Available online at www.sciencedirect.com
SCIENCE DIRECT
Environmental Modelling & Software 21 (2006) 976–990
Environmental Modelling & Software
www.elsevier.com/locate/envsoft

Sensitivity analysis based on regional splits and regression trees (SARS-RT)

Florian Pappenberger [a,*], Ion Iorgulescu [b], Keith J. Beven [a]

[a] *Environmental Science, Lancaster University, Lancaster LA1 4YQ, United Kingdom*
[b] *Ecole Polytechnique Fédérale de Lausanne, Switzerland*

Received 12 October 2004; received in revised form 15 January 2005; accepted 7 April 2005
Available online 10 August 2005

Abstract

A global sensitivity analysis with regional properties is introduced. This method is demonstrated on two synthetic and one hydraulic example. It can be shown that an uncertainty analysis based on one-dimensional scatter plots and correlation analyses such as the Spearman Rank Correlation coefficient can lead to misinterpretations of any model results. The method which has been proposed in this paper is based on multiple regression trees (so called Random Forests). The splits at each node of the regression tree are sampled from a probability distribution. Several criteria are enforced at each level of splitting to ensure positive information gain and also to distinguish between behavioural and non-behavioural model representations. The latter distinction is applied in the generalized likelihood uncertainty estimation (GLUE) and regional sensitivity analysis (RSA) framework to analyse model results and is used here to derive regression tree (model) structures. Two methods of sensitivity analysis are used: in the first method the total information gain achieved by each parameter is evaluated. In the second method parameters and parameter sets are permuted and an error rate computed. This error rate is compared to values without permutation. This latter method allows the evaluation of the sensitivity of parameter combinations and thus gives an insight into the structure of the response surface. The examples demonstrate the capability of this methodology and stress the importance of the application of sensitivity analysis.

Keywords: Regression tree; Sensitivity analysis; Random Forests; Uncertainty analysis; Calibration; Generalized likelihood uncertainty estimation; Regional sensitivity analysis

1. Introduction

Most hydrological models need some sort of calibration because inputs, such as parameters, model structure, boundary conditions and others, are uncertain (see for example and references therein Beven and Pappenberger, 2003; Duan et al., 1992; Pappenberger, 2004; Pastres and Ciavatta, 2005; Seibert and McDonnell, 2003). Thus it is not surprising that the attention of hydrological modellers has recently focused on the development of uncertainty frameworks such as GLUE (Beven and Binley, 1992), the Pareto Optimal set approach (Gupta et al., 1998; Yapo et al., 1996; Yapo et al., 1998), the Bayesian Forecasting System (Georgakakos and Krzysztofowicz, 2001; Krzysztofowicz, 2001, 2002a,b; Krzysztofowicz and Herr, 2001), NLFIT (Kuczera and Parent, 1998), PEST (Doherty, 2002), BATEA (Kavetski et al., 2003) and many more. Moreover, many decision support systems to accompany these developments have been proposed (Jolma and Norton, 2005; Pallottino et al., 2005; Reichert and Borsuk, 2005). Most of these approaches concentrate on the results of parameter uncertainty. However, many

* Corresponding author. Tel.: +44 1524 65201; fax: +44 1524 593985.
E-mail address: f.pappenberger@lancaster.ac.uk (F. Pappenberger).

1364-8152/$ - see front matter
doi:10.1016/j.envsoft.2005.04.010

methods neglect an investigation of parameter sensitivity or analyse only one-dimensional relationships, although a better understanding of parameter sensitivity could be beneficial to any uncertainty analysis (Oakley and O'Hagan, 2004).

It is well known that model results will generally be more sensitive to a change in one parameter than another (Melching, 1995) and this is at the heart of most optimisation strategies. The relationships can be especially complicated for complex models that have large numbers of parameters which interact in a highly non-linear way. Many traditional sensitivity analyses are based on the estimation of the local slope (McCuen, 1973a,b), which is computed around the best fitting parameter set (or some arbitrary "best criteria" point in parameter space prior to optimisation). These methods make the assumption that an optimum parameter set exists. However, this can be questioned, considering all the uncertainties of the modelling process (Beven, 2002, in press; Oakley and O'Hagan, 2004).

To avoid evaluation of only a local sensitivity, some form of global sensitivity method is needed (Chan et al., 1997). Very good summaries of various global methods are given by Saltelli et al. (2004), Helton and Davis (2003) and Oakley and O'Hagan (2004). Ratto et al. (2001) have demonstrated how variance methods like the Sobol indices, FAST, correlation matrix or principal component analysis (PCA) can be used to achieve a better understanding of parameter sensitivity. These methods allow for the investigation of higher order parameter interaction. Other sensitivity analysis methods include a differential analysis in which a Taylor series is fitted to the response surface and the variance is derived from it (see Helton and Davis, 2003 and references therein). However, the equations can become very rapidly complicated and the complexity of parameter interaction has to be predefined within this framework.

Regional sensitivity analysis (RSA, Hornberger and Spear, 1981; Spear and Hornberger, 1980) is an alternative approach which can be performed by making no assumptions about the shape of the response surface or covariation of parameters. The distinction between behavioural and non-behavioural model results is an essential part of the RSA and also the GLUE methodology. The behavioural set is the class of models for which it is not possible to disregard them as useful predictors of the system. For the sensitivity analysis, the cumulative distributions of subsets of the behavioural parameter space are investigated. This implicitly reflects local parameter interaction that gives rise to the behavioural sets of parameter taken. Significant differences mean a higher sensitivity towards the parameter which is investigated. Within the GLUE methodology, which is based on the RSA, the performance measure is plotted against one parameter and cumulative distribution functions are derived based on the likelihood weights. This idea has been further extended by Wagener et al. (2003) in their dynamic identifiability analysis (DYNIA) in which the slope of the cumulative distribution function is used as an indicator for parameter identifiability (and thus indirectly for parameter sensitivity). Unfortunately, it is not possible to explore multi-dimensional relationships explicitly with these concepts.

An extension to deal with high dimensional parameter spaces is the Tree Structured Density Estimation (TSDE) of Spear et al. (1994), which is employed to refine the sampling and promotes a more detailed investigation of parameter sensitivity and interaction. It is based on a Regression Tree method and computes sensitivity based on point density of behavioural models within hypercubes. Therefore, it is a theory that is aimed more for re-sampling purposes than for sensitivity analysis, although dominant behavioural modes can be extracted through the interpretation of the splits (Grieb et al., 1999; Iorgulescu and Beven, 2004; Spear et al., 1994; Spear and Hornberger, 1980). This methodology further focuses only on one optimised regression tree, although others may be considered as equally good predictors of the system (concept of equifinality). This concept has been developed for parameter sets (Beven and Binley, 1992), but does apply also to the model structure itself. The disregard of multiple tree structures has the effect of neglecting other possible interpretations. However, TSDE has the advantage of not being restricted to a certain sampling strategy such as the SOBOL or FAST method (Ratto et al., 2001; Saltelli et al., 2004). The response surface would not be a smooth function, and methods based on Gaussian Emulators cannot be applied (Oakley and O'Hagan, 2004) to which this method bears some similarities as it also interpolates a response surface. The type of regression tree applied in this paper gives a step response of the surface (Torgo, 1999a) and thus may be of different accuracy to Gaussian Emulators.

In this paper we expand the idea of TSDE based on the concept of *Random Forests* (Breiman and Cutler, 2004; Dietterich, 2000). We explore how the concept of Random Forests can be used to characterize model response surfaces and investigate parameter sensitivity. The method is demonstrated on a set of synthetic functions and in the application of a one-dimensional flood routing model. A comparison to other techniques is given.

2. Regression trees and Random Forests

A regression or decision tree consists of a sequence of questions which split the parameter space according to a response variable. This results in a set of branches which subdivide the parameter space into disjoint hyperspaces (leaves) of the predictor variable. A tree is grown until a termination criterion is reached and results in a maximum tree. Such a maximum tree has very often too many leaves and overfits the response surface which results in a poor predictive capability. Thus a pruning procedure is required to cut the tree back (see Esposito et al., 1997 and references therein).

The subdivision can be either crisp or fuzzy (Han et al., 2002; Ho, 1998) and will depend on the criteria used to determine each division. Regression trees derived in this way are in fact piece-wise constant or piece-wise linear functions. A large number of different regression tree methodologies exist, such as RETIS (Karalic and Cestnik, 1991), CART (Breiman et al., 1984) and M5 (Quinlan, 1992). A more detailed description of regression trees and their limitations is given in Torgo (1999a,b) and Breiman and Cutler (2004), and Breiman et al. (1984). A discussion related to applying these techniques to hydrological problems is given by Iorgulescu and Beven (2004) and Han et al. (2002).

