in five categories) represent the percentage by which these ranks overlap for each bootstrap sample. It can be seen that the ranks achieved with both methods not always agree and are different from Fig. 2.

in Fig. 4, which plots the range of Spearman rank correlations computed at each sub-sample (at the Lintgen station). The correlation is computed between the Nash–Sutcliffe measure and the parameters.

As in the previous example, the computed values can vary widely and indicate that values can be significantly correlated – or not. Depending on the sample efficiency and non-linearity of the response surface the variations can be larger or smaller. Best practice dictates that one should keep sampling until a stable sensitivity value is reached (see [44]), however, due to a lack of resources such as computer power this may always not be possible. This highlights that it is essential to indicate the uncertainty of the sensitivity measure.

Plot 5 is another additional example of how this uncertainty can be quantified. In this plot, the results of the Morris sensitivity analysis for the travel time are displayed. A high value of σ_M means that the elementary effect relative to this factor is significantly altered by changing the base point for computing elementary effects in the input factors space. In contrast, a low σ_M indicates very similar values of the elementary effect implying that the effect of the factor is almost independent of the values taken by the other fac-

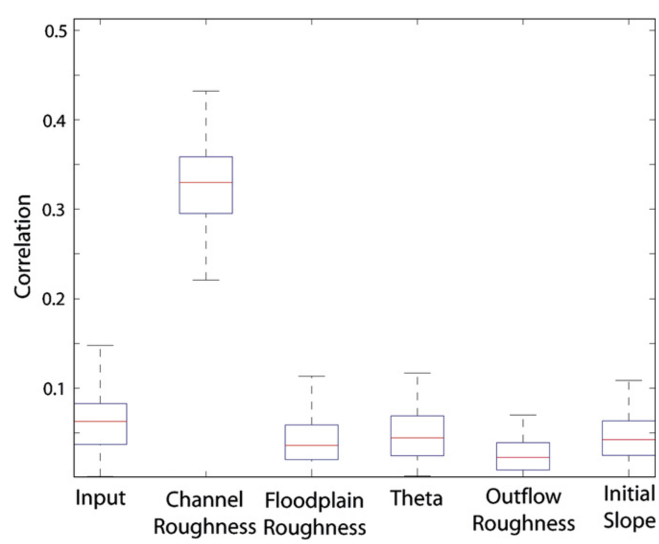

Fig. 4. Box and whisker plot of the Spearman rank correlation coefficients at Lintgen (Nash–Sutcliffe measure) computed for each sub-sample. The box has lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the box to show the extent of the rest of the data. The uncertainty ranges are very large, however, the channel roughness shows clearly to be the most important parameter (despite large uncertainties).

Table 3
Percentage of ranks of each factor after the regional sensitivity analysis method for the travel time evaluation criteria

Factor	Rank					
	1	2	3	4	5	6
Input (%)	5	90	5	0	0	0
Channel roughness (%)	95	5	0	0	0	0
Floodplain roughness (%)	0	0	0	93	0	7
Theta (%)	0	0	0	7	10	83
Outflow roughness (%)	0	0	0	0	90	10
Initial slope (%)	0	5	95	0	0	0

Input and channel roughness achieve the largest percentages at the ranks 1 and 2. The floodplain roughness is dominantly seen as the 4th most important factor. Theta and the outflow roughness are seen as the least important factors. The initial slope is predominantly ranked on 3rd place.

tors. μ_{M}^{*} is the average of absolute values of the elementary effect as explained earlier (see Section 3).

It is not necessarily the case that every method has large uncertainties associated with it, for example, the regional sensitivity analysis has a lower variance than all other methods (see Table 3). This is explained by the fact that a sub-sample does not change the cumulative distribution function of the parameters *vs.* the evaluation to a large extent. The sensitivity to uncertainty in the response surface can be seen as an advantage or disadvantage, depending on the application.

In Figs. 4 and 5 and Table 3, the channel roughness is the most important factor, but with significant uncertainties. The edges of the distributions of the uncertainty in the sensitivities can overlap and ranking can change. It must be pointed out that these findings do not contradict in principle, they are merely outcomes of different definitions of sensitivity and interacting effects. In fact, one can utilize the strength of each method to get a better understanding of the model and data. For example, Fig. 5 highlights the fact that all parameters have an effect on the output with more or less interaction (positive absolute elementary effect). This outcome is similar to the usage of multi-objective evaluation criteria in hydrological modelling [19] and can be embraced in a similar way (see next section on distribution of resources).

