
Article:
Willis, T, Wright, N and Sleigh, A (2019) Systematic analysis of uncertainty in 2D flood inundation models. Environmental Modelling & Software, 122. 104520. ISSN: 1364-8152

https://doi.org/10.1016/j.envsoft.2019.104520

Systematic Analysis of Uncertainty in 2D Flood Inundation Models

Thomas Willis[1], Nigel Wright[2], Andrew Sleigh[1]

1. University of Leeds
2. Nottingham Trent University

Abstract: Assessing uncertainty is a critical part of understanding and developing flood inundation models for use in risk assessment. Typically, uncertainties are investigated by comparing the effects of an ensemble of key model inputs, such as friction values and hydrographic uncertainties, on model outputs. In this study, an approach is adopted that also consider the uncertainty associated with the computational models. Using the LISFLOOD-FP code, which contains multiple methods for solving floodplain flow, two test cases with different hydraulic characteristics are used in a systematic uncertainty analysis. An ensemble of inputs including cell size, hydrological uncertainty, and representation of buildings are assessed for impact on modelling results. Results show the numerical complexity is a significant source of uncertainty in complex flow regimes, but this reduces in typical fluvial flood events. The method of assessing the modelling output is also found to be important in determining the overall influence of parameters.

Highlights:

- This paper analyses the major sources of uncertainty in flood inundation model and evaluates which source of uncertainty is the most significant.
- Two test cases, with different hydraulic properties, are used to provide broader conclusions to the uncertainty analysis
- The complexity of the physical model is shown to be highly influential on model results in both test cases.
- The significance of an input on model outputs is dependent on the assessment method used.

Keywords
Flood Inundation modelling, Uncertainty Analysis, Numerical Modelling, Hydraulics

1. INTRODUCTION

Flood inundation models are an essential tool in understanding the hydrodynamics of flood events and in assessing and analysing risk. Computed flood depths and extents are an essential requirement to determine areas at risk of damage from flooding, the effectiveness of flood mitigation measures and in the assessment of economic losses of flooding. In order to develop these data sets, numerical models have been created which use the shallow water equations to describe the movement of water in a river channel and across the floodplain. These numerical models form the basis of a computational model which will be used to simulate flooding. Additional datasets are required as inputs and to parametrise the numerical model to create the complete flood inundation model. Collectively, these components of the flood inundation model are a source of uncertainty that will impact on the accuracy of the results, and ultimately on the decisions that can be made from these datasets (Beven, 2008). To understand the significance of uncertainty in the models and input datasets, the contribution of each source of uncertainty must be assessed to determine which source of uncertainty provides the greatest impact

on the model results (Willems, 2012). This is underlined with the legal requirement of decision makers to understand computational models, particularly Catastrophe models, through legislation such as the UK Solvency 2 scheme for insurers (Boss, 2011). Broadly, this requires end users of computational models to understand the assumptions and uncertainties of the model used to create them. This in turn places an emphasis on modellers to justify often complex choices in model setup and to understand the uncertainty associated with these choices. Analysis of model uncertainty is therefore a critical requirement of current modelling approaches.

Model uncertainty is typically divided into four categories (Willems, 2012): input data uncertainty (the uncertainty related to input data sets such as hydrology and terrain data), parametric uncertainty (e.g. friction coefficient values), model structure uncertainty (the uncertainty associated with the numerical model), and model assessment uncertainty (such as the data and approaches used to validate the model). A number of studies have investigated the effects of these sources of uncertainty on model results (Warmink et al. (2011), Pappenberger et al. (2005), Hunter et al. (2005a), Willems (2012), Van Steenbergen et al. (2012)). A further number of studies have provided detailed analysis of the impact of individual components of modelling uncertainty, such as terrain data (Morgan et al 2016 , Tsubaki et al 2013, Stephens et al 2012), friction parameters (Pappenberger et al. (2008), Dung et al. (2011), Apel et al. (2009), Hall et al. (2005)), input boundary conditions (Domeneghetti et al. (2012), Aronica et al. (2012)) and grid cell size (Wildemeersch et al., 2014).