The splitting criterion used in this paper is the deviation from the mean of the two subsamples of each possible split. This criterion may not strictly hold for the problem in question and further assessment may be needed. For other distributions other criteria might be more effective (Breiman et al., 1984; Su et al., 2004; Torgo, 1999b). The following equation is minimized over all possible splits:

$$\mathrm{ssx}=\frac{\sum_{i\in L}(Y_i-\bar{Y}_L)^2}{n_L}+\frac{\sum_{i\in R}(Y_i-\bar{Y}_R)^2}{n_R} \tag{1}$$

Y, response variable ($\bar{Y}$ for mean); L, subset L; R, subset R; n, number of response variables in set L and R, respectively; ssx, sum of squared errors.

For environmental models the response variable can be flow (Iorgulescu and Beven, 2004) or any other variable which is computed directly or indirectly by the model. Thus it is also possible to use any model evaluation criteria. Splitting the parameter space into values greater and smaller than a threshold forms a subset. However, such a subdivision does not have to concentrate solely on the input parameters, but can also be performed on combinations of inputs – so called oblique trees (Iorgulescu and Beven, 2004).

A Random Forest is a set of regression trees that are created independently (Breiman, 2001a). Such a regression forest does not have the same explanatory power as a singular regression tree (Breiman, 1996, 2001a,b; Iorgulescu and Beven, 2004). However, it can be argued that a single tree which splits only accordingly to one criterion is very sensitive towards the training data set. A small change is this data set (as well as a change of the splitting criterion) can lead to different tree structures and thus to different explanations (Breiman, 1996).

In Breiman (1996) the trees are grown on random selections of training sets (bagging/bootstrapping). A training set is a subsample of the entire available simulation results. The internal error rate of each tree can be evaluated instantly by evaluating all samples which have not been used to grow the tree and computing their error rate (*Test Error Rate*).

$$\mathrm{TER}=\frac{\sum(Y_{\mathrm{tree}}-Y_{\mathrm{observed}})^2}{n} \tag{2}$$

TER, Test Error Rate; Y_{tree}, response variable given by the regression tree; Y_{observed}, response variable; n, number of response variables.

The number of trees to be grown depends on how quickly the distribution of the Test Error Rate stabilizes. In the examples given in this paper, $\frac{2}{3}$ of model realisations have been used to grow the tree and $\frac{1}{3}$ to compute the Test Error Rate. Other proportions can be considered and may have to be tested (Su et al., 2004). The proportion used in this paper reflects the traditional calibration and evaluation methodology in hydrological modelling.

In Breiman (2001a) this method is combined with randomizing the features/inputs which are chosen as split possibilities. This means that for each tree a training set is chosen randomly; then a random subsample of features/inputs is determined (see also Ho, 1998, 2002). From this the optimum split is chosen. Breiman (1996) shows that this random sampling is comparable with other algorithms such as Adaboost (Freund, 2001) but performs better with noisy data. Another example of random split selection is given by Dietterich (2000) who grows each tree by choosing randomly from the best 20 splits.

The latter is modified here to fit within the GLUE framework, by dividing into behavioural and non-behavioural tree structures. Every split, which gives a positive information gain is seen as potential split possibility (behavioural tree structure).

$$\frac{\sum_{i\in(L\cup R)}(Y_i-\bar{Y})^2}{n_L+n_R}\geq\frac{\sum_{i\in L}(Y_i-\bar{Y}_L)^2}{n_L}+\frac{\sum_{i\in R}(Y_i-\bar{Y}_R)^2}{n_R} \tag{3}$$

(notations as above).

It is impossible to split along all behavioural tree structures due to computational constraints. Therefore, a variation of Dietterich (2000) is applied, by sampling one split randomly taken from the distribution of ssx (Eq. (1)). However, further constraints are necessary to limit the computational burden: a split occurs only if the maximum distance of Y is greater than a threshold (*Minimum Threshold*):

$$\max(Y)-\min(Y)\geq \mathrm{MT} \tag{4}$$

$\max(Y)$, maximum value of Y; $\min(Y)$, minimum value of Y; MT, Minimum Threshold.

This criterion can be also termed *Threshold on Impurity* and has been applied in various forms to terminate the growth of regression and decision trees (Murthy, 1995).

A node is also not split further when it would result in a sample of smaller than 10 (*Minimum Sample Size*) (Friedman, 1977). The decision of setting these limits depends on the modelling target. For example, if one tries to predict water levels, it might be possible to define a *Minimum Threshold* under which the sensitivity is of less interest. In a flood inundation model one might decide that any variations within 10 cm are of minor interest, because this accuracy is not needed (out of similar reasoning a Maximum Threshold could be defined).

Tree quality is very often more influenced by good stopping or pruning rules than the splitting criteria (Breiman et al., 1984). The actual level of any threshold has to be investigated carefully and may differ significantly from default values given by Breiman (2001a,b) as has been shown by Segal (2004). A summary of various other methods for stopping and pruning and their applications can be found in Murthy (1995, chapter 2.4).

3. Regional sensitivity analysis

A Random Forest can be used to evaluate the sensitivity of parameters or parameter combinations.

The first method counts the **information gain** (*IG-Sensitivity*, Eq. (3)) which each variable achieves on average when it is split (*IG-Sensitivity*). In this paper each information gain is multiplied by the inverse Test Error Rate and all measures are normalized to a sum of 1. This is a modification of the original approach in which the sum of the decreases in a splitting criterion for each variable has been taken as an indication of variable importance (for more details, see Breiman, 2001a). This method concentrates on the one-dimensional importance of variables and can give additional information on the shape of the response surface as later analysis will show.

The second method will be called **permutational sensitivity** (*P-Sensitivity*). For example, if the importance of variable m is tested, then the variable is permuted for all data sets which have not been used for growing the tree and the prediction variable estimated with the regression tree. The difference of the prediction variable before and after permutation gives an indication of importance of this specific variable.

$$I = \frac{\sum \left| \mathrm{TER}_{\mathrm{permuted}} - \mathrm{TER}_{\mathrm{non\text{-}permuted}} \right|}{\mathrm{TER}_{\mathrm{non\text{-}permuted}}} \quad (5)$$

I, measure of importance; TER, Test Error Rate (see Eq. (2)); (notation as above).

The original framework of Breiman (2001a) can be extended by permuting combinations of parameters (*multiple effects*). A permutation of two parameters together will include the errors caused by the permutation of each single variable. If it can be assumed that the importance of each subset is additive following Sobol (1993), the importance of the combination of parameters can be quantified.

For example a second order effect is corrected by

$$\hat{I}_{1,2} = \frac{|I_{1,2} - (I_1 + I_2)|}{I_{total}} \quad (6)$$

$I_{1,2}$, measure of importance computed by permuting parameter 1 and 2; I_1, measure of importance computed by permuting parameter 1; I_2, measure of importance computed by permuting parameter 2; I_{total}, measure of importance computed by permuting all parameters.

In this equation all the lower order effects are subtracted from the higher order importance measure. Eq. (6) raises the important issue of parameter interaction and dependency. The errors of the output of physically based models are correlated within the parameter space. This will be especially true within the GLUE framework in which it is the parameter *set* that gives a behavioural model and therefore introduces some form of correlation. In this case Eq. (6) has to be treated with care due to its additive assumption and it might be necessary to analyse the high order effects without trying to remove the interaction terms. Similar concerns may be valid for correlated input parameters (see p25, Saltelli et al., 2004).

This method relies on local splits and thus is more sensitive towards local changes of the response surface (Iorgulescu and Beven, 2004). By combining this method with the original regional sensitivity analysis of Spear and Hornberger (Spear et al., 1994) a more detailed regional sensitivity analysis can be performed. The parameter space can be subdivided into hyperboxes and the influence of permutations in each subset (hyperbox) explored. This more detailed analysis is also reflected in the choice of the *Minimum Threshold* and the *Minimum Sample Size*, both of which reflect the influence of subheterogeneities on the overall model result.

4. Applications

Several equations have been chosen to test the methodology. Initially a simple model structure is tested which has been taken from chapter 6 of Saltelli et al. (2004). This is followed by an example using the Sobol g-function (Ratto et al., in press). Finally, the methodology is demonstrated on a one-dimensional flood inundation code to show an application to a hydrological example. In all examples presented, 1000–3000 different trees have been grown. The number of trees depends on the speed at which the distribution of the Test Error Rate stabilizes and has to be investigated separately for each example.

The results are further influenced by the sampling density. In these examples an average of 3000 parameter combinations have been evaluated with each model. The more sparse the sample, the more uncertain becomes the result of the sensitivity analysis (Hofer, 1999). This is of special importance as soon as high dimension parameter spaces are sampled. For example an eight parameter model with only 10 values in each direction would require a minimum of 100×10^6 function evaluations with no guarantee that all complex interactions are captured. Therefore, the total amount of model runs required may vary significantly and can be estimated by analysing changes in the cumulative distribution of the response variable as sample size is increased (Pappenberger et al., 2005).