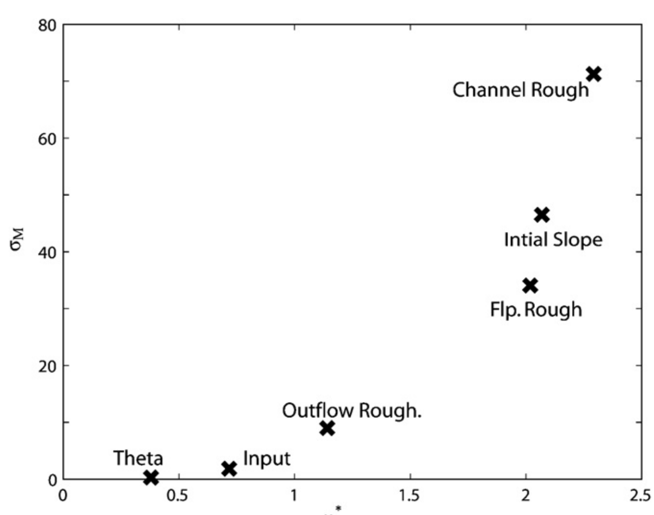

Fig. 5. Results of the Morris analysis of the travel time. This figure displays the mean of the absolute elementary effect (μ_{M}^{*}) and the standard deviation of multiple elementary effects (σ_{M}). The higher the absolute elementary effect the more important is the parameter. The absolute elementary effect is closely related to the total Sobol index.

4.3. Utilizing the variability in uncertain sensitivity rankings

All methods define sensitivity in different ways and thus can lead to different importance rankings. However, in many situations it may be still necessary to derive decisions from these distributions in order to concentrate research or resources on the most important parameters. There is a wealth of literature to include uncertainties (which can include the uncertainties of importance) into any decision making process. However, this is not the focus of this paper and we refer the reader to references such as Ben-Haim [4] or Yang et al. [62]. In this paper we simply demonstrate the possibility by computing an average sensitivity measure based on the different *ranks* over all the computed sensitivity analyses performed. We further assume that any average rank higher than 3.5 is too low to be considered in decisions about the distribution of any additional resources available. We explicitly ignore the possibility that the results of the various sensitivity measures might be correlated, since we wish similar rankings to be reinforced in the calculation of an overall measure.

In Fig. 6, this averaging (see section above) has been done for the maximum water levels. The most important parameters throughout the region for the maximum water levels are boundary conditions ('inflow magnitude' and 'manning outflow') and the 'channel roughness. At location 2, 'theta' is also of importance, which is probably explained by the fact that this location is subject to instabilities. The downstream boundary condition becomes important from location 4 onwards, although it would not be considered as important for future research efforts if the analysis concentrates on locations 5 and 7. Pappenberger et al. [47] had similar results for a sensitivity analysis along the cross-sections of the model for remotely sensed observations. However, the factor 'theta' did not show any importance, which is due to the different evaluation measures (presenting different hydrological regimes).

These pie charts would be an aid in the optimal prioritisation of resources for research or flood management to reduce uncertainty in model factors. One could for example select those parameters which are relevant for all stations or compute the average over all stations (see Table 4). Table 4 indicates that it is important to define clearly the aim of any further studies and that any global improvement due to increased understanding of any parameter, is not necessarily reflected locally.