By comparison uncertainty associated with the model structure is less frequently studied as it is difficult to analyse. Methods to describe the distribution of uncertainty associated with the model structure are difficult due to the number of factors that influence the uncertainty (Gupta et al., 2012). One of the key factors of the model structure uncertainty is the level of physical representation chosen for the conceptual model, which forms the basis of the numerical model. In hydrodynamic modelling, the complexity of the conceptual model and the level of physical representation, is typically associated with the number of terms used from the momentum component of the governing equations. The governing equations are the Shallow Water Equations (SWE), which are comprised of the conservation of mass equation (which describes the preservation of mass as water moves through an area) and the momentum equation (Equation 1). These equations describes the principal forces that conserve energy and control the movement and flow of water, and the number of terms used from it leads to different numerical models of the flood wave (Chow (1988), Bates et al. (2010)).

A number of studies that have explored the impact of the level of physical representation focus on benchmarking approaches, rather than structured uncertainty approaches (Hunter et al (2008), Neelz and Pender (2010), Neelz and Pender (2013)). These studies identified that models can be divided into two categories dependent on the underlying numerical model: simplified and full physics approaches, where the simplified approach removes terms from the momentum equation leaving 2-term or 3-term numerical models (Neelz and Pender, 2013). They further identified that model behaviour and interaction with other parameters vary greatly between these categories, and that in appropriate situations simplified approaches serve as well as full physics approaches. Further studies have used this benchmarking approach to cross compare further model structure uncertainty and parameter uncertainty (Liu et al., 2019). Typically, it has been difficult to include this aspect of the computational

model uncertainty in uncertainty analyses, as this has required the use of multiple model codes to represent the different levels of model complexity. These codes will not share common code structure, such as grid discretisation and time stepping approaches, meaning direct comparison is often difficult (Hunter et al 2008) and (Neelz and Pender, 2010). To overcome this issue, and to allow for a comprehensive approach to analysing modelling uncertainty, a computational model framework is required with multiple numerical models with varying levels of complexity. This approach is demonstrated in Neal et al (2012) using the LISFLOOD-FP code, which has multiple options for solving the 2D floodplain flow within a single model framework for a series of benchmarking problems. This further highlights the difference between the simplified and full physics approaches. Further sources of uncertainty are not considered in this study.

This study aims to continue the course of previous studies by investigating model structure uncertainty alongside other sources of uncertainty in flood inundation modelling. A systematic approach to analysing the uncertainty is used, where key model inputs including friction coefficients, model inflow, and cell size, are assessed as discrete inputs. Model structure uncertainty is included as a key model input by using the multiple floodplain flow solvers in the LISFLOOD-FP code. This approach provides a degree of control on the results, allowing the impact of varying the model inputs to be understood within the model (such as comparing depths at specific points), rather than considering only global model outputs.

A typical problem in uncertainty analysis, is the effect of test case bias, where the physical properties of the model domain may favour elements of a model. To reduce this impact, two test cases are selected with different hydraulic conditions. The first is a surface water flood in Glasgow, that has been used in previous studies (Hunter et al 2008, Aronica et al. (2012), Fewtrell et al 2008)), which is dominated by transcritical flow in a complex urban environment. The second is a river overtopping event, in a predominantly rural domain, a case study based on an event from the 2007 UK summer floods in Mexborough, South Yorkshire.

The approach adopted focuses on pragmatic modelling decisions, using parameters and values that would typically be used in a modelling exercise. This method is preferred to variance-based sensitivity analysis which provide quantitative analysis, but at a high computational cost (Hall et al 2005 and Savage et al 2016). The aim of this approach is to provide results that are relevant to model developers, model users and decision-makers. This paper is divided into the following sections: a methodology section with an introduction to the LISFLOOD-FP code and a review of the different solutions of floodplain flow, a description of the methods used to assess uncertainty, the presentation of the results from the two test cases, followed by a discussion and conclusions.