All the results are compared to scatter plots of the parameter space versus the response variable, and a comparison to traditional correlation analysis is performed. Helton and Davis (2003) have suggested several methods to compute correlation and show that different results can be achieved according to the coefficient chosen (see also Pastres et al., 1999). In this paper three different methods have been applied: Pearson Correlation, Spearman Rank Correlation, Kendall's Tau. The Pearson correlation is a standard method to evaluate the linear correlation between values by comparing the deviations from the mean. The Spearman Rank correlation coefficient is a non-parametric test, which compares the ranks of data. It has been used in various environmental related papers to show relationships between parameters or responses (e.g. Romanowicz and Beven, 2003). Kendall's Tau is even more non-parametric than Spearman's rank correlation. It does not use the numerical difference in rank, but the relative ordering of ranks. This paper does not show the equations to compute these correlation coefficients and the reader is referred to standard text books (e.g. Press et al., 2002 p641ff).

5. Structure 1

The first model structure is multiplication of two parameters with a third parameter which has zero influence. All parameters are sampled from a uniform distribution with values between −0.5 and 0.5 under the condition that $Y > 0$. The choice of the distribution and especially the sampling range will have an effect on the results of any global sensitivity analysis (Oakley and O'Hagan, 2004).

5.1. Equation

$$Y = X_1 X_2 + 0(X_3) \tag{7}$$

with

$$X_j \sim U(-0.5, 0.5)\big|_{Y>0} \tag{8}$$

Y, predicted variable; X_{1-3}, parameters.

5.2. Results

The results are split into three parts: first we show the scatter plots of the structure; then the results of the computation of the Spearman Rank Correlation coefficient; and finally the SARS-RT analysis.

5.2.1. Scatter plots

In the scatter plots parameters X_1 and X_2 are plotted against the predicted variable Y. Each dot represents one set of parameters. Fig. 1a shows a clear response for parameters X_1 and X_2 in respect to variable Y. It would be very difficult from these plots to make comments about the behaviour of parameter X_3. In the next figure the parameters are plotted against each other and the response variable Y is shown on a grey scale. Fig. 1b illustrates the strong dependency of the interrelationship between parameters X_1 and X_2 in respect to the predicted variable. It further reveals the gaps in the parameter space due to the restriction of the predicted variable Y.

5.2.2. Correlation analysis

In Table 1 the correlation analysis is presented.

It can be seen that the correlation coefficients detect a relationship between X_1 and X_2, which can also be seen from the scatter plots. No clear relationship to the predicted variable is established. This example demonstrates very neatly the failure of the correlation tests (Saltelli et al., 2004), because it does not reveal any structural information. Therefore, the detection of no relationship by the correlation coefficient cannot be used to reject the hypothesis of the existence of some relationship (thus some conclusions based on this may be partly invalid see e.g. Romanowicz and Beven, 2003). In view of this reasoning, the correlation analysis is not repeated in the following examples

5.2.3. SARS-RT

In the following we show the results of the SARS-RT analysis for structure 1. All figures of this analysis plot the outcome of an analysis with 1000 trees of the SARS-RT method. On each abscissa are the parameters or parameter combinations plotted. On the ordinate are the importance measure values. The distribution of the 1000 trees is shown as box and whisker plots. Each box shows the lower quartile, the median and the upper quartile as a line. These lines are extended to indicate the range of the rest of the data. Data points which extend over 1.5

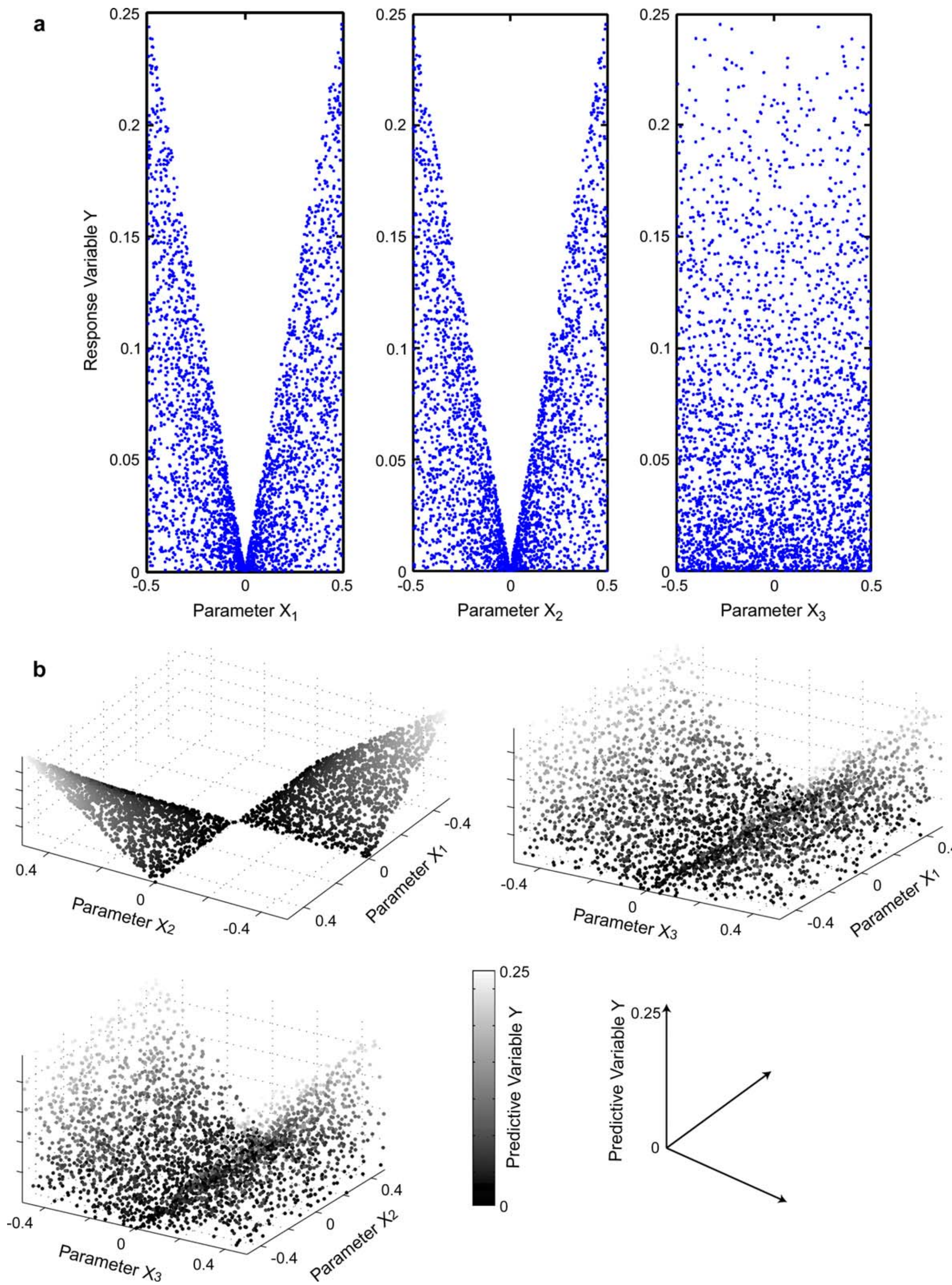

Fig. 1. (a) Scatter plot of Structure 1, which plots each parameter against the response variable Y. (b) Scatter plot of Structure 1, with the parameters plotted against each other and the predicted variable as grey scale.

Table 1
Correlation coefficients of Structure 1

Method	Parameter X_1			Parameter X_2			Parameter X_3			Response variable Y		
	Pearson	Spear	Kendall	Pearson	Spear	Kendall	Pearson	Spear	Kendall	Pearson	Spear	Kendall
Parameter X_1	1.00	1.00	1.00	**0.73**	**0.73**	**0.48**	0.03	0.03	0.02	0.16	0.19	0.13
Parameter X_2	–	–	–	1.00	1.00	1.00	0.00	0.00	0.00	0.13	0.15	0.11
Parameter X_3	–	–	–	–	–	–	1.00	1.00	1.00	0.09	0.09	0.06

times the length of the inner quartile range are marked by a cross.

Fig. 2 presents the results of the IG-Sensitivity test, which clearly demonstrates the importance of parameters 1 and 2 in respect of the response variable. A similar picture is given by Fig. 3, which plots the results of the P-sensitivity test for the first order effects. In Fig. 4 the latter plot is extended and the 2nd order effects shown. Second order effects are all those parameter combinations which involve two parameters. Both methods give numerically different results as they are normalized in different ways and moreover, the P-Sensitivity measure is only applicable for 1st order interactions.

