

Influence of uncertain boundary conditions and model structure on flood inundation predictions

Florian Pappenberger ^{a,*}, Patrick Matgen ^b, Keith J. Beven ^a, Jean-Baptiste Henry ^{c,d},
Laurent Pfister ^b, Paul Fraipont de ^d

^a *Environmental Sciences, Lancaster University, UK*

^b *Cellule de Recherche en Environnement et Biotechnologies, Centre de Recherche Public-Gabriel Lippmann, 41, rue du Brill, L-4422 Belvaux, Luxembourg*

^c *VTT Information Technology, PO Box 1201 02044 VTT, Finland*

^d *Service Régional de Traitement d'Image et de Télédétection, Strasbourg University, France*

Received 12 January 2005; received in revised form 22 September 2005; accepted 9 November 2005

Available online 19 May 2006

Abstract

In this study, the GLUE methodology is applied to establish the sensitivity of flood inundation predictions to uncertainty of the upstream boundary condition and bridges within the modelled region. An understanding of such uncertainties is essential to improve flood forecasting and floodplain mapping. The model has been evaluated on a large data set. This paper shows uncertainty of the upstream boundary can have significant impact on the model results, exceeding the importance of model parameter uncertainty in some areas. However, this depends on the hydraulic conditions in the reach e.g. internal boundary conditions and, for example, the amount of backwater within the modelled region. The type of bridge implementation can have local effects, which is strongly influenced by the bridge geometry (in this case the area of the culvert). However, the type of bridge will not merely influence the model performance within the region of the structure, but also other evaluation criteria such as the travel time. This also highlights the difficulties in establishing which parameters have to be more closely examined in order to achieve better fits. In this study *no* parameter set or model implementation that fulfils all evaluation criteria could be established. We propose four different approaches to this problem: closer investigation of anomalies; introduction of local parameters; increasing the size of acceptable error bounds; and resorting to local model evaluation. Moreover, we show that it can be advantageous to decouple the classification into behavioural and non-behavioural model data/parameter sets from the calculation of uncertainty bounds.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Flooding; Uncertainty analysis; Free surface flow; Sensitivity analyse; Modelling

1. Introduction

Flood inundation models play a central role in both real-time flood forecasting and in floodplain mapping. In flood forecasting, inundation models should be as precise as possible to predict the approach of a flood

correctly as well as to avoid false alarms. Flood mapping has to be accurate for a variety of reasons including decision making for local planning or the insurance industry. A full understanding of the model and the uncertainty in the modelling strategy is therefore important. All flood inundation models work with discharge and water level as upstream, downstream and/or internal boundaries. Moreover, models which use the St. Venant equations need to specify two boundary conditions for both the beginning and end of the model

* Corresponding author.

E-mail address: f.pappenberger@lancaster.ac.uk (F. Pappenberger).

region. Uncertainties in the boundary conditions are very often neglected unless the model is used in a real-time forecasting environment (see e.g. [1]). However, in these cases, the inundation model is at the receiving end of a model cascade, in which the input uncertainty stems mainly from the uncertainties in the rainfall input and of the rainfall-runoff predictions. Hall et al. [2] present an example which includes uncertainty in the input for a steady state case. To the knowledge of the authors, no study has previously investigated the impact of uncertainty in dynamic discharge from upstream as the boundary condition of a flood inundation model, although a large number of studies exist in other flood related areas for example flood damage analysis [3,4], flood warnings [5], flood frequency [6], and peak discharges for historical floods [7].

A one-dimensional flood inundation model based on the St. Venant equations (such as HEC-RAS, ISIS, MIKE11 or SOBEK) requires the specification of two boundary conditions at the upstream and downstream boundaries. Lacking direct measurements of velocity, most modellers assume an upstream discharge hydrograph, providing a link between measured depth and cross-sectional average velocity, as given. This is surprising, as it is known that input discharge can dominate flood inundation behaviour and most discharge records are still derived in many cases from rating-curves. These convert measured water levels to discharge (see Refs. [8–13]), on the basis of a number of discharge measurements made over time at different levels or some model of the relationship between depth and mean velocity [14].

Several different sources of uncertainty can be listed which limit the accuracy of rating curves and their application to any particular event (see the extensive literature review by Pelletier [15]). There is uncertainty in the sampling of the *cross-sectional area* during floods. It might be very difficult to measure in certain locations of the river [16,17]. More problematic is that a river cross-section is far from being stationary [18]. Modifications of the geometry after flood events are common [19–24]. In many cases *velocity and assumptions of the velocity distribution* are combined with the cross-section characteristics to derive discharge. However, velocity temporally and spatially fluctuates in complex patterns, which will cause difficulties for any measurement technique [14,15,25–27]. A further source is the measurement *instrument* [28–30]. It is not only the discharge determination that is subject to various uncertainties, but also the *water level*. In the case of the traditional rating curves a *model structure* has to be assumed which connects water level measurements to discharge. A large number of equations exist, of which some are physically based and some empirical (see e.g. [25,31–34]). Non-contact and continuous methods may allow the measurement of more data points and therefore may make

assumptions of a model structure redundant. They further may be able to quantify *all hydraulic processes* such as hysteresis (see Refs. [13,27,35,36] and references cited therein). Other processes not always apparent from the data include: growth and decay of aquatic vegetation; ice; log debris jams and sediment transport [10,11,13,37,38] or sudden changes in area and hydraulic radius when overtopping occurs. Not mentioned so far is the *location* of the measurement instrument, which sometimes forces the modeller to additional computations to derive a representative discharge for the model region. Most flood hydrographs are derived from rating curves, and more sophisticated measurement devices such as Acoustic Doppler instruments, which are also prone to some of the uncertainties mentioned [39–41], are still rare.

Implementation of a model also requires the specification of roughness parameters for all reaches. The values of roughness, and the way in which it changes with depth, will need to compensate for the limitations of the model in representing the effects of geometry and floodplain infrastructure in energy dissipation at the scale of the model discretisation. Thus, reach scale effective values of roughness are needed (see e.g. [42–46]). Since these cannot be measured directly, they should also be expected to be subject to significant uncertainty.

There will also often be uncertainty associated with internal boundary conditions for the modelled reach. In most recent studies, the uncertainty associated with structural elements such as bridges is ignored, although they can have a significant impact on the flood inundation predictions. This is usually justified by the argument that such uncertainty will be compensated for by an appropriate effective value of surface roughness. However, there are a large variety of different ways of implementing such structures in different modelling packages and the modeller will have to choose from approaches of varying complexity, each of which might require different compensating effective roughness values.

In this paper, we extend previous studies on parameter uncertainty [46] with consideration of the uncertainties in the upstream, downstream and internal (bridge) boundary conditions, and uncertainties in effective roughness values, in the one-dimensional unsteady flow model HEC-RAS [47]. One-dimensional models do not truly represent two-dimensional floodplain flow [48]. However, it has been shown that, in some topographical configurations, the results of lower dimensional models often cannot be distinguished from more complicated models [49,50]. Although results may not always be transferable directly to more complex approaches (for example surface roughness is of less importance in three-dimensional flow models), lessons can be learned from these simpler approaches [51].

The uncertainty study in this paper is based on the generalized likelihood uncertainty estimation (GLUE)

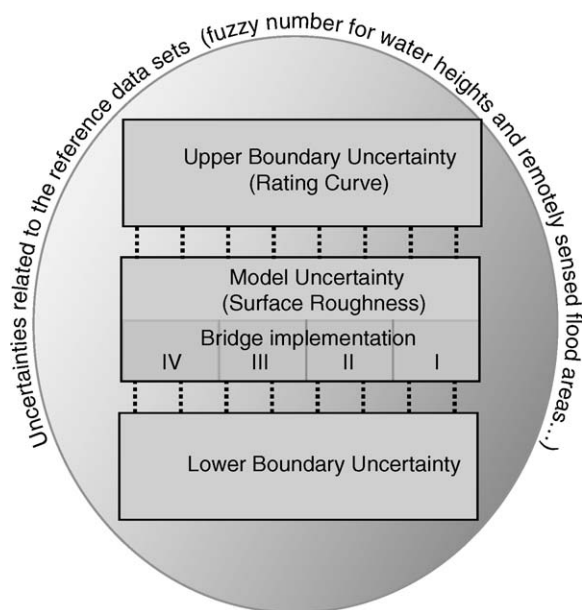


Fig. 1. Sketch of uncertainties considered in this paper. The uncertainty of the upper boundary is represented by a rating curve and cascades into the model. The model has several uncertainties such as the effective Manning surface roughness and various way of implementing the bridges in the model region. Also the uncertainties in the lower boundary have to be considered. The model is evaluated with various observed data. These data have additional uncertainties in the form of measurement and commensurability errors.

method [52] with data from the Alzette catchment (Luxembourg) extending the study of Matgen et al. [45].

In the first section of this paper, we propose a method to acknowledge rating-curve uncertainty with restricted available measurements. This is followed by an introduction of the model set-up and all available data (Section 3). The GLUE methodology and the evaluation measured used will be presented in section four. Finally the results will be discussed and conclusions drawn. Fig. 1 summarizes all the uncertainties considered in this paper.

2. Upstream boundary—quantification of uncertainty in rating curves

Several studies have attempted to quantify the uncertainty in stage–discharge rating curves [10,22,30,32,33,53]. The difficulty of the problem is increased when, during a flooding situation, the rating curve is extrapolated beyond the measurement range. The flow regime changes dramatically as soon as the river is overbank [25] and thus different rating curve behaviour can be expected [13,54].

Several models have been published amongst them the most widely used are the power law equations [10,11,32,33].

$$Q = c(a + h)^{\alpha} \quad (1)$$

Q is low ($\text{m}^3 \text{s}^{-1}$), H is water level (m) and a , c and α are calibration parameters.

And the Manning equation [10,11,31–33,55]:

$$Q = \frac{1}{n} R^a S^{0.5} \quad (2)$$

R is hydraulic radius, n is manning surface roughness, a is exponent of the hydraulic radius [measured between by ~ 0.5 and ~ 0.7 by 56] and S is slope.

Other approaches include the use of weir equations [39]; theoretical relationships derived from flume experiments [57,58]; neural networks [59–63]; or M5 regression trees, which approximate the data by a set of linear equations [63]. Furthermore, Chiu and others [14,27,64] have developed methods based on the velocity distribution of rivers.

The power law equation and the Manning equation will be applied in this study and therefore will be also used to extrapolate for flood levels (similar to the slope–area method of USGS and the study by Dawdy et al. [65]). In this approach, the two equations are used as equally valid representations of this extrapolation, because it is not possible to distinguish between the different model structures [66,67]. This is very much in line with the suggestion of DeGagne et al. [68] who advocate a multiple model approach for extrapolation. This non-identifiability is driven by poor data availability. Rantz et al. [55] argues that ‘the rating should not be extrapolated beyond twice the largest measured discharge except as a last resort’. However flood inundation is concerned with predicting extremes, therefore one might be forced to do this, otherwise no predictions will be made at all. Potter and others [69,70] use the amount of extrapolation needed as an indication for uncertainty.

In this example, only eight data points could be found in the records which have been used to establish the rating curve. The errors reported in the literature for the discharge measurements are from 1.8% to 8.5% with a large number of studies finding around 6% (see literature quoted in [13,71]). The errors which are quoted for the water level are around 3–14 mm, although this seems to be a very low value (see literature quoted in [13]). Due to the limited amount of data an ad hoc approach to fit the rating curve is proposed. The approach is a modified version of the Multicomponent Mapping by Pappenberger and Beven [72] in which data uncertainty has been expressed by pyramidal frustrums. To each data point a pyramidal frustrum of the height one has been assigned (Fig. 2). The height of the frustrum can be seen as a two-dimensional fuzzy membership function and thus the minimum distance from a line to the middle point gives a membership value. These values have been chosen according to the literature review and communication with the company responsible for taking these measurements.