2. METHODOLOGY

2.1 Hydraulic Model: LISFLOOD-FP

In order to explore model structure uncertainty, a model framework is required that allows the complexity of the model, and the level of physical representation, to be isolated from other factors in the numerical model. The LISFLOOD-FP code is a regular grid 2D code with different methods for solving floodplain

flow. The difference in the methods are based on varying levels of complexity in the numerical engine, and removing terms from the governing equations, to create simplified physics approaches.

Chow (1988) defined 3 physical wave structures types based on the 1D conservative momentum equation of the Shallow Water Equations (Equation 1), which can then be applied to 2D numerical models.

$$\frac{1}{A}\frac{\partial}{\partial x}\left(\frac{Q^2}{A}\right) + \frac{1}{A}\frac{\partial Q}{\partial t} + g\frac{\partial y}{\partial x} - g(S_o - S_f) = 0 \quad (1)$$

Convective term *Local term* *Pressure term* *Gravity and Friction Source Terms*

Diffusion Wave model (LISFLOOD-ATS)

Acceleration Wave model (LISFLOOD-ACC)

Full Physics Dynamic Wave model (LISFLOOD-Roe and Rusanov)

Where A is channel cross section area (m²), Q is discharge (m³/s), t is time, g is gravitational constant (m²/s), x and y are cell node directions, S_0 is the gravitational source term and S_f in friction source term. A fully physics based approach is made up of all terms in Equation 1. Wave models that do not include all these terms are described as simplified physics approaches. Two simplified physics approaches have been developed in LISFLOOD-FP, a diffusion wave approach which is represented by the removal of the convective and local acceleration terms, and the 'acceleration' wave which is represented by the removal of the convective acceleration term. The latter wave type formulation has been used more frequently in flood models in recent year, as a way of improving the physical representation of simplified physics approaches (Aronica et al 1998, Bates et al 2010, de Almedia et al 2012). In principle these simplified formulations should be only used in appropriate circumstances, such as when variations in the free surface and local velocities are negligible to the length scale of the problem being investigated (Hunter et al., 2007). This can be understood in the context of fluvial flooding where theoretically, the convective acceleration term is negligible compared to the effects on the overall movement of flood waters, and the diffusive wave become more dominant. Conversely, this means for scenarios with complex topography the diffusion wave formulation will no longer adequately represent the complexity of the hydraulics. However, simplified approaches have an advantage in that they can be solved by more direct numerical methods than those employed in solving the full momentum equation. This will in principle produce efficient computational code that is quicker to run than models that use the dynamic wave approach, at the expense of complete representation of the hydraulic processes (Hunter et al., 2008). As a result, there is a temptation to push the limits of the simplified approaches, beyond the theoretical limits of these methods.

There are 3 formulations used in the LISFLOOD-FP framework which utilise these physics models. LISFLOOD-ATS is an approximation to the diffusive wave formulation, using a decoupled Manning's

equation to update cell variables, in a finite difference approach (Bates and De Roo, 2000). The intercell unit discharge value, Q, is performed using Equation 2;

$$Q_x^{i,j} = \frac{h_{flow}^{5/3}}{n}\left(\frac{h^{i-1,j} - h^{i,j}}{\Delta x}\right)^{1/2} \Delta y \quad (2)$$

Where h is the water depth in a cell, n is Manning's friction coefficient, i and j are cell locations, x and y are grid cell size and h_{flow} is the difference between the highest water cell depth and elevation in two adjacent cells. The continuity equation is then used to sum the cell values to determine the new water depth values. The ATS formulation uses an adaptive timestep (Hunter et al., 2005b) and represents the lowest level of complexity for solving floodplain flow in LISFLOOD-FP.

LISFLOOD-ACC is based on the acceleration wave described in Equation 1, where the convective term is removed from the Equation 1 to create a numerical solution to this in the form of Equation 3;

$$Q_{t+\Delta t} = \frac{q_t - g h_t \Delta t \dfrac{\partial(h_z + z)}{\partial x}}{\left(1 + g h_t \Delta t n^2 q_t / h_t^{10/3}\right)} \quad (3)$$

Where z is the elevation in a cell. In order to recreate the acceleration wave, a semi-implicit formulation is utilised that approximates the upwinding methods used in finite volume approaches (de Almeida et al., 2012).