The SARS-RT analysis indicates the same results as the correlation computations in combination with the scatter plots. Therefore, it seems to be capable of reproducing standard results (in this case the visual analysis of the scatter plots and the linear correlation analysis). It shows that parameters 1 and 2 have an importance of roughly 0.35 for the IG-measure each, but that there is also a strong relationship between these two, and they both influence the predicted variable Y.

If we plot the information gain versus the split position for each parameter (Fig. 5), the previous scatter plots are resembled and positions of high and low activity become visible. The decline of the information gain measure towards the edges is clearly a function of this technique and not a change in sensitivity. This is a drawback and may pose problems.

Furthermore, if a subset analysis is performed with 4 boxes on each axis the pattern of the two-dimensional scatter plot emerges clearly. In Fig. 6 only the one-dimensional effects are shown for demonstration. Areas with stronger one-dimensional impacts can be identified as well as areas in which a more two-dimensional behaviour is expected. This plot does show the additional potential of this methodology: the relationships in each sub-box can be expressed in numerical terms and compared to other locations in the hyperspace. Therefore, all locations with changing relationships can be identified, which might give additional insight into the model behaviour. The cuts can become irregular by clustering areas with similar sensitivity.

The SARS-RT analysis has been shown to be capable of reproducing expected results and revealing structural elements of the response surface.

6. Structure 2

This example uses the Sobol g-function which is a non-linear and non-additive model. It has been used in previous studies to test global sensitivity analysis models (Ratto et al., in press; Saltelli et al., 2004).

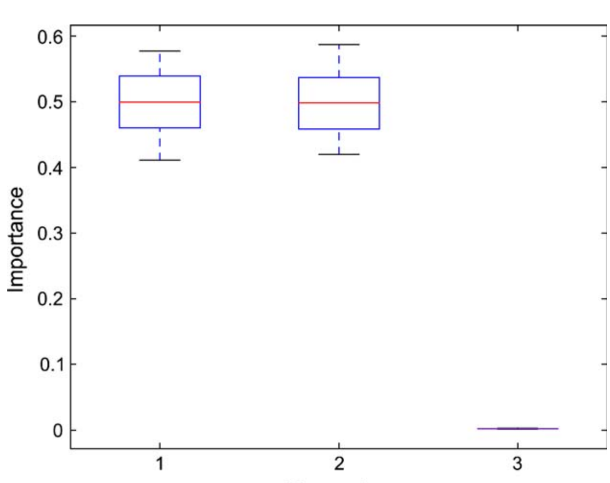

Fig. 2. Box and whisker plot of the IG-Sensitivity measure of Structure 1.

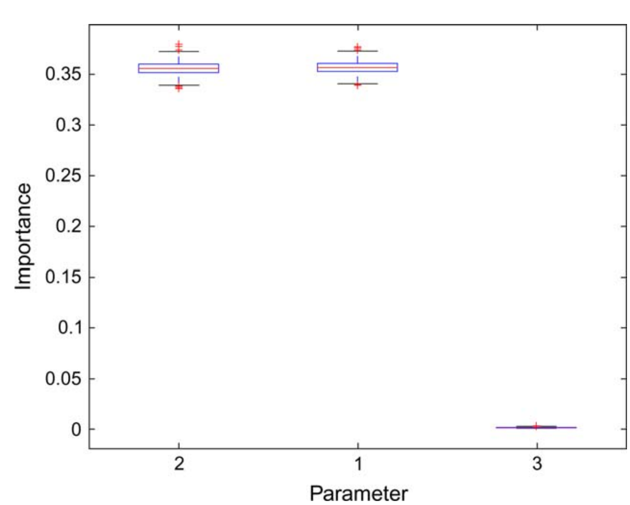

Fig. 3. Box and whisker plot of the P-Sensitivity measure of Structure 1 – 1st order effects.

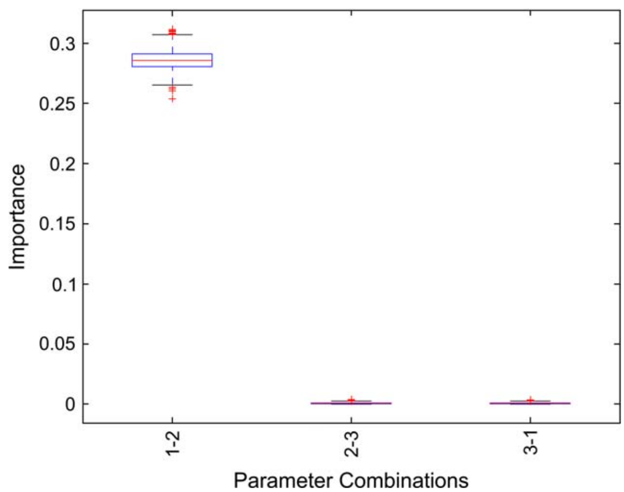

Fig. 4. Box and whisker plot of the P-Sensitivity measure of Structure 1 − 2nd order effects.

$$Y = \prod_{j=1}^{k} g_j(X_j) \quad (9)$$

where

$$g_j(X_j) = \frac{|4X_j - 2| + a_j}{1 + a_j} \quad (10)$$

with $a_j \geq 0$ and $X_j \sim U(0,1)$.

This allows the importance of parameters to be tuned by changing a. The smaller a_j is the more important is X_j and the larger is its influence on y.

6.1. Results

For this structure eight input factors have been considered ($k = 8$) with $a_j = [0,1,4.5,9,99,99,99,99]$. This allows a comparison with the Sobol method which has been published by Ratto et al. (in press). The results of the Sobol method can also be reproduced with the Simlab software package (follow links in Tarantola and Saltelli, 2004).

6.1.1. Scatter plots

The scatter plots (Fig. 7) indicate the sensitivity of the parameters as suggested earlier (only a subset of combinations is shown). This reinforces the argument that scatter plots should be seen as a useful complement to other evaluation techniques.

However, the scatter plots do illustrate very clearly a common problem in the analysis of multi-dimensional models: visual display is limited to a small number of dimensions and many models have more than 3 or 4 parameters. It is very difficult to understand and analyse multi-dimensional relationships in these cases. However, two-dimensional relationships show (e.g. top left corner) in this figure.

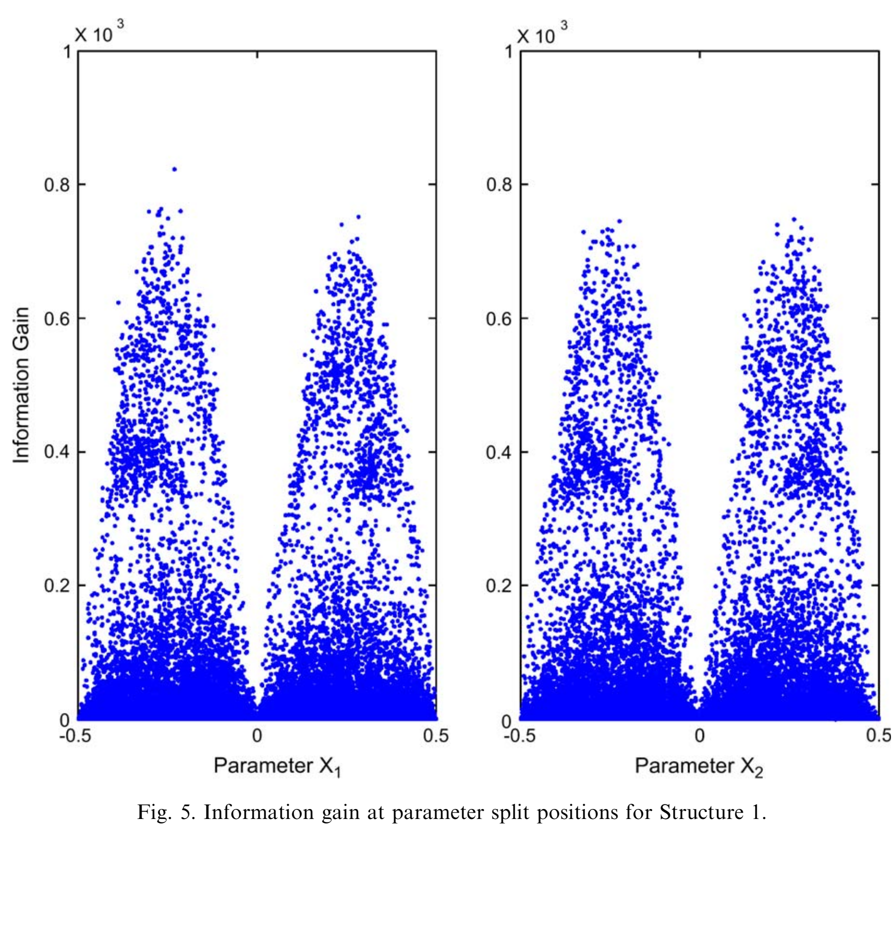

Fig. 5. Information gain at parameter split positions for Structure 1.