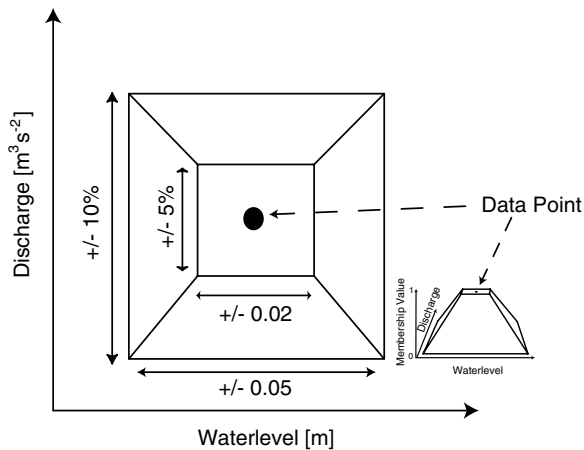


Fig. 2. Illustration of pyramidal frustrum at data points. An error in the form of a pyramidal frustrum can be quantified around each measurement. A membership value of 1 is given to each rating curve which is within $\pm 5\%$ of the measured discharge and simulates ± 0.02 m of the measured water level. Any rating curve with an error of more than 10% on discharge and more than 0.05 m on water level receives a membership value of 0. The values in between are scaled accordingly to the surface of the frustrum.

Parameter combinations for the rating curve functions have been obtained by Monte Carlo sampling and the membership values at each data point evaluated. A parameter set has been accepted to fit the current data set if a multiplication of all individual membership values is larger than 0. The combined membership value can be seen as likelihood measure for this particular realisation. This procedure is known as the generalized likelihood uncertainty estimation (GLUE) [43,52]. It allows that it might be difficult to distinguish between many parameter sets which give acceptable predictions (behavioural parameter sets). From this concept a likelihood measure can be computed for each parameter set which then can be used to compute uncertainty bounds in prediction, including predictions beyond the range of the data points available.

3. Description of the study area and model implementation

The study reach is on the river Alzette (Luxembourg) and is approximately 10 km long with an average slope of 0.001 m/m (Fig. 3). The region is modelled by the one-dimensional HEC-RAS model, which is based on the St. Venant equations. The equations are solved with an implicit finite difference scheme using a modified Newton–Raphson iteration technique [73,74]. This type of model is frequently criticised for not representing the real dynamics of floodplain flow. The problems of applying alternative higher dimension approaches are often ignored in such an argument. Models of higher dimension require an increased amount of parameters

and do not avoid the need to specify effective parameter values. If no more calibration data are available, the use of a higher dimension model will not in itself constrain the uncertainty. Moreover, many more complex models cannot be run on the spatial domain required. Key findings and methodologies can be easily developed and presented with one-dimensional models—and then be transferred if required.

The comprehensive data set includes pre-flood and flood SAR images, continuous water level measurements upstream and downstream of the river reach, surveyed high water marks and GPS control points of the maximum flood extent. A set of photographs taken during the flooding event (see Fig. 7) is also available. Two ERS-2 SAR and Envisat ASAR scenes acquired during the rising limb and at the peak discharge respectively cover the flooded area at two distinct stages of the event (a further detailed description which includes the pre-processing of the different data sources is given by [45]).

The geometry of the study region has been represented by 74 cross-sections which have been surveyed beforehand via a combination of field measurements (inside the river channel) and aerial LIDAR scanning (across the floodplain). The model has been simplified and only three different roughnesses are used to represent the left floodplain, the channel and the right floodplain. It would have been desirable to include further information of geometric uncertainty as it can have a major impact on flow paths and inundation extent [42,75]. However this is beyond the scope of this study. There are two road bridges within the river reach (in this text referred to as *bridge location* followed by an Arabic numeral), which cross the entire floodplain. Fig. 3 shows the location and images of the two bridges. The low cord of the first bridge is ~ 4.2 m above the river bed, the road is ~ 4.9 m above the river bed and the culvert has a width of ~ 12.8 m. The low cord of the second bridge is at ~ 3.25 m (above river bed), the road at ~ 4.75 m (above river bed) and the bridge culvert has a width of ~ 12.8 m. The bridges have been modelled with four different approaches (in this text referred to as *bridge implementation* followed by a Roman numeral).

In the first approach the bridges have been completely ignored (*bridge implementation method I*). In the second approach the bridges have been approximated by using additional effective Manning roughness coefficients (*bridge implementation method II*). In this case the model region (without downstream boundary) thus has five different Manning roughness values. In the third case the bridge geometry has been included in the cross-section and the wetted perimeter and area changed accordingly (*bridge implementation method III*). In the last representation the bridge is treated as an interior boundary condition and the different types of flow are represented by a combination of free and submerged rating curves (*bridge implementation method*

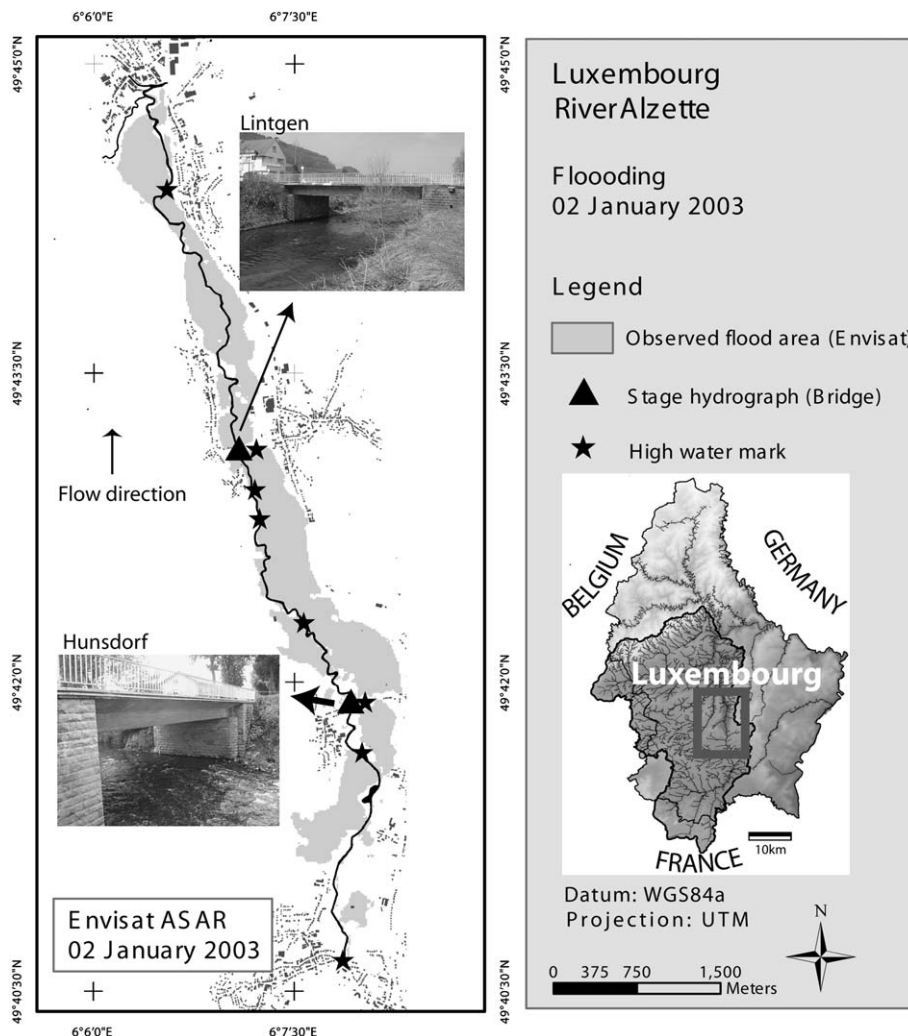


Fig. 3. Description of the Alzette catchment in Luxembourg. The location of the two bridges as well as some of the evaluation data (such as the location of the maximum water level measurements) available are marked on the figure.

IV). Four different parameters are used to specify the properties of these rating curves. At low flow Yarnell's equation is used, which needs the specification of Yarnell's pier loss coefficient. For pressure flow a pressure flow coefficient has to be specified and for submerged flow a weir equation parameter is given. Moreover, this family of rating curves needs the specification of a Bridge Entrance Loss Coefficient. The exact derivation and type of equations applied can be found in Barkau [73] and the parameter range has been chosen according to the suggested values in the user manual. The ranges of all parameter values have been based on estimates of their physical counterparts although it is recognised that it is effective values appropriate to the particular implementation of the model that are required [46,76–78]. One-dimensional flood inundation models are currently the only type of flood inundation models which have a large variety of different bridge implementations (as well as many other structures) implemented, used and tested.

4. Downstream boundary

Unfortunately, the uncertainty for the downstream boundary could not be implemented in the same way as the upstream boundary due to the 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. This boundary condition is only an approximation of the true rating curve but has the advantage to pass waves downstream [73]. The uncertainty in this boundary condition has been evaluated by allowing additional effective roughness values in the last two cross-sections, which are varied independently from all other cross-sections.

Table 1 lists the parameters which are used as calibration parameters in each model implementation. The parameter ranges have been deliberately chosen to

Table 1
Model parameters and sampling range

Parameter	Sampling range
<i>Input rating curve</i>	
Method I: Manning equation	
Slope	0.001–0.008
Manning roughness	0.01–0.15
Wetted perimeter exponent	0.4–0.6
Method II: Power law	
<i>C</i>	8–16
<i>A</i>	–0.3–0
Alpha	1–2
<i>Main model region</i>	
Manning roughness channel	0.01–0.5
Manning roughness left floodplain	0.01–0.5
Manning roughness right floodplain	0.01–0.5
<i>Bridges</i>	
Method I (bridges ignored)	
None	–
Method II (Manning roughness at bridge)	
Manning roughness	0.01–0.5
Method III (include geometry)	
Manning roughness	0.01–0.5
Method IV (complex bridge)	
Yarnell's loss coefficient	0.9–1.25
Pressure flow loss coefficient	1.2–1.8
Bridge entrance loss coefficient	0.3–0.6
Weir coefficient	2.0–3.2
Bridge entrance loss coefficient	0.3–0.6
<i>Downstream boundary</i>	
Manning roughness (channel and floodplains)	0.01–0.5

exceed ranges seen as physical acceptable to reflect the discrepancy between effective and 'real' parameters.

5. Description of the likelihood measures used within the generalized likelihood uncertainty estimation framework (GLUE)

The generalized likelihood uncertainty estimation (GLUE) procedure has been outlined in detail elsewhere [52]. We will provide only a brief description. Empirical results of studies in various fields of environmental modelling (for a summary see [79]) show that many different sets of parameters and model structures can fit evaluation data acceptably well and therefore, it may be impossible to decide between them as useful predictors of the system (the 'equifinality' concept of [80]). However, some realisations can be classified on the basis of the evaluation data as non-representative of the system or 'non-behavioural' and thus can be excluded from further analysis. GLUE has been criticised for the subjectivity of this distinction [5,81], but as Freer et al. [82] pointed out, subjectivity is unavoidable if no consistent

error structure can be assumed. Within GLUE a likelihood weight is assigned to each of the behavioural parameter sets, according to how well that parameter set has performed during a calibration period. The likelihood weights of the behavioural data set can be used to assess parameter sensitivity and to compute confidence limits on any model predicted variable [52,83]. This method requires the modeller to specify a prior sampling distribution of the parameter space including any known interrelationships between the parameters. These specifications are not always straightforward, because the model parameters are effective parameters, which will be affected by all sources of error in the modelling system. The sampling strategy will inevitably influence the result and is governed also by the available computer power. It is a particular problem with parameters that can be allowed to vary spatially in a distributed model, since this will result in very high dimensional parameter spaces and the possibility of quite different spatial patterns giving similar predictions for good physical reasons. The choice of an appropriate likelihood measure is very important in the GLUE methodology. This will depend on the type of observational data available for a given application. It would be expected that the analysis is strongly influenced by the choice of the likelihood measure, though this is not necessarily the case as Freer et al. [84] have proven with different likelihoods in an application of the rainfall-runoff model TOPMODEL. An interesting variation to this approach has been reported by Aronica et al. [42], Franks et al. [85], and Blazkova and Beven [86] who have successfully used fuzzy evaluation criteria. In this paper, we follow a similar approach in which the limits of acceptable error for each observation are set before any model runs are made, taking account of any differences in meaning between observed and model predicted variables (the *commensurability* error, see Beven [66]). Table 2 summarises the evaluation measures used in this study for the different types of observational data described below.