A full physical solution is available in the LISFLOOD-FP code, which uses a Godunov-based finite volume method for solving the SWE on a regular grid. A full description of the method can be found elsewhere (Villanueva and Wright, 2006), but the basic approach to the method involves updating cells using a form of Equation 4;

$$U_i^{t+1} = U_i^t - \frac{\Delta t}{\Delta x}\left[F_{i+\frac{1}{2}} - F_{i-\frac{1}{2}}\right] \quad (4)$$

Where t is the time step interval U is the conserved variables of depth and velocity to be updated at the next time step for cell i and F is the intercell flux to update the variables.

The Godunov method recreates a Riemann problem at the intercell boundary, for which there are a number of numerical methods available to solve this problem, and to calculate the intercell flux. Here, two are considered with different levels of numerical complexity. The first is a linear approximation to the Reimann problem - the Roe solver (Roe, 1981) which has been implemented in LISFLOOD-FP and other 2D regular grid flood inundation models (Neal et al. (2012) and (Villanueva and Wright, 2006)). The second is a simplified upwind method, the Rusanov solver, which approximates the Riemann problem using information determined in the calculation of the timestep. This solver has been used previously for solving 2D flood problems (Simoes, 2011) and according to Toro represents the simplest approach to an upwind boundary solver (Toro 2001).

By using these two full physics approaches, LISFLOOD-Roe and LISFLOOD-Rusanov, insight into the significance of the complexity of the numerical method can also be understood and can be cross-compared to the different levels of physical representation. Other approaches to solving the SWE including higher order solvers may provide more robust and accurate solutions to these equations but

are not considered in this study, due to the scale and computational costs of the test cases. These approaches could be used as a potential area of future study.

These different solutions to solving floodplain flow in LISFLOOD-FP will be collectively referred to as modules.

2.2. Uncertainty analysis methodology

The systematic approach is based on discretising each uncertain input into levels, and cross-comparing each level of all the input parameters, as in a Monte Carlo approach. For each input a maximum and minimum value for an input is chosen and is divided into discrete values (or levels). Unlike a Monte Carlo approach, the level of an input reflects values that might be typically considered by a modeller, rather than representing value based on the distribution of that input (as in Pappenberger et al. (2008)). This is similar to the approach of Hunter et al (2008). Although an approach of sampling the inputs from an underlying input uncertainty distribution is not used, the upper and lower limit of the input values chosen will implicitly assume the range of values in a uniform distribution.

For the two test cases used, the same principal inputs are used in the input ensemble; friction values, hydrograph inputs, cell size, random error in the Digital Elevation Model (DEM) and approaches to representing flow through a building. The values used are shown in Table 1.

Table 1. Parameter type, range and levels for the systematic assessment

Parameter Type	**Range**	**Levels**			**Notes**
Model Complexity	Full SWE to diffusion wave	4 levels based on the LISFLOOD-FP modules (ATS,ACC,Roe, Rusanov)			Further detail is provided in Section 2
Cell Size	2m – 4m for urban tests, 20m-40m for rural tests	2 levels			Based on work by Fewtrell et al (2011)
Hydrograph	20% of calculated hydrograph	5 levels (-20%,-10%, 0,+10%,+20%)			Value based on Di Baldassarre et al (2012)
DEM error	0cm - 15cm	2 levels - Original surface and 15cm degraded			LiDAR vertical RMSE
Building Representation	BR, BP, BB	3 levels (Building Block, Building Porosity, Building roughness)			Based on Schubert and Sanders (2012)
Friction Value	Low Friction - 0.008-0.020 (13 levels with a difference of 0.01 n)	Surface	Low	High	The value of the spatially distributed friction coefficient is split into 13 levels. Each level has a corresponding high
		1.	0.008	0.015	
		2.	0.009	0.02	
		3.	0.010	0.025	