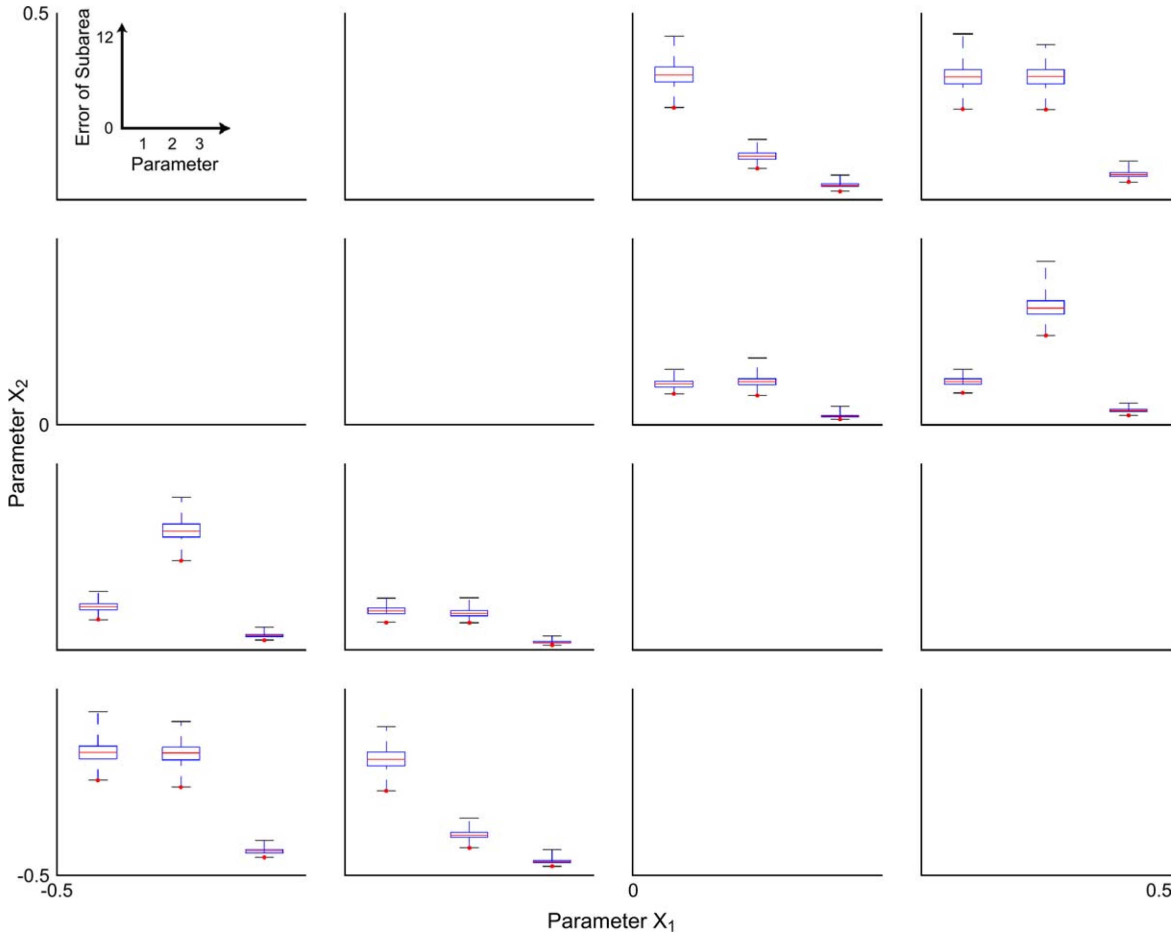

Fig. 6. Subplot analysis of Structure 1 with subdivision into 4 × 4 × 4 hyperboxes. First order effects for parameters 1 and 2.

6.1.2. SARS-RT

Figs. 8 and 9 show the first order effects of the *IG-measure* and the *P-measure* (multiple effects are not shown). The relative importance of the parameters behaves as expected with the first parameter as being most important and the second parameter as the second most important. This result is similar to the previous analysis of Ratto et al. (in press), which comes to similar conclusions (Table 2). In this analysis Ratto et al. demonstrated that the Sobol method is capable of detecting that parameter sensitivity reduces with increase in *a*. In the Sobol method the variance is split and thus the ordinate would show a variance-based measure. However, the values shown on the ordinate Figs. 8 and 9 are not *variance* in the statistical sense. Rather they can be assumed to be relative importance of the parameters to each other. A sum of one would be the result of adding up the sensitivities of all parameters and parameter combinations (as they have been normalized). This example does show that this analysis can reproduce classical sensitivity methods although it is based on an irregular sampling scheme.

All examples given so far have been computed with 'synthetic' equations. However, hydrology and hydraulics deal very often with complex data and models. The next example will be a typical model used in flood forecasting.

7. Structure 3

Structure 3 is the result of applying the one-dimensional flood routing model HEC-RAS (U.S. Army Corps Engineers) which solves the one-dimensional Saint Venant equations to route flood flow. The impact of parameter uncertainty on one-dimensional flood inundation models has been investigated in various previous studies (Horritt and Bates, 2003; Matgen et al., 2004; Pappenberger et al., 2003; Wiles and Levine, 2002) and one sensitivity analysis of a more complex flood inundation model exists (Hall et al., 2005). HEC-RAS has been applied on the Alzette catchment (Luxembourg) with 7 different parameters which consist of: 3 different surface roughnesses, 3 parameters which control the magnitude of the model input and one parameter controlling the accuracy of the numerical solution (see Table 3). The Monte Carlo based generalized likelihood uncertainty estimation (GLUE) has been performed with

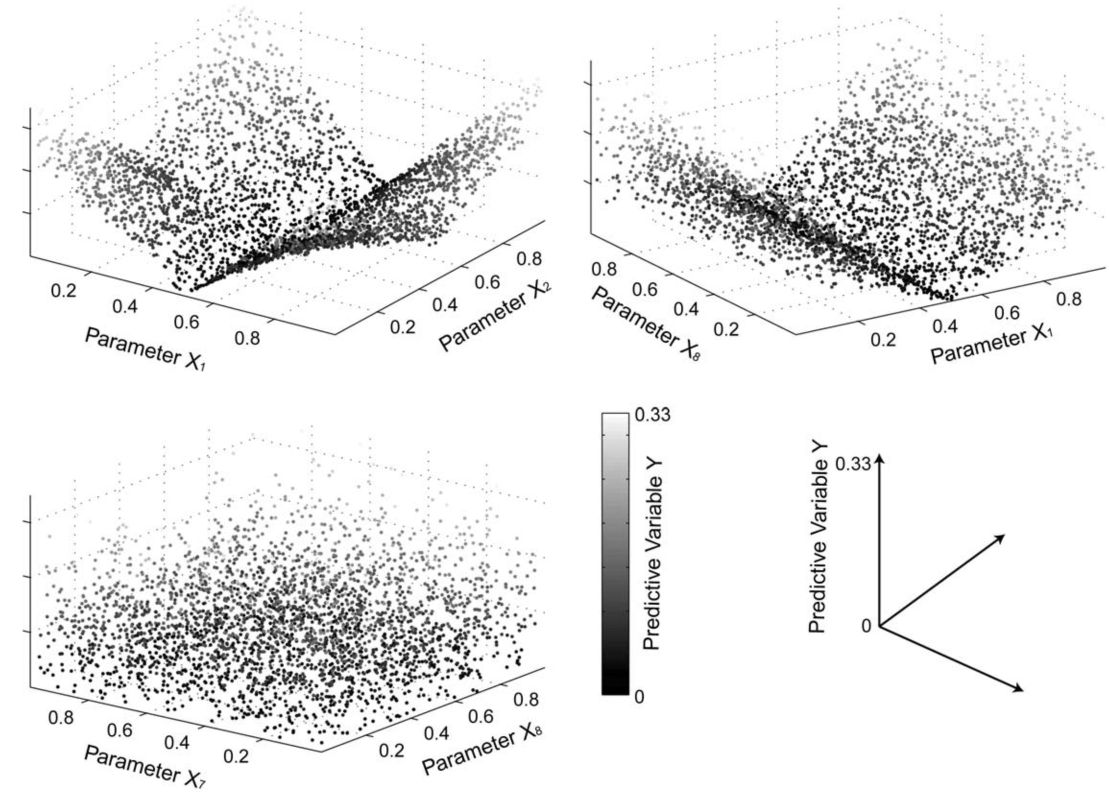

Fig. 7. Scatter plot of Structure 2, which plots each parameter against the response variable Y.