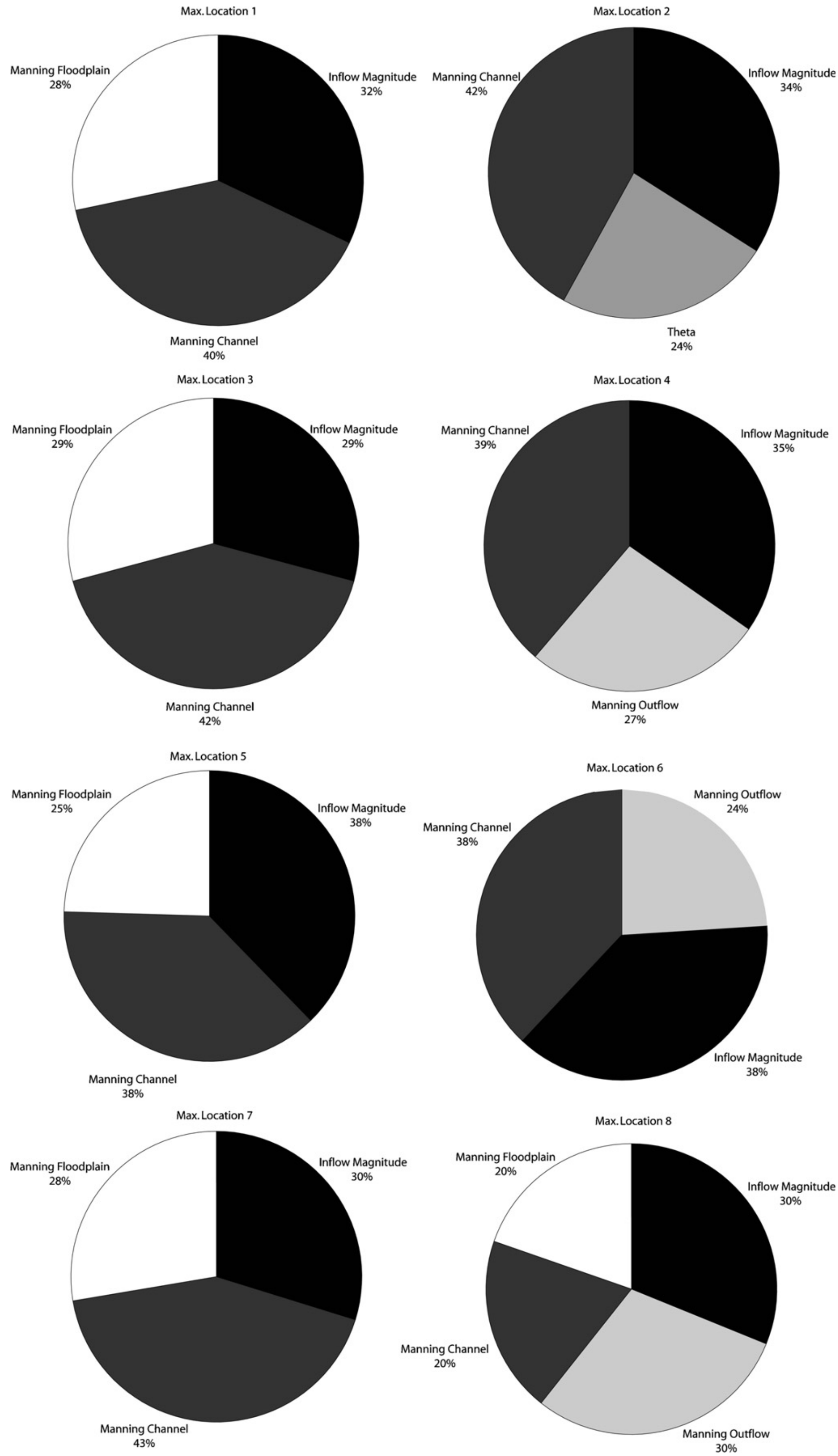

Fig. 6. Importance of the six factors along the locations of the maximum water mark. The higher the percentage the more impact will research and better knowledge of this parameter have to constrain the uncertainties in the model output.

Table 4
Importance of the six factors for all locations of the maximum water mark

Parameter	Manning floodplain	Inflow magnitude	Manning channel	Theta	Manning outflow
Importance	16.25	33.25	37.5	3	10

The higher the percentage the more impact will research and better knowledge of this parameter have to constrain the uncertainties in the model output.

5. Discussion

The main source of uncertainty in the determination of sensitivity is in the non-linearity of interactions between factors as reflected in the response surface for a given performance measure. Even with unrestricted computer power and a model with low CPU requirements, any model without a simple analytical solution will have some areas of unknown response. The question this raises is whether we really need such detailed model response investigations or if a relatively uncertain approximation would be sufficient. Some methods such as the computation of the Sobol main effect will be most likely insensitive to non-detected peaks in the surface (if the rest of the response surface is sampled adequately). However, flooding is very often affected by local phenomena such as overtopping of a dyke, which may affect only a small proportion of the factor space. The question is then whether the resources spent on getting a better estimate may be more efficiently invested in acquiring more data to improve the model representation. In any real application, both approaches to improve predictions have to be used together and iteratively. The collection of data leads to a model with uncertainties, the investigation of the sensitivities and uncertainties of that model should lead to a refined measurement campaign, which in turn leads to a new and improved model set-up. This is a form of the learning process about places described by Beven [8]. Unfortunately, in the case of predicting flood inundation, such an iterative cycle may be difficult to achieve as extreme events such as floods are rare.