Several data sets are available for model calibration as described above (Table 2 summarizes the evaluation measures used in this study). However, each of these sets has its own characteristic and is derived in different ways. Therefore, it is necessary to describe the application of each set individually. All data sets have been used in a fuzzy membership framework. In this way one cannot take only account of the measurement errors but also for the commensurability errors [66] (although the latter might be difficult to quantify).

5.1. Envisat ASAR and ERS-2 SAR

A number of previous studies have applied SAR (Synthetic Aperture Radar) imaging techniques (pixel size 12.5 m) for estimating the extent of flood inundation [46,50,87,88]. However, the classification of flooded and

Table 2
Evaluation data of Alzette catchment

Name	Description
Envisat ASAR	Synthetic Aperture Radar image from satellite platform, Alternating Polarization mode with VV and VH polarizations, 12.5 m pixel, image classification through histogram thresholding
ERS-2 SAR	Synthetic Aperture Radar image from satellite platform, VV polarization, 12.5 m pixel, image classification through histogram thresholding
Photographs	Oblique photographs taken from exposed ground positions
Maximum water level	Post-flood measurements of maximum water levels
Hydrographs	Continuous stream stage recording at bridge location 1 and 2, pressure transducers measure the backpressure in orifice lines set into the stream channel
Travel time	Travel time of flood peak from upstream to downstream boundary

non-flooded pixels is not always certain. Wind effects, vegetation and wet-lands make an accurate delineation difficult. In this paper a simple threshold approach has been used to quantify a fuzzy membership function of flooding along each cross-section which expresses the uncertainty of flooding extent (see [45]). Profiles of pixel values at several cross-sections of the river floodplain (average floodplain width 300 m) are drawn and confronted with the GPS control points of the maximum lateral flood extent. This allows determining the threshold value of the radar backscattering coefficient for the binary classification of flooded and non-flooded pixels. This approach provides the reference flood map. To reflect our lack of knowledge about the real flood extent, the threshold value of the radar backscattering is slightly changed to delineate other plausible flood maps. Thus, for each cross-section, a membership function has been derived for the left and right floodplain reflecting the probability of flooding based on our analysis. Fig. 4 illustrates this methodology which has also been applied

in similar ways by Pappenberger et al. [46] and Romanowicz and Beven [44].

This membership function is also constrained in the direction of the channel to account for inundation predictions that are too low. However, the membership functions have been independently built for the left and right floodplain inundation extent. Thus the membership functions for the left and right floodplain are not necessarily equivalent to each other and a predicted water level may achieve different likelihood values on the left and right floodplain. The difference can come from ignorance of processes (e.g. two dimensionality of floodplain flow), error in the DTM or the inundation image and has to be accounted for in the model evaluation process. Therefore, the union between the two membership values on each cross-section has been computed and used as performance value at each cross-section. A similar effect can be observed in the downstream direction with models showing significant differences in performance values at consecutive cross-sections. This

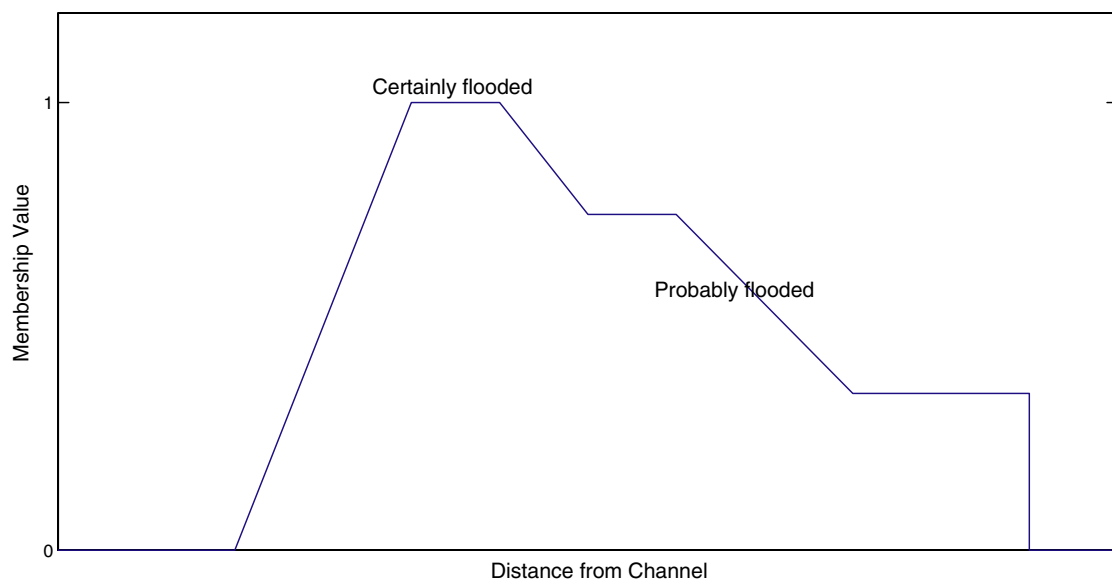


Fig. 4. Membership function to compute the performance of inundation predictions at each cross-section. This function includes assumptions of the probability of pixels being flooded. The distance is given from the centre of the river and the membership values decrease towards the edges of the floodplain.

error has been taken into account by applying a moving average (size three) to the evaluation results. The size of this window has been determined subjectively by visual analysis of the data.

5.2. Photographs

The photographs of the flooded area have been mapped back on the Digital Terrain Model. The edge of the inundation area was determined by visual comparison of the photographs (taken at an oblique incidence angle) with geo-referenced aerial photographs of the area (taken at vertical incidence before the event). Geometrical features that were recognised on both images allowed the delineation of the edge of the flooded area. Since the delineation of the edge of the flooded area is based on the recognition of geometrical features, the accuracy of the shoreline derived from the photographs obviously relates to the presence or absence of such features. In the worst case, no geometrical features are available and in these areas an error of 30 m has been assumed. Similar to the SAR, a fuzzy membership function has been derived and computed. Comparisons with the model results take account the time at which the photos were taken. The exact derivation of performance measures and the processing of the remotely sensed images is explained in more detail by Matgen et al. [45].

5.3. 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. This commensurability error is a combination of the error in the DTM and the GPS, which as been used for this measurement. However, the 1 m limit is very large and the geometry should be refined to decrease this error in future studies.

5.4. Hydrographs (water level)

The hydrographs have been measured at the two bridges and are analysed via the Multicomponent Mapping method [72]. This method enables the specification of subjective error bounds around the measurements and the computation of a performance value. In this case a maximum error of 1.5 h has been allowed in the time direction and a maximum error of 10% for the water level predictions. These values have been chosen to incorporate the commensurability (location of the water level measurement in the bridge structure and the quasi one-dimensional representation in the model) as well as the measurement error.

5.5. Travel time

Unfortunately, in this case it is not possible to use the downstream hydrograph as boundary condition or evaluation data, because a small tributary flows into the main river before the last cross-section. However, the confluence does not peak at the same time as the main river, so it was possible to take the travel time of the peak flow. 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 different datasets can be used to constrain the uncertainty in model predictions and to evaluate the importance of different parameters in producing acceptable or behavioural models. In this paper, the Sensitivity Analysis based on Regional Splitting (SARS-RT) has been employed. SARS-RT is based on the construction of Regression Forests and allows the determination of the importance of parameters and parameter combinations in different regions of the model space. It is described in detail in Pappenberger et al. [89]. This method can help to identify the sensitivity of a model parameter towards the model results and thus helps to identify important and influencing parameters. The more important a parameter is the higher is the importance value.

6. Results

The results of the rating curve uncertainty analysis are first presented, followed by an analysis of the influence of this uncertainty on the flood inundation model. Then the uncertainty analysis of the bridge structure is presented. Finally, the results of all evaluation criteria are drawn together.

6.1. Rating curve uncertainties

The uncertainty in the rating curve will directly affect the estimate of discharge which is used as inflow into the model region. In Fig. 5, the scatter plots of the behavioural parameter set of the power law equation is presented. Each dot in this figure presents a rating curve, which is behavioural at each data point. All parameters show a large range of equifinality and a similar plot can be drawn for the Manning equation. In the Manning equation the Manning roughness and the channel slope have a power law relationship, which is seen in Eq. (2). The Manning equation has been developed for uniform flow and it cannot be expected to hold fully for non-uniform conditions or as flow goes overbank [25]. Therefore, the true uncertainty bounds are probably larger than presented in this plot—especially in the light of all uncertainties quoted in Section 1. It would therefore be

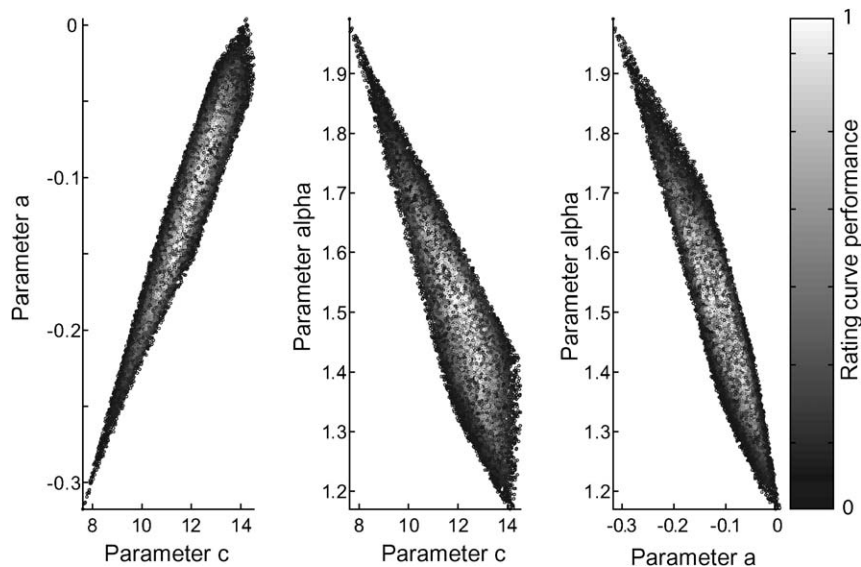


Fig. 5. Scatter plot of parameters of polynomial equation (see Eq. (1)). Each dot represents one set of parameters for the rating curve of Eq. (1). The parameters are plotted against each other and the performance, which each rating curve achieves in comparison to the measured data, is grey coded. The higher the value (lighter the colour) the better the overall performance (maximum of 1).

expected that the uncertainty bounds in the rating curve will widen even further at the top end of the rating curve.