these model parameters (the reader is referred to Beven (2001) for details on method description). The model has been evaluated on a satellite picture of flood extent with a performance measure (Y). The performance measure has been normalized and is scaled from 0 to 1 (with 1 being the best performing models). All model results which perform worse than a linear interpolation from the upstream to the downstream boundary have been classified as non-behavioural and excluded from further analysis. The reader is referred to Matgen et al. (2004) and Pappenberger et al. (2004) for a more detailed description of the model, the model set-up and the evaluation data applied. This paper does not attempt to give a full understanding of Structure 3 and the set-up; it rather demonstrates the possibility of applying SARS-RT within a common model calibration and evaluation exercise. In addition, it demonstrates that the *Minimum Threshold* can have a real meaning. For example, if we were to use water levels in this example, the threshold

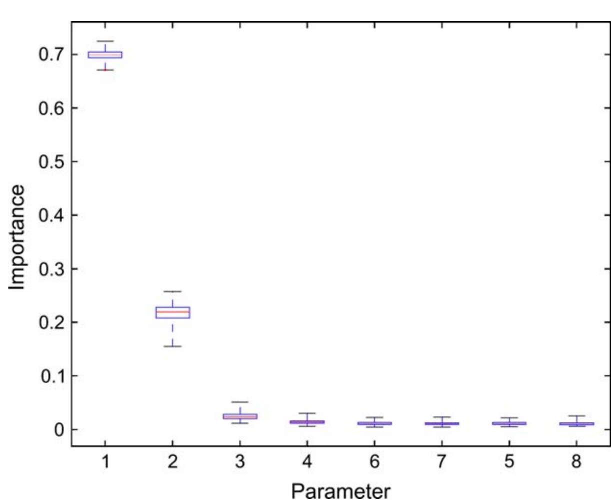

Fig. 8. Box and whisker plot of IG-Sensitivity measure of Structure 2.

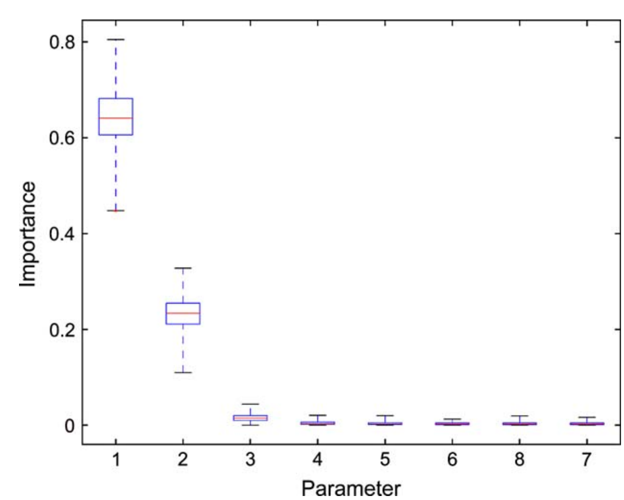

Fig. 9. Box and whisker plot of the P-Sensitivity measure of Structure 2 − 1st order effects.

Table 2
Sobol values for the main effects of Structure 2

Parameter	1	2	3	4	5	6	7	8
Main effect	0.7743	0.1941	0.0257	0.0078	0.0001	0.0001	0.0001	0.0001

represents the water level range in which we assume that the sensitivity of the our model predictions is not of interest. For example, there may be measurement scale versus prediction scale issues for a variable that result in a range of acceptable model behaviours (Beven, in press).

The sampling range of all parameters is uniform over the whole range shown (only the behavioural parameter range is shown). Parameter 4 could only have the values of 0.6, 0.7, 0.8, 0.9 and 1.0 due to constraints in the model set-up. Around ~4000 model realisations have been computed.

7.1. Results

The scatter plots are not shown for this example, as it is difficult to display this multi-dimensional response surface. The surface demonstrates highly non-linear relationships.

7.1.1. SARS-RT

Figs. 10–14 show the results of the SARS-RT analysis in the same order as for the previous model structures.

A correlation analysis only manages to identify a relationship between parameters 5 and 7, but fails to identify further structures (not shown here). Multiple effects and thus combinations of parameters rather than individual parameters dominate the response of this model. The Manning roughness of the channel is the most important factor and clearly interacts with the left flood plain friction. In contrast, the right flood plain friction seems to be of minor importance when it is evaluated against a global inundation measure (it still can be of importance on a local scale). The SARS-RT analysis reveals further interplay between channel roughness, left flood plain roughness and input parameters 5 and 6. This reflects our physical understanding which would suggest that input and roughness have to be related in order to achieve behavioural model results. This example seems to reveal that parameter 4 plays only a minor role in controlling the performance measure (it still plays an important role in controlling the accuracy of the numerical solution).

It would have been very difficult to extract such relationships only from scatter plots, whereas this SARS-RT analysis allows to speculate further on the structure of the response surface – especially when it is difficult to visualize.

Table 3
Parameters of the HEC-RAS model

No.	Parameter	Sample range
1	Manning surface roughness of channel	0.01–0.2
2	Manning surface roughness of right flood plain	0.01–0.2
3	Manning surface roughness of left flood plain	0.01–0.2
4	Theta (controlling accuracy of numerical implicit scheme)	0.5–1.0
5	Rating Curve Parameter (controlling magnitude of inflow)	0.05–0.2
6	Rating Curve Parameter (controlling magnitude of inflow)	0.2–0.7
7	Rating Curve Parameter (controlling magnitude of inflow)	0.001–0.01

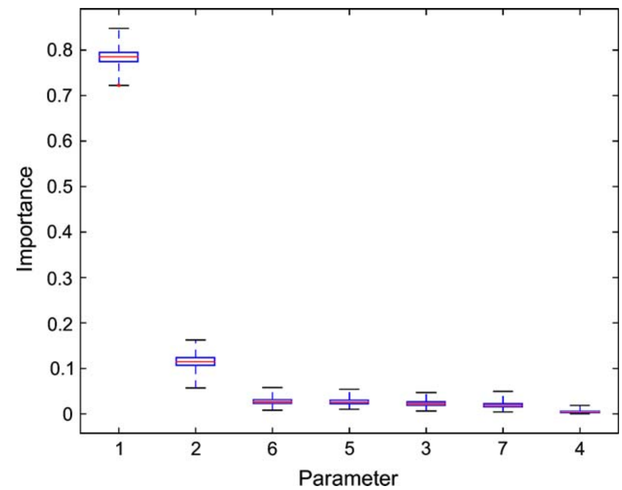

Fig. 10. Box and whisker plot of the IG-Sensitivity measure of Structure 3.

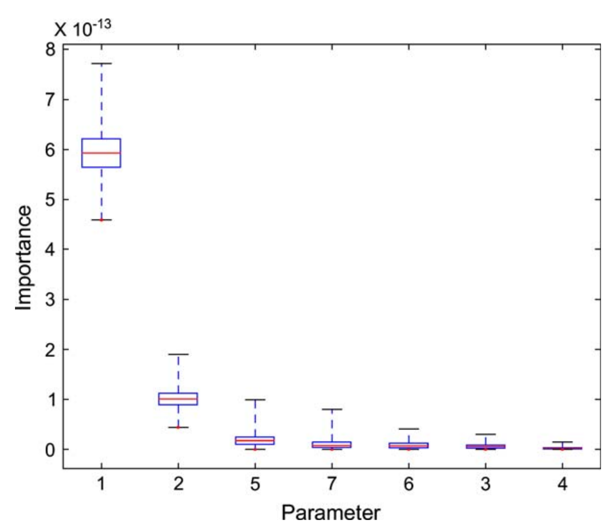

Fig. 11. Box and whisker plot of the P-Sensitivity measure of Structure 3 – 1st order effects.

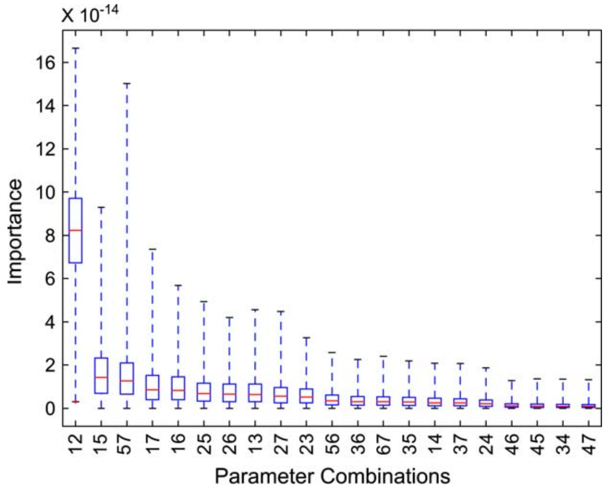

Fig. 12. Box and whisker plot of the P-Sensitivity measure of Structure 3 – 2nd order effects.

8. Conclusion

In this paper the sensitivity analysis of Breiman (2001a) has been extended to analyse multi-dimensional relationships of parameter response surfaces. The method is based on growing multiple regression trees.