Multi-method global sensitivity analysis can be used as exploratory tool. It gives an improved understanding of the model and an insight into which model parameters are important thus gives a deeper insight into the programmed structure. Moreover, it can be directly used to improve estimates of model uncertainties by sampling the important areas of the response surface more efficiently. For example the generalized likelihood uncertainty estimation framework [7] requires the identification of behavioural model simulations. There are two ways to classify models as behavioural or non-behavioural: either based on a model performance, which could include soft data [55] or by declaring certain model simulations as physically not realistic and inconsistent with our conceptual understanding of the system. Sometimes, such inconsistencies are not apparent and can only be revealed by a sensitivity analysis. For example, if a factor is revealed as being important, when it clearly should not be, then this would be inconsistent with our conceptual understanding even if, as the analyses presented here show, such inconsistencies may be difficult to quantify in the absence of adequate data. In our study the initial slope of the downstream boundary condition seems to have an unduly high influence on the travel time (Fig. 5) and suggests that further investigation into the implementation of this boundary condition should be performed.

6. Conclusion

GSA is a valuable tool in understanding flood inundation models and deriving decisions on strategies to reduce model uncertainty. In this paper, a sensitivity analysis of a one-dimensional flood inundation model (HEC-RAS) on the River Alzette has been presented. It is impossible to define sensitivity in a unique way and different methods can lead to a difference in ranking of importance of model factors. In this paper we applied five different methods (Sobol, K–L entropy, Morris, regionalised sensitivity analysis and regression) and compared the outcomes on selected examples. It could be shown that for example that the Sobol' method and the entropy based method can lead to different ranking of importance of the parameter factors. This is complicated by the fact that different evaluation measures can lead to a difference in ranking, too. This paper further investigates the uncertainty in the estimation of the sensitivity measures based on a bootstrap sample. It can be shown that there is considerable uncertainty in the determination of the sensitivity indices. The response of two of the referees of this paper was to suggest that it should be enhanced with a critical and rigorous analysis of the different sensitivity analysis methods to show the advantages and disadvantages of each method. Our results suggest, however, that such an analysis is not really possible. The different methods are simply different. They just reveal different things about the nature of the highly complex interactions in the model space. This is because the number of input factors considered is small and there is none of them which is insignificant, i.e. each of the input factors included in the analysis has *some* effect on the model behaviour, with considerable interactions between factors. Each measure considered in the paper weights differently various aspects of the interaction structure.

This has the following implications on some of the methodologies applied:

1. Sobol' main effects: these ignore interactions, so it has to be expected that its ranking might differ form any other measure considered here, in particular channel rough-

ness seems involved in strong interactions which are not apparent in the estimated main effects, using the Nash–Sutcliffe performance measure.
2. The main value of the Morris method in screening cannot be appreciated here, since there is nothing to screen in this model; while the actual ranking of the factors having *non-negligible* effect on the output (all factors here) is often quite uncertain.
3. Entropy measure seem the least robust, where by initial slope has similar probability of being ranked 2 or 6.
4. RSA is not particularly sensitive to uncertainty in the response surface and this is surely an advantage; however, the statistical power of the Smirnov statistics is not so strong in defining rankings among factors with similar Smirnov statistics (see e.g. [63], where the number of important factors identified by Smirnov is always smaller than those of other measures: "Overall ANOVA and Sobol's method were shown to be superior to RSA and PEST. Relative to one another, ANOVA has reduced computational requirements and Sobol's method yielded more robust sensitivity rankings").

There is an analogy here with similar arguments in the assessment of model uncertainties (e.g. [36,37,7,10,11]). If environmental models do not conform to the constraints of the assumptions of a formal analysis (as is the case with the flood inundation explored here), then how best do we assess the information content of data in testing sensitivities and estimate the uncertainties of the models we use.

Saltelli et al. [53, p. 49] have addressed this issue and the consequent argument that if different ordering of the factor importance arises through the application of different SA methodologies, why bother doing SA at all? They suggest that the discrepancies between methods are a result of having an ill-posed SA question. Refining the definition of the setting, they suggest, will resolve the discrepancy. They consider two settings: factor prioritisation and factor fixing. Factor prioritisation is defined by the question of which factors one should concentrate on first (e.g. by getting a better understanding of the uncertainty ranges) in order to reduce the uncertainty in the output most. In factor fixing the aim is to determine, which factor can be fixed without affecting the model output significantly (for more details see [50]). Saltelli and Tarantola [54] advocate measures of the main effect for the first setting and measures of the total effect for the second.