All parameters have been estimated for flow conditions considerably below the peak of this flood event. Hence, parameter sensitivity and uncertainty will most probably not hold at extrapolation (see Section 1). However, the latter is frequently done. Figs. 6 and 7 show the

effect of the uncertainty in the rating curve by plotting the envelope of all behavioural rating curves and the spread of the inflow hydrograph, respectively. The lower boundaries of both equations are nearly identical and the upper boundary of the power law is higher than the Manning equation. At the upper limit of the discharge range considered the uncertainty range of the

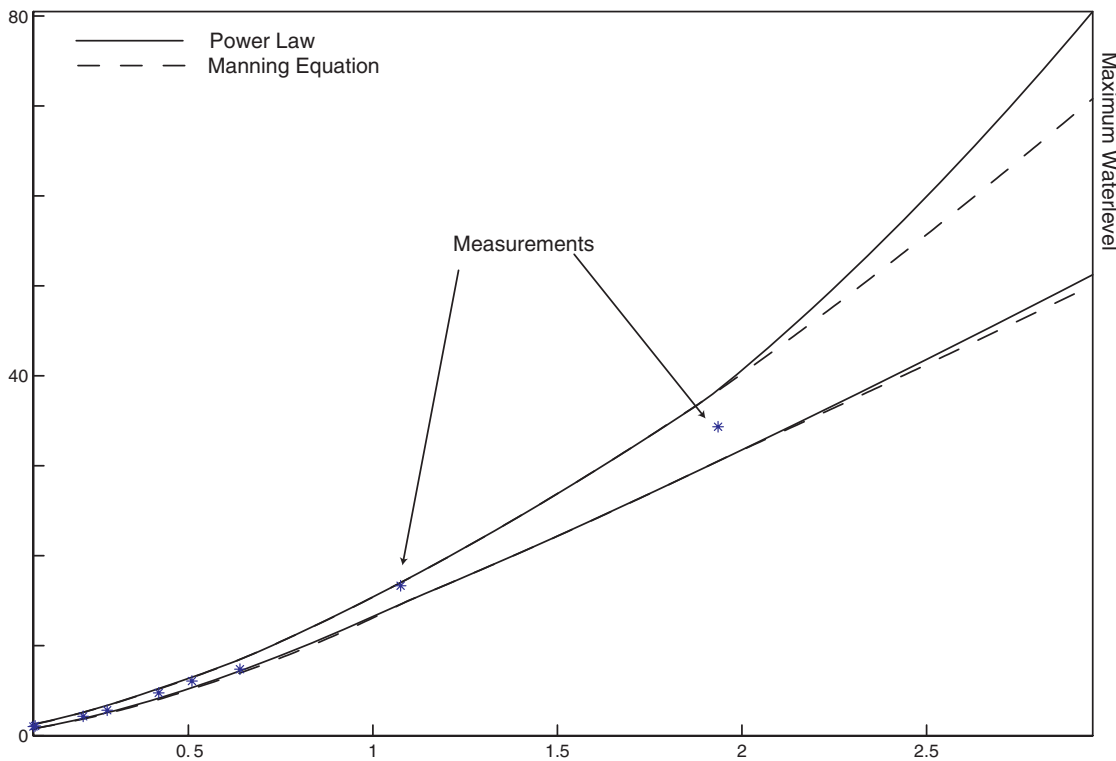


Fig. 6. Envelope curve of rating curves of the Power law and the Manning equation. The envelope is taken from the 5% and 95% percentiles of all rating curves, which achieve a performance greater then 0.

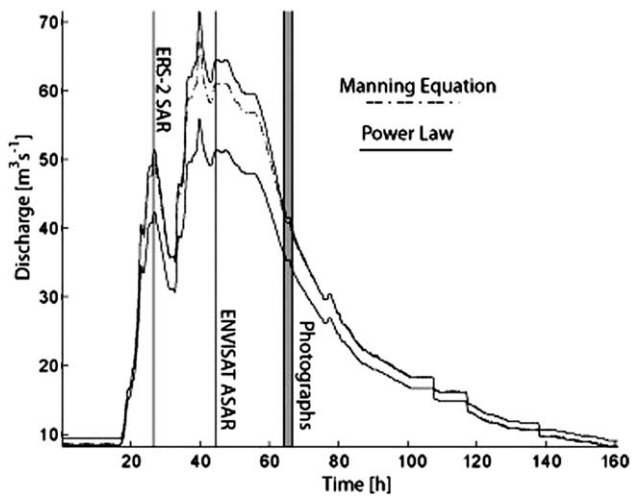


Fig. 7. 5% and 95% percentile of the inflow hydrograph for the Manning equation and the power law. The figure illustrates the range of inflow created through the uncertainty of the rating curve. The figure also shows the timing of the remotely sensed evaluation data, which are available for this analysis.

Manning equation is roughly 11 and 15 $\text{m}^3 \text{s}^{-1}$ for the power law. This is considerable knowing that the mean maximum discharge for this event is around $60 \text{ m}^3 \text{s}^{-1}$. Such a high value can have significant impact in addition to the other sources of uncertainty investigated in

this paper. In many countries flood warning levels are attached to certain water levels or discharges and in such cases an uncertainty of this magnitude may be considered as very high by decision makers. This already indicates the importance of the uncertainty in rating curves for local inundation predictions and in the following section the impact on the flood inundation will be investigated. The uncertainty analysis of the input rating curve is used in the further modelling exercise.

6.2. Sensitivity of water level predictions to input uncertainty

The influence of the input uncertainty of the rating curves on water level predictions has been evaluated by employing the SARS–RT analysis technique. The sensitivity analysis has been performed for the timing of the three remotely sensed images, and the uncertainty of the rating curve has been included by using the inflow discharge at the time of each image as an input parameter. However, this is only a partial reflection of the sensitivity, as one expects a shift of the importance values over time which is not considered. All figures in this paragraph only plot the major effect of each parameter and no effects due to parameter interaction or correlation (see [90] for details). Fig. 8 presents the median of the importance measure of the SARS–RT analysis at the

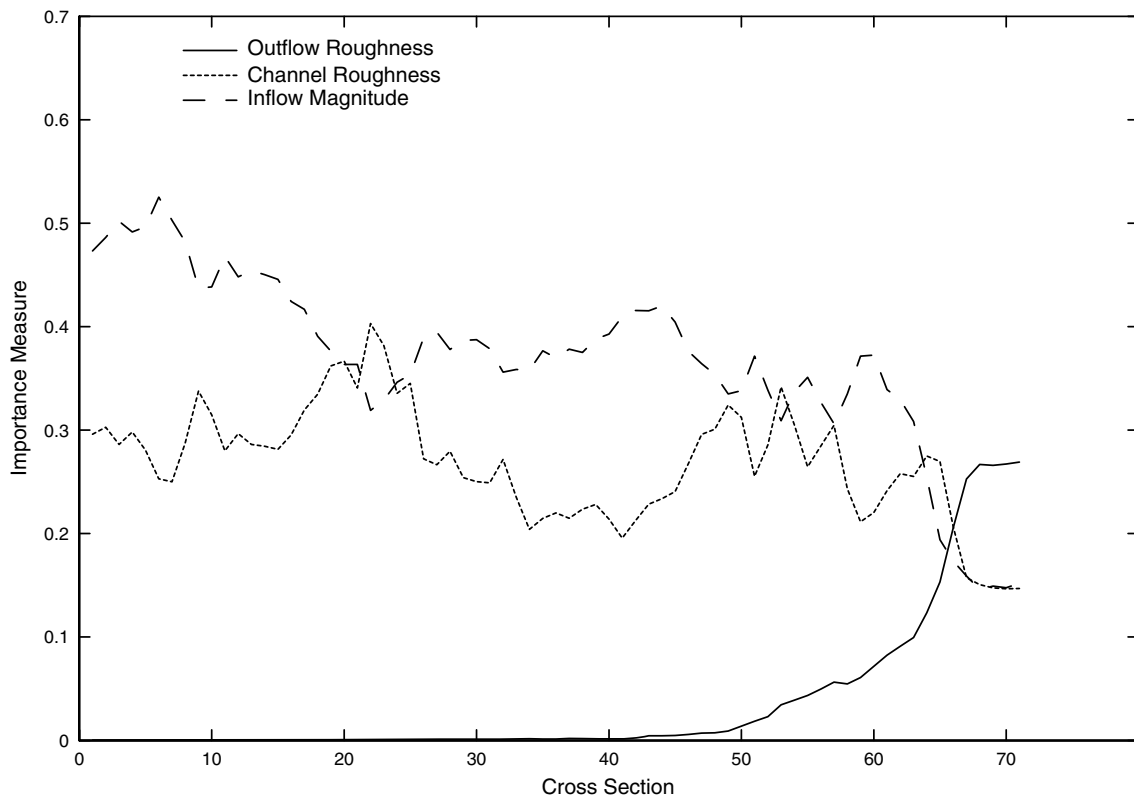


Fig. 8. Median of the SARS sensitivity analysis along all cross-sections for bridge implementation I at the time of the ERS-2 SAR image. The lower the importance measure the less sensitive is the parameter.

time of the ERS-2 SAR image and using bridge implementation I. In this plot, the sensitivity of all model parameters plus the inflow against the water level in each particular section has been tested. The higher the value on the ordinate the more sensitive is the parameter towards the water level at this particular cross-section.

In this figure only the three most sensitive parameters are presented and it shows that the model region is mainly influenced by the Manning channel roughness and the inflow magnitude. The value of the Inflow magnitude importance measure declines towards the end of the reach and drops sharply at the last 10 cross-sections. The channel roughness remains important over the entire model region (this is equivalent to the findings of Hall et al. [2] with a raster based model), but is less important than uncertainties in the inflow magnitude. The floodplain roughness is not important at any cross-section of this model and hence is not shown in Fig. 8. At the downstream sections, the outflow roughness becomes increasingly important. The influence of the downstream boundary stretches over more than 20 sections. The large influence of the upstream and downstream boundary on a large amount of the cross-sections differs from the results by Hall et al. [2]. In their paper, the influence of the uncertainty in the inflow is small and changes little along the river. However, this publication uses a different model (Lisflood) and varies different factors. Moreover, only a $\sim 5\%$ uncertainty on the discharge magnitude is considered and the inundation is computed under steady state conditions. Moreover, the modelled reach does not contain the structures that have been included here in the representation of the Alzette. The fluctuations of the difference between the importance of the channel roughness and the inflow magnitude are probably due to changes in channel width. This indicates that another major factor (geometry uncertainty) does have an important effect on parameter sensitivity and uncertainty [75].

Bridge implementation II (not plotted here) shows very similar sensitivities to implementation I and indicates only minor and local additional significance of the roughness at the bridge location. This suggests that one local roughness value does not have a large influence on the entire model reach and has restricted effects on a local scale. The lower downstream condition influences the model region again for roughly 20 sections.

The influence of the downstream boundary for bridge implementation III and IV reaches further upstream (~ 30 sections). This illustrates the non-linear effect of the bridge implementations as there is no difference between the lower cross-sections of the different bridge implementations. Moreover, the constriction of the river at the bridge seems to have significant impact on the influence of the uncertainty at the upstream boundary. Fig. 9, which has again only the three major effects plotted, demonstrates this effect and shows a sharp

fall in the importance at the bridge location. The channel roughness becomes influential after the first bridge, which is also true to a lesser extent for bridge implementation III. This indicates that the flow is more controlled by the constriction of the river than the bridge parameterisation.

These results show that the relative importance of uncertainties is very dependent on the hydraulic conditions in the reach and that the influence of the boundary conditions is so large that they have to be considered as additional calibration factors.

6.3. Sensitivity of water level predictions to bridge implementation

It is possible to examine the bridge effects more directly, by subtracting the water levels at each cross-section from the water level at the first cross-section (or any other reference water level). In Fig. 10, this is shown for the time of the ERS-2 SAR image, in which the envelopes of this analysis, split by the four different bridge implementations, are plotted. Bridge implementations I and II do not show any difference in their behaviour, which suggests that the application of a single different roughness value at one cross-section has a minimal effect. However, the other two implementations (III and IV) show the potential to create large backwater effects at the first bridge and no large variations are shown at the location of the second bridge. The difference is due to the size of the bridge opening, which is about 50% larger at the second bridge compared to the first bridge. The back-water effect will be mainly due to the parameter choice for each bridge implementation. Implementations III and IV heavily constrain the cross-sectional area of the flow. A more extreme choice of Manning surface roughnesses for bridge implementation II might have created a similar effect.