With the Random Forest it is possible to evaluate a goodness of fit (*Test Error Rate*) and perform sensitivity analysis. For the first method the information gain for each split of the branches of the tree is multiplied by the inverse Test Error Rate and analysed in respect of the splitting parameter. For the second method, the parameters have been permuted and the resulting error computed. This can be done with only one input parameter (as in the original of Breiman, 1996) or with multiple parameter combinations.

The possibility of investigating the sensitivity of subregions or hyperboxes has been introduced. However,

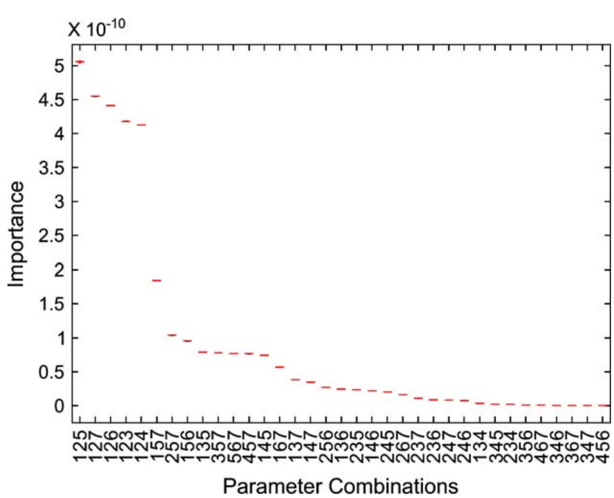

Fig. 13. Box and whisker plot of the P-Sensitivity measure of Structure 3 – 3rd order effects.

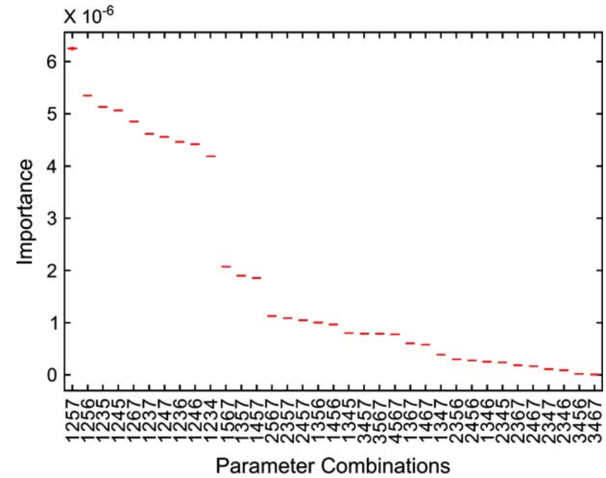

Fig. 14. Box and whisker plot of the P-Sensitivity measure of Structure 3 – 4th order effects.

this methodology needs to be explored further. This is very similar to the TSDE of Spear et al. (1994), which is a tool mainly developed to support re-sampling. It has to be stressed that the method in this paper has not been used for re-sampling purposes, although it might be applicable. This re-sampling may be especially important to increase the efficiency of uncertainty analyses.

The TSDE and this method (SARS-RT) differ from other global sensitivity analysis (Ratto et al., 2001, in press; Saltelli et al., 2004). The analysis concentrates on local splits and differences rather than global properties and handles the local interactions and non-linearities aspect in complex models. However, mostly global results are presented and thus in this application the method has to be classified as a regional sensitivity analysis with global properties. Moreover, this technique does not require a definite sampling pattern or an approximation to such a pattern. This is significantly different from other methods and is more realistic for many environmental models, which may be, for example, constrained by instabilities of the numerical solutions or where models that are considered behavioural are retained for analysis. However, this methodology should be seen as an addition to the well explored and defined techniques of Saltelli et al. (2004) as the results of several methods should always be compared (Babendreier and Castleton, 2005).

In this paper two different model structures of different complexity have been investigated to demonstrate the properties of this method. The (synthetic) structures have been chosen in a way that it was possible to understand or analyse their properties without the help of SARS-RT. All model structures contained non-linearities as well as gaps in the response surface. All results have been compared to scatter plots of the parameter space as well as mathematical understanding of the equations. A good agreement between this type of visual analysis and the numerical results has been achieved.

In each example it has been possible to rank the sensitivity of each parameter according to importance. This ranking is in agreement to the mathematical understanding and the visual analysis of the scatter plots of the equations. Thus it can be hoped that it also holds for more complex model formulations.

Furthermore, multi-dimensional relationships have been explored. It has been demonstrated that SARS-RT can assist in a better understanding of the shape and properties of response surfaces. It has been possible to identify complex interactions in multi-dimensional space and the results are comparable to other established frameworks such as the Sobol method.

SARS-RT has been successfully applied to a hydraulic model code in identifying response surface structures. It did provide useful information of the dependencies and the importance of model parameters. However, this method should not been seen as an ad hoc approach. It rather indicates places (hyperboxes) in which further investigation is necessary. It should be used in conjunction with other methods and offers support when one does not see the wood for the trees.

In the traditional calibration/optimisation framework, parameter sensitivity is always seen as a by-product, which is very often not explored further. However, recent models have become more and more complex and interactions increasingly non-linear. This paper has demonstrated that a simple one-dimensional analysis of model results may hide important information and lead to misleading interpretation of model parameter sensitivity. A complete model analysis should therefore include some form of sensitivity analysis which acknowledges the complexity of the applied model structures.

Acknowledgements

Thanks to Patrick Matgen of the Centre de Recherche Public – Gabriel Lippmann (CRP-GL), Grand Duchy of Luxembourg for the data for the one-dimensional flood routing model. We further would like to thank Hannah Cloke (Kings College, London, England), Paul Smith (Lancaster University, Lancaster, England), Marco Ratto (Joint Research Centre of the European Commission, Ispra, Italy) and two anonymous reviewers for their valuable comments. This research was supported by the NERC Grant NER/L/S/2001/00658.

Reference

Babendreier, J.E., Castleton, K.J., 2005. Investigating uncertainty and sensitivity in integrated, multimedia environmental models: tools for FRAMES-3MRA. Environmental Modelling & Software 20 (8), 1043–1055.

Beven, K.J., 2001. Rainfall-Runoff Modelling: The Primer. Wiley, New York.

Beven, K.J., 2002. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. Hydrological Processes 16 (2), 189–206.

Beven, K.J., A manifesto for the equifinality thesis. Journal of Hydrology, in press.

Beven, K.J., Binley, A., 1992. The future of distributed models: model calibration and uncertainty prediction. Hydrological Processes 6, 279–298.

Beven, K.J., Pappenberger, F., 2003. Towards the hydraulics of the hydroinformatics era – by M.B. Abbott, V.M. Babovic and J.A. Cunge, Journal of Hydraulic Research, vol. 39. 2001, Issue 4, pp. 339–349 – discussion. Journal of Hydraulic Research, 41(3), 331–333.

Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.

Breiman, L., 2001a. Random forests. Machine Learning 45 (1), 5–32.

Breiman, L., 2001b. Statistical modeling: the two cultures. Statistical Science 16, 199–215.

Breiman, L., Cutler, A., 2004. Random Forests. <http://oz.berkeley.edu/users/breiman/RandomForests/>.

Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth Pub. Co.

Chan, K., Saltelli, A., Tarantola, A., 1997. Sensitivity analysis of model output: variance-based methods make the difference. In: Andradottir, S., Healy, K.J., Withers, D.H., Nelson, B.L. (Eds.), Winter Simulation Conference. <http://www.informs-cs.org/wsc97papers/0261.PDF>.

Dietterich, T.G., 2000. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learning 40 (2), 139–157.

Doherty, J., 2002. PEST Model-Independent Parameter Estimation. <http://www.sspa.com/pest/download.html>.

Duan, Q., Sorooshian, S., Gupta, V., 1992. Effective and efficient global optimisation for conceptual rainfall-runoff models. Water Resources Research 28, 1015–1031.

Esposito, F., Malerba, D., Semeraro, G., 1997. A comperative analysis of methods for pruning decision trees. IEEE Transactions on Pattern Analysis and Machine Intelligence 19, 476–491.

Freund, Y., 2001. An adaptive version of the boost by majority algorithm. Machine Learning 43 (3), 293–318.

Friedman, J.H., 1977. A recursive partitioning decision rule for nonparametric classifiers. IEEE Transactions on Computers, 404–408.

Georgakakos, K.P., Krzysztofowicz, R., 2001. Probabilistic and ensemble forecasting. Journal of Hydrology 249 (1–4), 1–1.

Grieb, T.M., Hudson, R.J.M., Shang, N., Spear, R.C., Gherini, S.A., Goldstein, R.A., 1999. Examination of model uncertainty and parameter interaction in a global carbon cycling model (GLOCO). Environment International 25 (6–7), 787–803.

Gupta, H.V., Sorooshian, S., Yapo, P.O., 1998. Toward improved calibration of hydrologic models: multiple and noncommensurable measures of information. Water Resources Research 34 (4), 751–763.