Within the setting of the equifinality thesis, however, it is evident that this is a simplification of the sensitivity issue. The possibility of necessarily recognising multiple acceptable models also implies that local sensitivities might determine which combinations of parameter values and other factors produce acceptable results, and that these local sensitivities and factor interactions might vary through the model space. This is one reason why different methods produce different sensitivity rankings of the factors. In essence, neither local sensitivities alone, nor global sensitivities can totally reflect the non-stationary complexities of the interactions that produce acceptable models. In real applications, the different methods can only be a guide, to be used with care and circumspection. Here, a simple methodology to incorporate the uncertainty in the sensitivity evaluations in a decision making process is presented. While, it will not generally be possible to give a unique answer to the specification of sensitivities and we still have to explain the distribution of factor sensitivities to the end-users of our model. Such a framework should ideally be embedded into a code of practice for modelling (see [43] for an example).

Acknowledgements

We thank Georges Müller of the Service de la Gestion de l'Eau for providing some of the data used in this study. Comments by Hannah Cloke (King's College, London, UK) and Tobias Krüger (Lancaster University, UK) improved this paper. Moreover, Paul Bates (Bristol University) and three anonymous reviewers gave very valuable comments which greatly improved the manuscript. This research was performed as part of a multi-disciplinary programme undertaken by the Flood Risk Management Research Consortium (http://www.floodrisk.org.uk) in Research Priority Area 9 (Risk and Uncertainty). The Consortium is funded by the UK Engineering and Physical Sciences Research Council under grant GR/S76304/01, jointly with NERC, the Joint Defra/EA Flood and Coastal Erosion Risk Management R& D programme, the Scottish Executive, the Rivers Agency (Northern Ireland) and UK Water Industry Research. Development of extensions to the GLUE methodology has been supported by NERC grant NER/L/S/2001/00658. Florian Pappenberger was further supported by the PREVIEW program (FP6).

References

[1] Aronica G, Bates PD, Horritt MS. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. Hydrol Process 2002;16(10):2001–16.
[2] Baker S, Ponniah D, Smith S. Survey of risk management in major UK companies. J Profess Issues Eng Educ Pract 1999;125(3):94–102.
[3] Bates PD, Horritt MS, Aronica G, Beven K. Bayesian updating of flood inundation likelihoods conditioned on flood extent data. Hydrol Process 2004;18(17):3347–70.
[4] Ben-Haim Y. Information-gap decision theory: decisions under severe uncertainty. Series on decision and risk, vol. x. San Diego, CA/London: Academic Press; 2001. 330 pp.
[5] Beven KJ. Prophecy, reality and uncertainty in distributed hydrological modelling. Adv Water Resour 1993;16(1):41–51.
[6] Beven KJ. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrol Process 2002;16(2):189–206.
[7] Beven KJ. A manifesto for the equifinality thesis. J Hydrol 2006;320(1–2):18–36.