This difference can be also demonstrated on the two stage hydrographs, which have been measured at the bridge locations. Fig. 11 shows the hydrograph evaluation measures (average over all time steps) at each bridge site expressed in terms of the numbers of simulations in different classes of performance. Bridge implementations I and II are very similar and do indicate more behavioural model realisations at the upper end of the distribution. Model simulations with bridge implementations III and IV result in far fewer behavioural models. Bridge implementation IV even fails to achieve any behavioural parameter sets at the location of bridge one. This is partially due to the fact that neither pressure flow nor bridge overtopping occurred in this event (the deck was easily cleared) and bridge formulation IV seems to constrain the through flow to a far too large extent. This of course questions whether the more complex bridge formulations can be distinguished from the simpler ones if none of the above flow

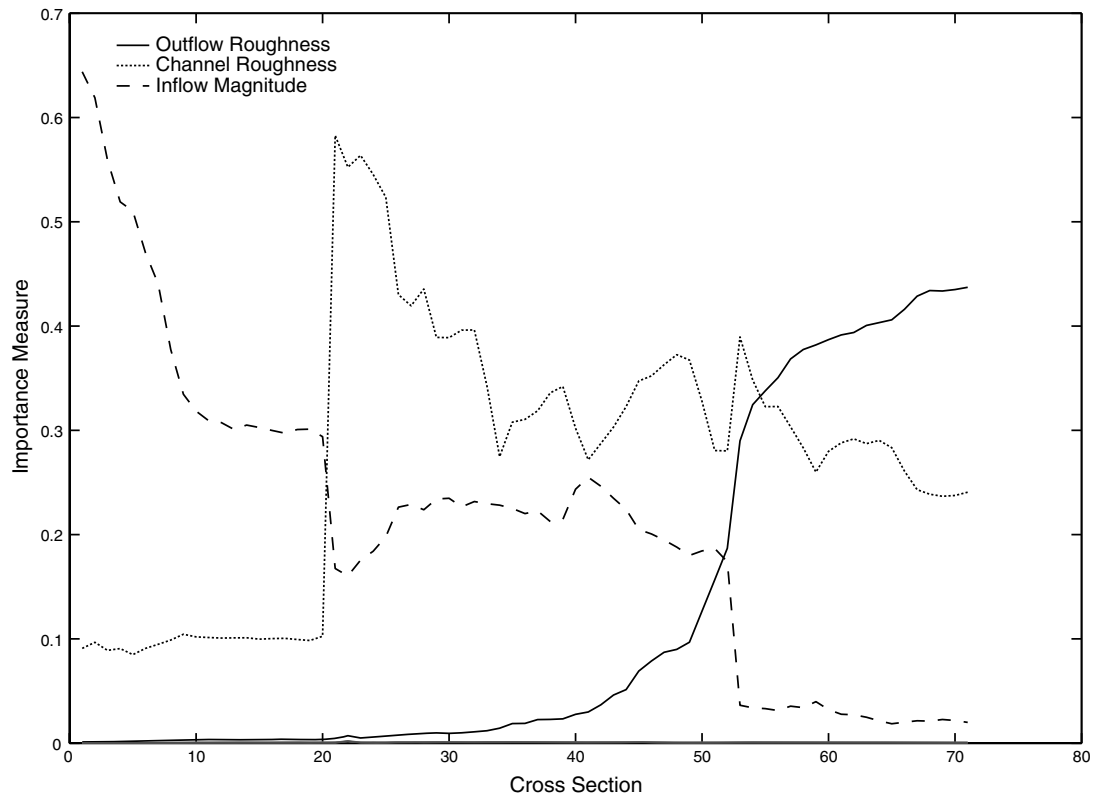


Fig. 9. Median of the SARS sensitivity analysis along all cross-sections for bridge implementation IV at the time of the ERS-2 SAR image. The lower the importance measure the less sensitive is the parameter.

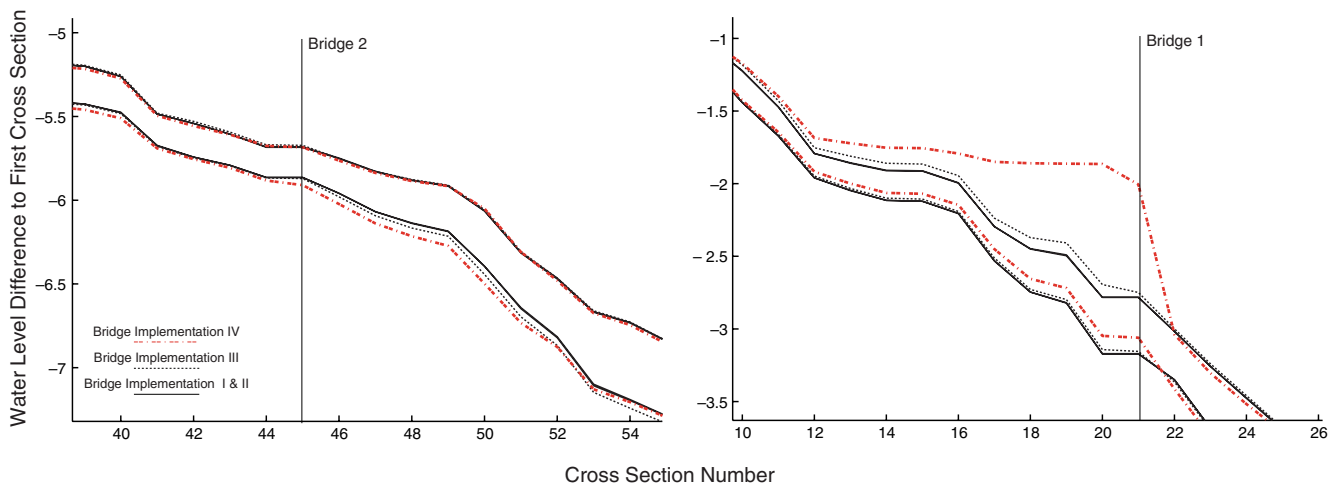


Fig. 10. Range of predictions of difference of water levels to first cross-section over all behavioural models at the time of the ERS-2 SAR image. The difference between the water levels predicted at each cross-section (numbered on abscise from upstream to downstream) and the first cross-section is computed for each model realisation. The 5% and 95% envelope curves of all these computations are plotted on this figure.

regimes occurs. In other words, the quality and quantity of data prohibits an entirely fair comparison.

6.4. Spatial and temporal shift of behavioural surface roughnesses

Such a comparison of distributions can be also used to investigate changing behaviour in Manning surface

roughness values. It is possible to compute the cumulative distribution function (cdf) at each cross-section for the Manning surface roughness based on the performance of the model as evaluated with the fuzzy membership functions at each cross-section. These cdfs can be compared to each other e.g. by computing correlation. For example it is possible to compute the correlation between the cdf of cross-section 50 with the cdf of

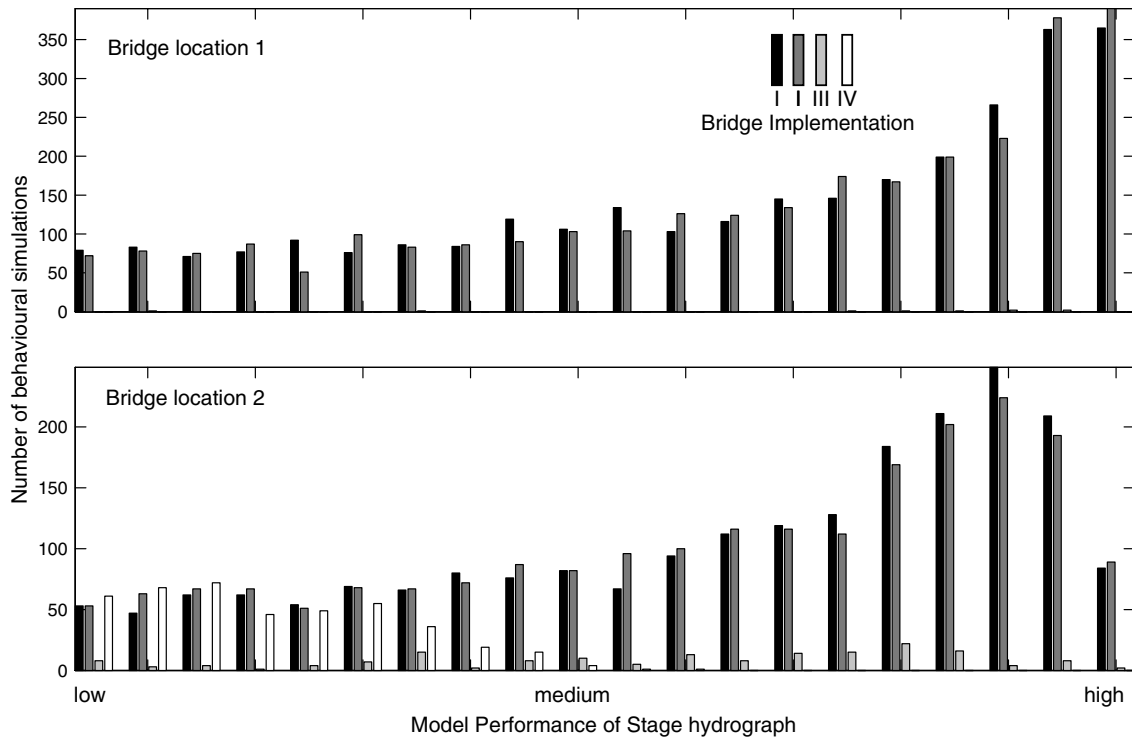


Fig. 11. Histogram of the of the model performances of the stages at the two bridges against the number of behavioural models after evaluation. The results for all four bridge implementations are shown.

cross-section 51. The same can be done in time in comparing performance for the different SAR-derived inundation maps. The result of such an analysis is shown in

Fig. 12 for both the space and time variations. In the upper subplot of Fig. 12, the roughness cdf at each section has been compared to the cdf of the following sec-

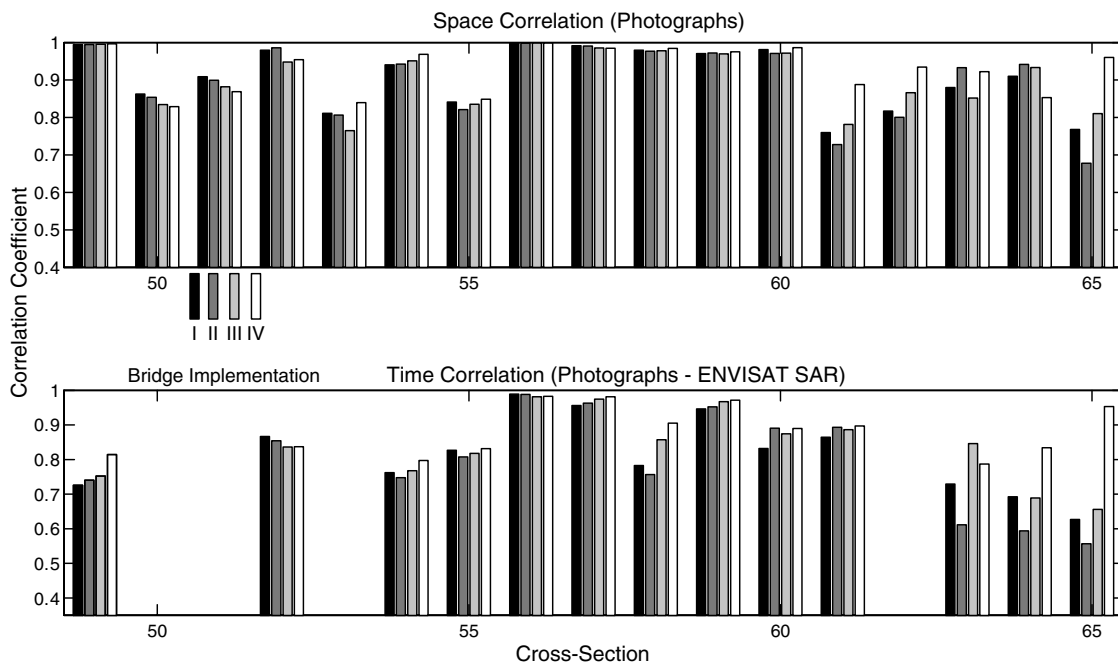


Fig. 12. Correlation between the cumulative distribution functions (cdf) in spatial and temporal direction. A cdf has been computed at each cross-section and for each remotely sensed evaluation picture. The correlation between these cdf's has been computed and plotted for each bridge implementation.

tion. Large values means that there is a strong correlation between the cdfs. In this figure, a significant change in distribution suggests that different Manning roughnesses are required for those cross-sections. For example in Fig. 12, the distribution at cross-sections 50 and 51 differ and thus it might be possible that different effective Manning roughnesses have to be applied. In contrast a value of close to one is computed for cross-sections 56 and 57, which seem to suggest that their distributions are similar and no additional local roughness values are required.