Hall, J.W., Tarantola, S., Bates, P.D., Horritt, M.S., 2005. Distributed sensitivity analysis of flood inundation model calibration. ASCE Journal of Hydraulic Engineering 131 (2), 117–126.

Han, D., Cluckie, I.D., Karbassioun, D., Lawry, J., Krauskopf, B., 2002. River flow modelling using fuzzy decision trees. Water Resources Management 16 (6), 431–445.

Helton, J.C., Davis, F.J., 2003. Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems. Reliability Engineering & System Safety 81 (1), 23–69.

Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (8), 832–844.

Ho, T.K., 2002. A data complexity analysis of comparative advantages of decision forest constructors. Pattern Analysis and Applications 5 (2), 102–112.

Hofer, E., 1999. Sensitivity analysis in the context of uncertainty analysis for computationally intensive models. Computer Physics Communications 117 (1–2), 21–34.

Hornberger, G.M., Spear, R.C., 1981. An approach to the preliminary analysis of environmental systems. Journal of Environmental Management 12, 7–18.

Horritt, M.S., Bates, P.D., 2003. Evaluation of 1D and 2D numerical models for predicting river flood inundation. Journal of Hydrology 268 (1–4), 87–99.

Iorgulescu, I., Beven, K.J., 2004. Nonparametric direct mapping of rainfall-runoff relationships: an alternative approach to data analysis and modeling? Water Resource Research 40, doi: 10.1029/2004WR003094.

Jolma, A., Norton, J., 2005. Methods of uncertainty treatment in environmental models. Environmental Modelling & Software 20 (8), 979–980.

Karalic, A., Cestnik, B., 1991. The Bayesian approach to tree-structured regression. In: Proceedings of the ITI-91, Cavtat, Croatia.

Kavetski, D.N., Franks, S.W., Kuczera, G., 2003. Confronting input uncertainty in environmental modelling. In: Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Advances in Calibration of Watershed Models. American Geophysical Union, Washington, pp. 49–68.

Krzysztofowicz, R., 2001. The case for probabilistic forecasting in hydrology. Journal of Hydrology 249 (1–4), 2–9.

Krzysztofowicz, R., 2002a. Bayesian system for probabilistic river stage forecasting. Journal of Hydrology 268 (1–4), 16–40.

Krzysztofowicz, R., 2002b. Probabilistic flood forecast: bounds and approximations. Journal of Hydrology 268 (1–4), 41–55.

Krzysztofowicz, R., Herr, H.D., 2001. Hydrologic uncertainty processor for probabilistic river stage forecasting: precipitation-dependent model. Journal of Hydrology 249 (1–4), 46–68.

Kuczera, G., Parent, E., 1998. Monte Carlo assessment of parameter uncertainty in conceptual catchment models: the Metropolis algorithm. Journal of Hydrology 211 (1–4), 69–85.

Matgen, P., Henry, J.-B.F., Pappenberger, F., Pfister, L., de Fraipont, P., Hoffmann, L., 2004. Uncertainty in calibrating flood propagation models with flood boundaries derived from synthetic aperture radar imagery. XXth ISPRS Congress. International Society for Photogrammetry and Remote Sensing, Istanbul, Turkey.

McCuen, R.H., 1973a. Component sensitivity: a tool for the analysis of complex water resources systems. Water Resources Research 9 (1), 243–247.

McCuen, R.H., 1973b. The role of sensitivity analysis in hydrologic modelling. Journal of Hydrology 18, 37–53.

Melching, C.S., 1995. Reliability estimation. In: Singh, V.P. (Ed.), Computer Models of Watershed Hydrology. Water Resources Publications, Littleton, pp. 69–118.

Murthy, S.K., 1995. On Growing Better Decision Trees from Data, PhD thesis, Johns Hopkins University Baltimore, <http://www.tigr.org/%7Esalzberg/murthy_thesis/thesis.html>.

Oakley, J.E., O'Hagan, A., 2004. Probabilistic sensitivity analysis of complex models: a Bayesian approach. Journal of the Royal Statistical Society Series B – Statistical Methodology 66, 751–769.

Pallottino, S., Sechi, G.M., Zuddas, P., 2005. A DSS for water resources management under uncertainty by scenario analysis. Environmental Modelling & Software 20 (8), 1031–1042.

Pappenberger, F., 2004. Uncertainty in Flood Inundation Models. Lancaster University, Lancaster, 325 pp.

Pappenberger, F., Beven, K., Horritt, M., Bates, P., 2003. Comparison of the uncertainty in ISIS and HEC-RAS within the Generalized Likelihood Uncertainty Estimation framework conditioned on historical data, EGS XXVII General Assembly, Nice, France.

Pappenberger, F., Beven, K.J., Horritt, M., Blazkova, S., 2005. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. Journal of Hydrology 302 (1–4), 46–69.

Pappenberger, F., Matgen, P., Beven, K.J., Henry, J.-B.F., Pfister, L., de Fraipont, P., 2004. The influence of rating curve uncertainty on flood inundation predictions, Flood Risk Assessment, Bath.

Pastres, R., Chan, K., Solidoro, C., Dejak, C., 1999. Global sensitivity analysis of a shallow-water 3D eutrophication model. Computer Physics Communications 117 (1–2), 62–74.

Pastres, R., Ciavatta, S., 2005. A comparison between the uncertainties in model parameters and in forcing functions: its application to a 3D water-quality model. Environmental Modelling & Software 20 (8), 981–989.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2002. Numerical Recipies in C++. Cambridge University Press, Cambridge.

Quinlan, J.R., 1992. Learning with continuous classes. In: Proceedings of the Fifth Australian Joint Conference on Artificial Intelligence. World Scientific, pp. 343–348.

Ratto, M., Saltelli, A., Tarantola, A., Young, P. Improved and accelerated sensitivity analysis using state dependent parameter models, in press.

Ratto, M., Tarantola, A., Saltelli, A., 2001. Sensitivity analysis in model calibration: GSA-GLUE approach. Computer Physics Communications 136, 212–224.

Reichert, P., Borsuk, M.E., 2005. Does high forecast uncertainty preclude effective decision support? Environmental Modelling & Software 20 (8), 991–1001.

Romanowicz, R., Beven, K.J., 2003. Estimation of flood inundation probabilities as conditioned on event inundation maps. Water Resources Research 39 (3), doi: 10.1029/2001WR001056.

Saltelli, A., Tarantola, A., Campolongo, F., Ratto, M., 2004. Sensitivity Analysis in Practice – A Guide to Assessing Scientific Models. John Wiley & Sons, Chichester.

Segal, M.R., 2004. Machine Learning Benchmarks and Random Forest Regression, Center for Bioinformatics & Molecular Biostatistics.

Seibert, J., McDonnell, J., 2003. The quest for an improved dialog between modeler and experimentalist. In: Duan, Q., Gupta, H., Sorooshian, S., Rousseau, A.N., Turcotte, R. (Eds.), Advances in Calibration of Watershed Models. American Geophysical Union, Washington.

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

Spear, R.C., Grieb, T.M., Shang, N., 1994. Parameter Uncertainty and Interaction in Complex Environmental Models. Water Resources Research 30 (11), 3159–3169.

Spear, R.C., Hornberger, G.M., 1980. Eutrophication in Peel Inlet, II. Identification of critical uncertainties via generalised sensitivity analysis. Water Resource Research 14, 43–49.

Su, X., Wang, M., Fan, J., 2004. Maximum likelihood regression trees. Journal of Computational and Graphical Statistics 13 (3), 586–598.

Tarantola, A., Saltelli, A., 2004. The Simlab software package.

Torgo, L., 1999a. Inductive Learning of Tree-based Regression Models, University of Porto, Porto, <http://www.liacc.up.pt/~ltorgo/PhD/>.

Torgo, L., 1999b. Reliable Error Estimates for Pruning Regression Trees, University of Porto, Porto, <http://www.liacc.up.pt/~ltorgo/Papers/REPRT/REPRT-Title.html>.

U.S. Army Corps Engineers, HEC-RAS, http://www.hec.usace.army.mi.

Wagener, T., McIntyre, N., Lees, M.J., Wheater, H.S., Gupta, H.V., 2003. Towards reduced uncertainty in conceptual rainfall-runoff modelling: dynamic identifiability analysis. Hydrological Processes 17 (2), 455–476.

Wiles, J., Levine, N., 2002. A combined GIS and HEC model for the analysis of the effect of urbanization on flooding: the Swan Creek watershed, Ohio. Environmental & Engineering Geosience 8 (1), 47–61.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1996. Automatic calibration of conceptual rainfall-runoff models: sensitivity to calibration data. Journal of Hydrology 181 (1–4), 23–48.

Yapo, P.O., Gupta, H.V., Sorooshian, S., 1998. Multi-objective global optimization for hydrologic models. Journal of Hydrology 204 (1–4), 83–97.