[8] Beven KJ. Working towards integrated environmental models of everywhere: uncertainty, data, and modelling as a learning process. Hydrol Earth Syst Sci 2007;11(1):460–7.
[9] Beven KJ, Binley A. The future of distributed models – model calibration and uncertainty prediction. Hydrol Process 1992;6(3):279–98.
[10] Beven KJ, Smith PJ, Freer JE. Comment on ‘‘Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology’’ by Peitro Mantovan and Ezio Todini. J Hydrol 2007. doi:10.1016/j.jhydrol.2007.02.023.
[11] Beven K, Smith P, Freer J. Comment on ‘‘Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology’’ by Pietro Mantovan and Ezio Todini. J of Hydrol 2007;338(3–4):315–8.
[12] Campolongo F, Cariboni J, Saltelli A. An effective screening design for sensitivity analysis of large models. Environ Model Softw 2007;22(10):1509–18.
[13] Campolongo F, Tarantola S, Saltelli A. Tackling quantitatively large dimensionality problems. Comput Phys Commun 1999;117(1–2):75–85.
[14] Coyle D, Buxton MJ, O’Brien BJ. Measures of importance for economic analysis based on decision modeling. J Clin Epidemiol 2003;56(10):989–97.
[15] Crosetto M, Tarantola S. Uncertainty and sensitivity analysis: tools for GIS based model implementation. J Geogr Inform Syst 2001;15:415–37.
[16] Cullen AC, Frey HC. Probabilistic techniques in exposure assessment: a handbook for dealing with variability and uncertainty in models and inputs, vol. xvi. New York/London: Plenum Press; 1999. 335 pp.
[17] Fieberg J, Jenkins KJ. Assessing uncertainty in ecological systems using global sensitivity analyses: a case example of simulated wolf reintroduction effects on elk. Ecol Model 2005;187(2–3):259.
[18] Fraedrich D, Goldberg A. A methodological framework for the validation of predictive simulations. Eur J Oper Res 2000;124(1):55–62.
[19] Freer J, Beven KJ, Peters N. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: Duan QY, Gupta H, Sorooshian S, Rousseau A, Turcotte R, editors. Calibration of watershed models. Washington: American Geophysical Union; 2003. p. 69–88.
[20] Frey HC, Patil SR. Identification and review of sensitivity analysis methods. Risk Anal 2002;22(3):553–78.
[21] Hall JW, Tarantola S, Bates PD, Horritt MS. Distributed sensitivity analysis of flood inundation model calibration. J Hydraul Eng-ASCE 2005;131(2):117–26.
[22] Helton JC, Cooke RM, McKay MD, Saltelli A. Sensitivity of model output: SAMO 2004 – editorial of special issue. Reliab Eng Syst Safety 2006;91:1105–8.
[23] Helton JC, Davis FJ. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliab Eng Syst Safety 2003;81(1):23–69.
[24] Horritt MS. A methodology for the validation of uncertain flood inundation models. J Hydrol 2006;326(1–4):153–65.
[25] Hunter NM. Flood inundation modelling. PhD thesis. Bristol: School of Geography, Bristol University; 2006.
[26] Hunter NM, Bates PD, Horritt MS, De Roo PJ, Werner MGF. Utility of different data types for calibrating flood inundation models within a GLUE framework. Hydrol Earth Syst Sci 2005;9(4):412–30.
[27] Iman RL, Helton JC. An investigation of uncertainty and sensitivity analysis techniques for computer-models. Risk Anal 1988;8(1):71–90.
[28] Ionescu-Bujor M, Cacuci DG. A comparative review of sensitivity and uncertainty analysis of large-scale systems. I: Deterministic methods. Nucl Sci Eng 2004;147(3):189–203.
[29] Jones RN. Analysing the risk of climate change using an irrigation demand model. Climate Res 2000;14(2):89–100.
[30] Kleijnen JPC. Statistical validation of simulation-models. Eur J Oper Res 1995;87(1):21–34.
[31] Kleijnen JPC, Helton JC. Statistical analyses of scatterplots to identify important factors in large-scale simulations. 2: Robustness of techniques. Reliab Eng Syst Safety 1999;65(2):187–97.
[32] Knight DW, Shiono K. Channel and floodplain hydraulics. In: Anderson MG, Walling DE, Bates PD, editors. Floodplain processes. New York: John Wiley and Sons; 1996.
[33] Krykacz-Hausmann B. Epistemic sensitivity analysis based on the concept of entropy. In: Prado P, Bolado R, editors. Proceedings of SAMO2001; 2001. p. 31–5.
[34] Kullback S, Leibler RA. One information and sufficiency. Ann Math Stat 1951;22(1):79–86.
[35] Liu H, Chen W, Sudjianto A. Relative entropy based method for global and regional sensitivity analysis in probabilistic design. In: Proceedings of DETC’04 ASME 2004 international design engineering technical conferences and computers and information in engineering conference; 2004.
[36] Mantovan P, Todini E. Hydrological forecasting uncertainty estimation: Incoherence of the GLUE methodology. J Hydrol 2006;330:368–81.
[37] Mantovan P, Todini E, Martina MLV. Reply to comment by Keith Beven, Paul Smith and Jim Freer on ‘‘Hydrological forecasting uncertainty assessment: incoherence of the GLUE methodology’’. J of Hydrol 2007;338(3–4):319–24.
[38] Mokhtari A, Frey HC. Recommended practice regarding selection of sensitivity analysis methods applied to microbial food safety process risk models. Hum Ecol Risk Assess 2005;11(3):591–605.
[39] Morris MD. Factorial sampling plans for preliminary computational experiments. Technometrics 1991;33(2):161–74.
[40] Nash JE, Sutcliffe JV. River flow forecasting through conceptual models. Part 1: A discussion of principles. J Hydrol 1970;10:282–90.
[41] Pappenberger F. Uncertainty in flood inundation models. Lancaster: Lancaster University; 2004. 325 pp.
[42] Pappenberger F, Beven K. Functional classification and evaluation of hydrographs based on multicomponent mapping. Int J River Basin Manage 2004;2(2):89–100.
[43] Pappenberger F, Beven KJ. Ignorance is bliss – or 7 reasons not to use uncertainty analysis. Water Resour Res 2006;42(5). doi:10.1029/2005WR00482.
[44] Pappenberger F, Beven KJ, Horritt MS, Blazkova S. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. J Hydrol 2005;302:46–69.
[45] Pappenberger F, et al. Risk and uncertainty – tools and implementation – flood risk management research consortium. Lancaster: University Lancaster; 2005. www.floodrisk.org.uk and www.flood-risk.net.
[46] Pappenberger F, Iorgulescu I, Beven KJ. Sensitivity analysis based on regional splits (SARS-RT). Environ Model Softw 2006;21(7):976–90.
[47] Pappenberger F et al. Influence of uncertain boundary conditions and model structure on flood inundation predictions. Adv Water Resour 2006;29(10):1430–49.
[48] Phillips A, Janies D, Wheeler W. Multiple sequence alignment in phylogenetic analysis. Mol Phylogenet Evol 2000;16(3):317–30.
[49] Ratto M, Tarantola S, Saltelli A. Sensitivity analysis in model calibration: GSA-GLUE approach. Comput Phys Commun 2001;136(3):212–24.
[50] Ratto M, Young PC, Romanowicz R, Pappenberger F, Saltelli A, Pagano A. Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology. Hydrol Earth Syst Sci 2007;11:1249–66.
[51] Romanowicz R, Beven KJ, Tawn J. Evaluation of predictive uncertainty in nonlinear hydrological models using a Bayesian approach. In: Barnett VT, Turkman KF, editors. Statistics for the environment. II. Water reslated issues. Chichester: John Wiley and Sons; 1994. p. 297–318.
[52] Saltelli A, Sobol IM. About the use of rank transformation in sensitivity analysis of model output. Reliab Eng Syst Safety 1995;50(3):225–39.