However, such conclusions should be only used as a guide since the results will depend on the evaluation measures chosen with the associated difficulty of assessing appropriate limits of acceptability for all the available observational data. What is clear is that the model is influenced by a complex interactions of parameters, which will make it difficult to pin-point how parameter values should be changed locally [71]. This becomes especially clear at the cross-sections 61–65. The bridges are located several kilometres upstream of these sections and nevertheless the similarity of the distributions of the four bridge implementations varies at these particular sections. This may be the result of an interaction between the downstream boundary condition and the amount of water which can pass through the bridge structures. A more consistent pattern between the cross-sections exists for the cross-sections 50–55. This suggests that the model performance may be improved by assuming a separate effective Manning Roughness value for these sections, despite the fact that the real surface cover and channel morphology at these sections do not differ significantly from the previous or later sections. Another reason for this grouping could be an especially high spatial correlation of the performance measure (it has been difficult to determine the shoreline in this area) and not of the surface roughness itself. In fact we would like to link this disaggregation to local characteristics of land-use and floodplain infrastructure. In this case, the density of houses is slightly increased in this area. This exhibits an interesting discrepancy as, on the one hand it has been shown by Werner et al. [91] that floodplain roughnesses based on land-use do not improve overall model performance and on the other hand that our model seems to suggest exactly such adjustments to improve model performance. As Werner et al. [91] pointed out, the interaction between the different factors influencing model performance is highly complex and non-linear.

The lower subplot of Fig. 12 provides a comparison between different time (non-existing bars indicate missing data either at the time of the aerial photographs or the ENVISAT ASAR). The reader is reminded that the photographs have been taken during the recession and the radar image at the peak (Fig. 7). Some of the cross-

section values seem to argue for effective roughness values that vary, whereas others do not allow this conclusion for the two time step images available. This variation could be due to different water levels as these images are taken at different times or due to other processes which are connected to the difference between peak flow and recession. Comparisons between the other combinations of remotely sensed images lead to similar conclusions. Again, the correlation between water depth and surface roughness is inconsistent, as it seems to be high at some cross-sections but not at others.

6.5. Comparison of all model results

The methods of the previous section can be applied to all other evaluation criteria as well. In this chapter, the analysis of all evaluation criteria is drawn together and an overall picture of model behaviour is presented.

In Table 3, the number of behavioural model realisations for each evaluation criteria is presented. The last column shows the number of numerically stable model runs from the total number of runs (~20,000 for each bridge formulation). This table shows that the model simulations which contain the bridge as a geometrical structure (implementation method III), have a very low number of numerically stable solutions. This is probably due to the rapidly changing conveyance properties at this location. Unfortunately, it was not possible to acquire a better explanation, why so many model simulations are unstable. This would require the source code, which was not available at the time of this research.

The different model structures show a significant difference in the number of acceptable runs at the two stage measurement sites (bridge locations 1 and 2) and the one measure based on maximum elevations. This maximum elevation measure is shortly before the first bridge structure and thus can be explained by the previous analysis of the bridge behaviour. Of further interest is the time of the peak at the downstream boundary. From Table 3 it becomes apparent that bridge formulation IV has a larger number of 'good' runs and further analysis seems to suggest that this is connected to backwater effects at the first bridge. However, previous analysis suggests that the timing of bridge formulation IV is not very good as it does not satisfy the stage hydrographs at the bridges. In contrast the first two bridge formulations have a large number of behavioural simulations for the stage hydrographs.

Table 3 also displays the total number of accepted model results, which fulfil all the criteria listed in this particular table. A similar table can be shown for the spatial inundation images (Table 4). In this table, the number of the behavioural runs at each inundation image and for each bridge implementation is listed.

Table 3
Number of behavioural runs for all likelihood measures beside inundation extent

Bridge implementation	Travel time	Max water level location 1	Max water level location 2	Max water level location 3	Max water level location 4	Max water level location 5	Max water level location 6	Max water level combined	Stage Hunsdorf (bridge 1)	Stage Lintgen (bridge 2)	Combined number of behavioural runs	Total number of stable runs
I	558	9613	9612	9608	9609	9600	9525	533	2597	2595	6	9668
II	77	10,198	10,194	10,194	10,194	10,191	10,183	75	2843	2094	7	10,231
III	140	1486	709	2350	2284	2280	2105	88	4	102	1	3186
IV	519	9328	27	8791	8165	8177	7262	0	0	280	0	9250

Table 4
Number of behavioural simulations of inundation extent

	Bridge I	Bridge II	Bridge III	Bridge IV
ERS-2 SAR	3628	7150	7291	6615
ENVISAT ASAR	19	33	331	514
Photographs	626	923	2474	8761
All images	15	27	41	42
All likelihood measures	0	0	0	0

Again bridge formulation IV results in a large number of runs accepted, and furthermore ENVISAT ASAR results in a very low number of behavioural runs.

It is perhaps not surprising that out of ~80,000 simulations with different parameter set *no* set survives this rigorous analysis of acceptability across all performance measures (see Table 4), even though all model set up and likelihood measures have been based on values which could be considered as generous by practitioners. Unfortunately this means that the current model set-up has limited predictive power and has to be used with caution in prediction for future events. This is naturally not a preferred outcome from a complicated modelling exercise. Nevertheless, it has been shown that large amounts of information about the model can be retrieved by an uncertainty and sensitivity analysis—although a precautionary approach has to be taken in how this information is applied.

7. Discussion

The GLUE methodology applied here requires the specification of a range of acceptable values for each observation which distinguishes between the behavioural and non-behavioural sets of models. Models that result in predictions outside that range are rejected [66]. Ideally, a physically-based model should perform behaviourally for all evaluation measures otherwise the adequacy of the physically-basis of the model can be questioned. However, this requires that the effective observational errors allowed are also physically consistent. It is often the case that when images of inundation are projected back onto the available geometry, then the interpolated inundation heights are not physically consistent, even allowing for the difficulties in finding the inundation boundary. In this case it may be very difficult to find a model that provides simulation results that are globally consistent with the observations. There are then four possible responses: investigate those regions of the flow domain where there are consistent anomalies between model predictions and range of observations; avoid using data we do not believe or that is doubtful; introduce local parameters if there are particular local anomalies (as at the bridges); make error bounds wider in some way where data is doubtful; or, if none of the

above can be justified (no reason to doubt anomalous data), then resort to local evaluations in assessing local uncertainties (for the latter see [92]).

Related to this is the interesting discussion of whether a model has to be acceptable for all evaluation measures or if this may be only restricted to one of the evaluation measure of particular interest (e.g. inundation level at a hospital). The model should ideally be behavioural for all evaluation criteria, due to the physical basis of this modelling approach. However, this may not be a very practical approach when the computation of risk levels is needed and the engineer may have to give his best educated guess. This requires a choice of which set of performance criteria have to be satisfied for different types of predictions and which set of performance measures are used to actually compute the uncertainty bound. Both sets do not have to be identical and in many cases it will be difficult to make this decision due to the non-linear interactions of the model. For example in this particular case, even though no simulations are behavioural for all evaluation measures, a flood hazard map can be derived for the time of the Envisat ASAR. As it is close to the maximum inflow discharge, the minimum requirements are that the parameter set has to be behavioural for each maximum water level (Table 3) as well as for the Envisat ASAR image (Table 4). However, the cumulative distribution function for each cross-section will be computed by using the performance measure at *each* cross-section rather than for *all* cross-sections. Fig. 13 illustrates this risk map with the 5% and 95% percentile. The previous discussion illustrates that such a map has to be interpreted with care as it is based on only a fraction of the available performance measures. In any publication or presentation this should be made clear.

Similar explicit reporting is needed for any sensitivity analysis. It can be shown that the relative importance of any factor depends largely on the hydraulic conditions in the reach. Therefore, decisions to omit factors from further analysis due to low sensitivities, or the attempt to decrease the overall model uncertainty through knowledge gain of the most sensitive factors, has to be based on a large variety of hydraulic scenarios.

We demonstrate that the GLUE approach allows for the possibility of model failure. It has to be especially highlighted that the most complex bridge implementation did produce a much higher number of behavioural model runs of the timing of the peak discharge between upstream and downstream boundary. This seems to be inconsistent with all the stage evaluation data. It suggests additional neglected processes within the model reach which have not been considered in setting up the model. The equifinality of the different bridge implementations opens the question about the choice of the correct one. According to Young et al. [93] the most

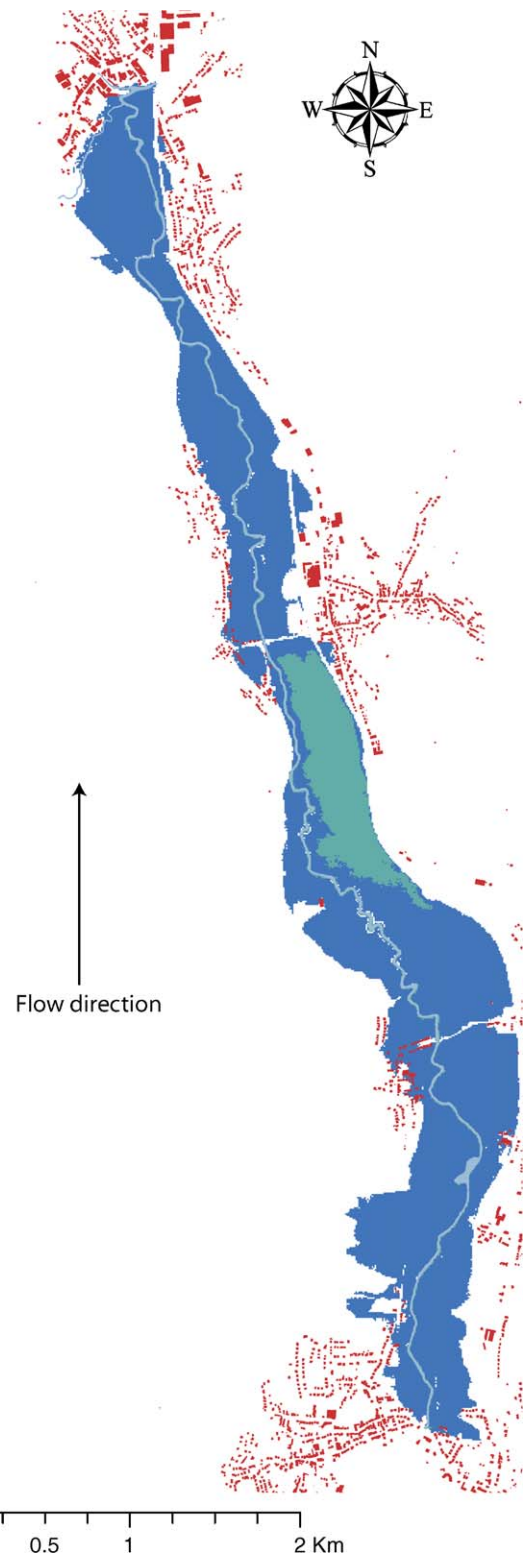


Fig. 13. Flood hazard prediction for the inundation extent at the time of the Envisat ASAR image. The 5% and 95% of flood hazard is plotted along the Alzette river. This flood hazard map could be used for flood warning or flood mapping purposes.

simplest implementation which can be justified by the data should be chosen, whereas Abbott and other [94] would argue that the most physically correct implemen-

tation should be chosen. We leave it to the reader to subscribe to one of these views.

Reporting inadequacy of any modelling exercise is not common practise. In fact, it is normal to ignore this possibility in calibrating a model by optimisation and in statistical estimations of modelling uncertainty (see Discussion in [66]). However, this also raises the question of how the modelling strategy can be improved in the light of the model failure in this case. Our current model is obviously inadequate to represent the current evaluation data within our assumptions. The most obvious solution is to increase the number of parameters (e.g. assign more roughness values) and add processes, such as the height dependency of the Manning roughness, or increase dimensionality of the model structure. With unlimited computer power and a large number of Ph.D. students it may be possible to explore all these configurations. There are three possible outcomes out of such a labour intensive modelling exercise.

Firstly, we may still end up with no model configuration which is constrained within the error of the observations. This may mean on the one hand, that we still have not found a proper representation of the model region and on the other hand that there are fundamental flaws in our evaluation data and assumptions. Both results would be the seed for an additional extensive measurement program to identify missing processes or faulty data. The model failures guide the type of measurements which have to be conducted and thus hint at knowledge gaps in the understanding of hydrology and hydraulics.

Secondly, we may end up with a large number of different model formulations, which fit all our criteria (equifinality of model structures). If we believe that such a physical model is an approximation of reality, then we could assume that there are several physical processes which can represent the data and thus we have to go into the field and determine those data which would best discriminate between the alternative models. However, we should expect that prediction uncertainties without additional constraints might remain high, because due to the nonlinearities of the system any extrapolation would result in a large variety of predictions from the set of behavioural models.

Thirdly, we may have only one or a small number of similar model structures and set-ups which represent the data. In this case we may have partially succeeded and can establish a strong hypothesis that this model region has a certain hydraulic configuration, at least under the conditions of the evaluation event(s) (including any interactions between input data error realisation and effective parameter values). Such a constrained set should be used with care in extrapolating to other, perhaps more extreme, events taking proper account of the input uncertainties.

8. Conclusions

The effects of uncertainty of the input boundary conditions on inundation predictions are very often neglected. This is despite the fact that the inflow in most inundation models is very often derived from rating curves which have been established with limited measurements. These curves are liable to considerable uncertainties because of difficulties in determining the cross-sectional area, velocity, model structure and regression method. In this paper, we introduce a methodology to acknowledge this uncertainty which is based on the Generalized Likelihood Uncertainty estimation framework [52] and the Multicomponent Mapping [72]. The inundation model used in this study requires also the specification of Manning surface roughness coefficients. Three roughness coefficients (one channel, one left floodplain and one right floodplain) have been specified for 74 cross-sections of the Alzette river reach. The uncertainty in the downstream boundary has also been included in this analysis. Three different remote sensing images are available for model calibration. It has been assumed that any model which predicts the flood inundation better than a linear interpolation is behavioural and produces equally acceptable results. The region modelled also contains two bridges, which have been modelled in four different ways of increasing complexity.

An analysis of rating curve uncertainties leads to an uncertainty of the input of 18–25% at peak discharge. This input uncertainty has a significant impact on the model performance as well as on acceptable effective Manning channel roughness values, as does uncertainty in the downstream boundary. However the extent of the impact in upstream or downstream direction depends on the bridge implementation chosen. The input uncertainty has naturally a larger impact on the model as long as the type of bridge formulation does not dominate the inundation extent and the flow regime. The relative importance of any source of uncertainty depends largely on the hydraulic conditions in the reach. Therefore, decisions to omit factors from further analysis due to low sensitivities or the attempt to decrease the overall model uncertainty through knowledge gain of the most sensitive factors should be based on a large variety of hydraulic scenarios.

Different implementations of bridge structures can lead to significantly different predictions of inundation extent. Bridges have large impacts on flood inundation at high flow and the details of the geometry are generally not adequately specified in hydraulic models. The most complex bridge formulation failed to produce any acceptable model simulations when evaluated with respect to the stage hydrograph situated on the bridge with the steeper slope. It could not be verified how far this is related to the difference in the real location of

the stage hydrograph at the bridge and its counterpart in the model with significant distances between cross-sections. However, all other bridge formulations do lead to multiple acceptable simulations (indicating some equifinality of model structures).

It has been shown how the results of the simulations can be used to suggest how the effective values of the Manning surface roughness could be spatially disaggregated based on local model performances. However, there was no obvious correlation with local floodplain characteristics and therefore this approach has been only partially successful due to the difficulty in unambiguously identifying the source of any discrepancy or similarity.

Unfortunately, none of the sampled model structure/parameter set combinations was acceptable for all the evaluation criteria and thus the predictive power of the model is limited. In other words, this modelling approach failed, though this conclusion is necessarily conditioned on the limited number of model runs that were possible for each configuration. We argue that ideally, a physically based model should be behavioural for all evaluation criteria justified by the goal of the model and it is possible, that given an unlimited number of runs a number of such models might have been found. However, when the model runs available are not acceptable on all criteria it can be questioned whether uncertainty in predictions (discharges, risk maps) based on only certain evaluation measures have value. This is a fundamental problem and could occur easily with more complex flood inundation model structures (e.g. two-dimensional representations of the floodplain). We propose four different responses to this problem: closer investigation of anomalies with a possible rejection of some evaluation measures; introduction of local parameters at locations where sensitivity analysis suggests different values are justified (though at the cost of an increased dimensionality of the model space); increasing the size of the acceptable error bounds based on reasonable justification; and resort to the use of local evaluation measures in making local predictions. Moreover, we make the case that it can be advantageous to decouple the classification into behavioural and non-behavioural data sets from the calculation of uncertainty bounds. In practice, a modeller may be forced to produce an educated guess of flood hazard maps, and thus it might be necessary to compute uncertainty maps based on a fraction of the available performance measures. In such an approach a decision has to be made as to which performance measures have to be behavioural and which performance measures are used to calculate the risk maps.

The results of this study lead to the overall conclusion that it is vital to consider the uncertainty in rating curves, channel roughness and downstream boundary in flood forecasting and flood mapping. The overall

uncertainty would be reduced significantly if the uncertainty in these parameters and variables could be reduced.

Acknowledgements

The Monte Carlo simulations have been computed on workstations of the Lancaster ES department via the Condor system (<http://www.cs.wisc.edu/condor/>). We would like to thank everyone who contributed *empty computer cycles* and in particular Charles Blakeley, for his support. We would also like to thank Georges Müller of the Service de la Gestion de l'Eau, for providing some of the data used in this study. Florian Pappenberger has been funded by the Flood Risk Management Research Consortium (<http://www.flood-risk.org.uk/FRMRCRecentArticles.htm>). Development of extensions to the GLUE methodology has been supported by NERC Grant NER/L/S/2001/00658. Many thanks to Hannah Cloke and Renata Romanowicz, for their valuable comments. We further thank two anonymous reviewers for their valuable comments which improved the article significantly.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.advwatres.2005.11.012](https://doi.org/10.1016/j.advwatres.2005.11.012).

References

- [1] Pappenberger F, Beven KJ, Hunter N, Gouweleeuw B, Bates P, de Roo A, et al. Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European flood forecasting system (EFFS). *Hydrol Earth System Sci* 2005;9(4): 381–93.
- [2] 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.
- [3] Ahmad S, Simonovic SP. Integration of heuristic knowledge with analytical tools for the selection of flood damage reduction measures. *Canad J Civil Eng* 2001;28(2):208–21.
- [4] Apel H, Thieken AH, Merz B, Bloschl G. Flood risk assessment and associated uncertainty. *Nat Hazards Earth System Sci* 2004;4(2):295–308.
- [5] Melching CS. Reliability estimation. In: Singh VP, editor. *Computer models of watershed hydrology*. Littleton: Water Resources Publications; 1995. p. 69–118.
- [6] Kuczera G. Comprehensive at-site flood frequency analysis using Monte Carlo Bayesian inference. *Water Resour Res* 1999;35(5): 1551–7.
- [7] Cook JL. Quantifying peak discharges for historical floods. *J Hydrol* 1987;96(1-4):29–40.
- [8] Barrows HK. Surface-water supply of New England, 1906. *Water-Supply Paper*, US Geological Survey, 1907;201: 13–23.