[53] Saltelli A, Tarantola A, Campolongo F, Ratto M. Sensitivity analysis in practice – a guide to assessing scientific models. Chichester: John Wiley and Sons; 2004.
[54] Saltelli A, Tarantola S. On the relative importance of input factors in mathematical models: safety assessment for nuclear waste disposal. J Am Stat Assoc 2002;97(459):702–9.
[55] Seibert J, McDonnell JJ. On the dialog between experimentalist and modeler in catchment hydrology: use of soft data for multicriteria model calibration. Water Resour Res 2002;38(11) [art. no. 1241].
[56] Sobol IM. Sensitivity estimates for nonlinear mathematical models. Math Model Comput Exp 1993;1:407–14.
[57] Spear RC, Hornberger GM. Eutrophication in peel inlet. II. Identification of critical uncertainties via generalised sensitivity analysis. Water Resour Res 1980;14:43–9.
[58] Wagener T, McIntyre N, Lees MJ, Wheater HS, Gupta HV. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrol Process 2003;17(2):455–76.
[59] Ward DL, Schultz AA, Matson PG. Differences in swimming ability and behavior in response to high water velocities among native and nonnative fishes. Environ Biol Fishes 2003;68(1):87–92.
[60] Wilson MD. Evaluating the effect of data and data uncertainty on predictions of flood inundation. Southampton: University of Southampton; 2004. 276 pp.
[61] Wilson MD, Atkinson PM. Prediction uncertainty in elevation and its effect on flood inundation modelling. In: Atkinson PM, Foody GM, Darby S, Wu F, editors. GeoDynamics. John Wiley and Sons Ltd.; 2005. http://www.geocomputation.org/2003/Papers/Wilson_1_Paper.pdf.
[62] Yang JB, Wang YM, Xu DL, Chin KS. The evidential reasoning approach for MADA under both probabilistic and fuzzy uncertainties. Eur J Oper Res 2006;171(1):309–43.
[63] Tang Y, Reed P, Wagener T, van Werkhoven K. Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation. Hydrol Earth Syst Sci 2007;11:793–817.