- [9] Corbett DM. Stream-gage procedure: a manual describing methods and practices of the US Geological Survey. Water-Supply Paper, US Geological Survey, 1943;888: 245.
- [10] Herschey RW. Streamflow measurement. London: E&FN Spon; 1995.
- [11] Kennedy EJ. Discharge ratings at gaging stations. In: Techniques of water-resources investigations Book 3. USG Survey, editor; 1984. p. 58 [chapter A10].
- [12] World Meteorological Organization. Manual on stream gaging, volume II: computation of discharge. Operational hydrology report Np13; 1980, WMO.
- [13] Schmidt AR. Analysis of stage-discharge relations for open-channel flow and their associated uncertainties. Urbana: University of Illinois; 2002. p. 328.
- [14] Chen YC, Chiu CL. A fast method of flood discharge estimation. *Hydrol Process* 2004;18(9):1671–84.
- [15] Pelletier PM. Uncertainties in the single determination of river discharge: a literature review. *Canad J Civil Eng News* 1988;15(5):834–50.
- [16] Wahl KL. Accuracy of channel measurements and implications in estimating streamflow characteristics. *J Res, US Geol Survey* 1977;5(6):811–4.
- [17] Sefe FTK. A study of the stage–discharge relationship of the Okavango River at Mohembo, Botswana. *Hydrolog Sci J–J Des Sci Hydrolog* 1996;41(1):97–116.
- [18] Callede J, Kosuth P, Guyot JL, Guimaraes VS. Discharge determination by acoustic Doppler current profilers (ADCP): a moving bottom error correction method and its application on the River Amazon at Obidos. *Hydrolog Sci J–J Des Sci Hydrolog* 2000;45(6):911–24.
- [19] Chow VT, Maidment DR, Mays CW. Applied hydrology. New York: McGraw-Hill; 1988.
- [20] International Organisation for Standardization. Collection of data for determination of errors in measurement by velocity area methods, 1088. International Organisation for Standardization, editor. Geneva; 1973.
- [21] International Organisation for Standardization. Calculation of the uncertainty of a measurement of flow-rate, 5168. International Organisation for Standardization, editor. Geneva; 1978.
- [22] International Organisation for Standardization. Guide to the expression of uncertainties in measurements. International Organisation for Standardization, editor. Geneva; 1995.
- [23] International Organisation for Standardization. Slope area method, 1070. International Organisation for Standardization, editor. Geneva; 1992.
- [24] Gawne KD, Simonovic SP. A computer-based system for modeling the stage–discharge relationships in steady-state conditions. *Hydrolog Sci J–J Des Sci Hydrolog* 1994;39(5):487–506.
- [25] Knight DW, Shiono K. Channel and floodplain hydraulics. In: Anderson MG, Walling DE, Bates PD, editors. Floodplain processes. New York: John Wiley & Sons; 1996.
- [26] Bonacci O. Influence of turbulence on the accuracy of discharge measurements in natural streamflows. *J Hydrology* 1979;42(3–4): 347–67.
- [27] Chiu CL, Chen YC. An efficient method of discharge estimation based on probability concept. *J Hydraul Res* 2003;41(6):589–96.
- [28] Herschey RW. Current meter calibration: individual rating versus group rating, No. 134. IAHS Pub; 1982. p. 25–36.
- [29] Carter RW, Anderson LE. Accuracy of current–meter measurements. *Am Soc Civil Eng J* 1963;89(No. HY4):105–15.
- [30] Herschey RW. The uncertainty in a current meter measurement. *Flow Measur Instru* 2002;13:281–4.
- [31] Stevens JC. A method estimating stream discharge from a limited number of gaugings. *Eng News* 1907;58(3):52–3.
- [32] International Organisation for Standardization. Liquid flow measurement in open channel—part 2: determination of the stage-discharge relationship ISO Standard 1100/2-1982. In: Measurement of liquid flow in open channels—ISO Standards Handbook 16, I.O.f. Standardization, editor. Geneva; 1983. p. 154–186.
- [33] Herschey RW. Flow measurement. In: Herschey RW, editor. Hydrometry: principles and practices. Chichester: Wiley; 1998. p. 9–83.
- [34] Clarke RT. Uncertainty in the estimation of mean annual flood due to rating-curve indefiniteness. *J Hydrol* 1999;222(1–4):185–90.
- [35] Mishra SK, Seth SM. Use of hysteresis for defining the nature of flood wave propagation in natural channels. *Hydrol Sci J–J Des Sci Hydrolog* 1996;41(2):153–70.
- [36] Mishra SK, Singh VP. Hysteresis-based flood-wave analysis using the concept of strain. *Hydrol Process* 2001;15(9):1635–51.
- [37] Syvitski JP, Morehead MD. Estimating river-sediment discharge to the ocean: application to the Eel margin, northern California. *Mar Geology* 1999;154(1–4):13–28.
- [38] Knight DW, Brown FA. Resistance studies of overbank flow in rivers with sediment using the flood channel facility. *J Hydraul Res* 2001;39(3):283–301.
- [39] Herschey RW. Flow measurement. In: Herschey RW, editor. Hydrometry. New York: Wiley; 1999. p. 181–97.
- [40] Costa JE, Spicer KR, Cheng RT, Haeni FP, Melcher NB, Thurman EM. Measuring stream discharge by non-contact methods: a proof-of-concept experiment. *Geophys Res Lett* 2000;27(4):553–6.
- [41] Spicer KR, Costa JE, Placzek G. Measuring flood discharge in unstable stream channels using ground-penetrating radar. *Geology* 1997;25(5):423–6.
- [42] Aronica G, Hankin B, Beven KJ. Uncertainty and equifinality in calibrating distributed roughness coefficients in a flood propagation model with limited data. *Adv Water Res* 1998;22(4):349–65.
- [43] Beven KJ, Carling P. Velocities, roughness and dispersion in the lowland River Severn. In: Carling P, Petts G, editors. Lowland floodplain rivers. New York: John Wiley & Sons; 1992. p. 71–93.
- [44] Romanowicz R, Beven KJ. Estimation of flood inundation probabilities as conditioned on event inundation maps. *Water Resour Res* 2003;39(3). art. no.-1073.
- [45] Matgen P, Henry JBF, Pappenberger F, Pfister L, de Fraipont P, Hoffmann L. Uncertainty in calibrating flood propagation models with flood boundaries derived from synthetic aperture radar imagery. In: XXth ISPRS Congress, Istanbul, Turkey: International Society for Photogrammetry and Remote Sensing; 2004.
- [46] Pappenberger F, Beven KJ, Horritt M, Blazkova S. Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations. *J Hydrol* 2005;302(1–4):46–69.
- [47] US Army Corps Engineers, HEC-RAS. Available from <http://www.hec.usace.army.mil>.
- [48] Hunter NM, Bates PD, Horritt MS, De Roo APJ, Werner MGF. Utility of different data types for calibrating flood inundation models within a GLUE framework. *Hydrol Earth System Sci* 2005;9:412–30.
- [49] Bates PD, Anderson MG, Hervouet JM. Initial comparison of 2-dimensional finite-element codes for river flood simulation. Proceedings of the institution of civil engineers-water maritime and energy 1995;112(3):238–48.
- [50] Horritt MS, Bates PD. Evaluation of 1D and 2D numerical models for predicting river flood inundation. *J Hydrol* 2003;268(1–4):87–99.
- [51] Lane SN. Hydraulic modelling in hydrology and geomorphology: a review of high resolution approaches. *Hydrol Process* 1998;12(8):1131–50.
- [52] Beven KJ, Binley A. The future of distributed models: model calibration and uncertainty prediction. *Hydrol Process* 1992;6:279–98.
- [53] Clarke RT, Mendiondo EM, Brusa LC. Uncertainties in mean discharges from two large South American rivers due to rating

- curve variability. *Hydrol Sci J–J Des Sci Hydrolog* 2000;45(2):221–36.
- [54] Wormleaton PR, Merrett DJ. An improved method of calculation for steady uniform-flow in prismatic main channel flood-plain sections. *J Hydraul Res* 1990;28(2):157–74.
- [55] Rantz SE, et al. Measurement and computation of streamflow, volume 1: measurement of stage and discharge, volume 2: computation of discharge—US Geological Survey, Water Supply Paper 2175; 1982. Available from <http://water.usgs.gov/pubs/wsp/wsp2175/>.
- [56] Manning R. On the flow of water in open channel and pipes. *Trans Institut Civil Eng Ireland* 1891;20:161–207.
- [57] Clemmens AJ, Wahl TL, Bos MG, Replogle JA. Water measurement with flumes and weirs. ILRI Publication 58, International Institute for Land Reclamation and Improvement: Wageningen, The Netherlands; 2001.
- [58] Gwinn WR, Parsons DA. Discharge equations for HS, H, and HL flumes. *J Hydraul Division–ASCE* 1976;102:1704–6.
- [59] Deka P, Chandramouli V. A fuzzy neural network model for deriving the river stage–discharge relationship. *Hydrol Sci J–J Des Sci Hydrolog* 2003;48(2):197–209.
- [60] Supharatid S. Application of a neural network model in establishing a stage–discharge relationship for a tidal river. *Hydrol Process* 2003;17(15):3085–99.
- [61] Chandramouli V, Raman H. Multireservoir modeling with dynamic programming and neural networks. *J Water Resour Planning Manag–ASCE* 2001;127(2):89–98.
- [62] Tawfik M, Ibrahim A, Fahmy H. Hysteresis sensitive neural network for modeling rating curves. *J Comput Civil Eng* 1997;11(3):206–11.
- [63] Dibike YB, Solomatine DP. River flow forecasting using artificial neural networks. *Phys Chem Earth (B)* 2001;26(1):1–7.
- [64] Chen YC, Chiu CL. An efficient method of discharge measurement in tidal streams. *J Hydrol* 2002;265(1–4):212–24.
- [65] Dawdy DR, Lucas W, Wang WC. Physical basis of stage–discharge ratings. In: Eighth international symposium on stochastic hydraulics. Beijing, China: Balkema, A.A.; 2000.
- [66] Beven KJ. A manifesto for the equifinality thesis. *J Hydrol*, in press.
- [67] Beven KJ. Towards an alternative blueprint for a physically based digitally simulated hydrologic response modelling system. *Hydrol Process* 2002;16(2):189–206.
- [68] DeGagne MPJ, Douglas GG, Hudson HR, Simonovic SP. A decision support system for the analysis and use of stage–discharge rating curves. *J Hydrol* 1996;184(3–4):225–41.
- [69] Potter KW, Walker JF. A model of discontinuous measurement error and its effect on the probability distribution of flood discharge measurements. *Water Resour Res* 1981;17(5):1505–9.
- [70] Potter KW, Walker JF, editors. Modeling the error in flood discharge measurements. In: El-Shaarawi AH, Esterby RS, editors. *Developments in water science*, vol. 17. Elsevier: New York; 1982.
- [71] Leonard J, Mietton M, Najib H, Gourbesville P. Rating curve modelling with Manning’s equation to manage instability and improve extrapolation. *Hydrol Sci J–J Des Sci Hydrolog* 2000;45(5):739–50.
- [72] Pappenberger F, Beven K. Functional classification and evaluation of hydrographs based on multicomponent mapping. *Int J River Basin Manag* 2004;2(2).
- [73] Barkau RL. UNET one-dimensional unsteady flow through a full network of open channels, user’s manual. Davis: US Army Corps of Engineers, Hydrologic Engineering Center; 1997.
- [74] Fread DL, Harbaugh TE. Open channel profiles by Newton’s iteration technique. *J Hydrol* 1971;13(1):70–80.
- [75] Wilson MD. Evaluating the effect of data and data uncertainty on predictions of flood inundation. Faculty Eng, Sci Math–School Geography. Southampton: University of Southampton; 2004. p. 276.
- [76] Bashford KE, Beven KJ, Young PC. Observational data and scale-dependent parameterizations: explorations using a virtual hydrological reality. *Hydrol Process* 2002;16(2):293–312.
- [77] Beven KJ. Linking parameters across scales—subgrid parameterizations and scale-dependent hydrological models. *Hydrol Process* 1995;9(5–6):507–25.
- [78] Lane SN, Richards KS. High resolution, two-dimensional spatial modelling of flow processes in a multi-thread channel. *Hydrol Process* 1998;12(8):1279–98.
- [79] Beven KJ. *Rainfall-runoff modelling: the primer*. New York: Wiley; 2001.
- [80] Beven KJ. How far can we go in distributed hydrological modelling? *Hydrol Earth System Sci* 2001;5(1):1–12.
- [81] Thiemann M, Trosset M, Gupta H, Sorooshian S. Bayesian recursive parameter estimation for hydrologic models. *Water Resour Res* 2001;37(10):2521–35.
- [82] Freer J, Beven KJ, Peters N. Multivariate seasonal period model rejection within the generalised likelihood uncertainty estimation procedure. In: Duan H, Gupta S, Sorooshian A, Turcotte R, editors. *Calibration of watershed models*. Washington: American Geophysical Union; 2003. p. 69–88.
- [83] Romanowicz R, Beven KJ, Tawn JA. Evaluation of predictive uncertainty in nonlinear hydraulic models using a Bayesian Approach. In: Barnett V, Turkman KF, editors. *Statistics for the environment 2, water related issues*. New York: Wiley & Sons; 1994. p. 297–317.
- [84] Freer J, Beven KJ, Ambroise B. Bayesian estimation of uncertainty in runoff prediction and the value of data: an application of the GLUE approach. *Water Resour Res* 1996;32(7):2161–73.
- [85] Franks SW, Gineste P, Beven KJ, Merot P. On constraining the predictions of a distributed model: the incorporation of fuzzy estimates of saturated areas into the calibration process. *Water Resour Res* 1998;34(4):787–97.
- [86] Blazkova S, Beven KJ. Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty). *Water Resour Res* 2002;38(8). Art. no. 1139.
- [87] Horritt M. A statistical active contour model for SAR image segmentation. *Image Vision Comput* 1999;17:213–24.
- [88] Horritt MS. Calibration of a two-dimensional finite element flood flow model using satellite radar imagery. *Water Resour Res* 2000;36(11):3279–91.
- [89] Pappenberger F, Iorgulescu I, Beven KJ. Sensitivity analysis based on regional splits and regression trees (SARS–RT). *Environ Modell Softw*, in press.
- [90] Saltelli A, Tarantola A, Campolongo F, Ratto M. *Sensitivity analysis in practice—a guide to assessing scientific models*. Chichester: John Wiley & Sons; 2004.
- [91] Werner MGF, Hunter NM, Bates PD. Identifiability of distributed floodplain roughness values in flood extent estimation. *J Hydrol* 2005;314(1–4):139–57.
- [92] Choi HT, Beven KJ. Multi-period and multi-criteria model conditioning to reduce prediction uncertainty in distributed rainfall-runoff modelling within the GLUE framework. *J Hydrol*, submitted for publication.
- [93] Young P, Parkinson S, Lees M. Simplicity out of complexity in environmental modelling: Occam’s razor revisited. *J Appl Statist* 1996;23:165–210.
- [94] Abbott MB, Babovic VM, Cunge JA. Towards the hydraulics of the hydroinformatics era. *J Hydraul Res* 2001;39(4):339–49.