	High Friction 0.015-0.075 (13 levels with a difference of 0.05)	4.	0.011	0.03	friction surface value and low friction surface value
		5.	0.012	0.035	
		6.	0.013	0.04	
		7.	0.014	0.045	
		8.	0.015	0.05	
		9.	0.016	0.055	
		10.	0.017	0.06	
		11.	0.018	0.065	
		12.	0.019	0.07	
		13.	0.02	0.075	

As described in Section 2, model complexity is represented by one of the four different floodplain solver formulations used in LISFLOOD-FP to solve floodplain flow. The lowest level is the two-term diffusion wave based ATS solver, followed by the three-term ACC solver. For the two full SWE models, the simplified interflux solver used in LISFLOOD-Rusanov is the third tier, whilst the more complex and robust Roe solver is the upper level of the model structure uncertainty.

Grid cell size is a critical input to the computational model, and the precise consideration of this is determined by two competing factors, minimum feature representation and computational runtime (Savage et al., 2016). Previous studies have identified that a rule of thumb for a minimum cell size is that main flow paths should be represented by three cells width (Fewtrell et al., 2011). For urban areas, the main flow path for surface flow will be the road network, which will have a minimum width of around 6m, so a 2-meter cell size is selected as the lowest level. A maximum value of 4 meters is used for the urban test case presented here, creating a 2-level input to represent the uncertainty in grid cell size. The maximum value is selected to test the assumption of Fewtrell et al. (2011). For the rural test case this is increased to 20-40m due to the computational time of the model runs but is justified as the road network is less critical in floodplain flow path development. At this scale, catchment scale geometry and topology will have a significant effect on flood event dynamics, which is suitably represented at the 20-40m resolution model.

Inflow boundary conditions can often represent the most significant uncertainty in flood inundation and risk modelling (Pappenberger et al., 2006). Evaluating the full range of uncertainties associated with hydrological inputs, including modelling approaches and is beyond the scope of this work. The description of uncertainties associated with gauge data and rating curve uncertainties provides a more focused approach to assessing inflow uncertainties. These have been analysed in previous studies and demonstrated to have a range of about ±20 – 25% of a calculated flow rate (Domeneghetti et al., 2012). As the two test cases presented here are based on historic events, this seems a suitable approach, as the events use observed inflows as boundary condition, which would be subject to these errors. The rating curve uncertainty was applied to the inflow boundary conditions with the upper and lower limits set at ±20%, creating a 5-level input for this model input. Whilst this range of levels does not take into

account precise details about either measurement procedure, it is felt to be sufficiently wide to cover the range of potential approaches to using gauge data to develop model inflow.

The DEM used in the models are all based on LiDAR derived data sets, which have a RMSE of the relative true ground rating of ±15cm for elevation and 5cm for lateral position (Cobby et al., 2001). This is incorporated into the input parameter space in a similar way to the approach of Hunter et al (2008) and Tsubaki and Kawahara (2013), by degrading the original DEM randomly and on a cell-by-cell basis, according to the RMSE value for height and lateral position.

An important part of developing flood models for urban areas is to consider how buildings should be represented in the model. Buildings walls will act as impermeable barriers to flood waters and are critical in defining flow paths during flood events, by directing and restricting flow routes. During a flood event, water that enters a building is often stored temporarily, which remove water from the floodplain. This can be critical in urban areas, where the number of buildings may significantly impact the volume of water stored during an event. Ideally, the method used to represent buildings should incorporate both of these macro scale processes. Three approaches are used to represent buildings, based on the approaches first suggested in Sanders et al. (2008) and Schubert et al. (2008): Building Block method, (BB), Building Resistance method (BR), and Building Porosity method (BP). The position of the building is determined using Ordnance Survey (OS) Landline vector map dataset. The building block method is defined as raising the elevation of a cell that fall within the footprint of the building by 6m. This will represent the impermeable structure of the building edge. The building resistance (BR) method contains no physical representation but a high friction value for all cells in the building footprint, to allow for some representation of the flow through the building. The Manning's coefficient is fixed at 0.1 in all test cases, which reduces water discharge levels through the building footprints. The BR method therefore represents the storage aspect of flood waters in buildings. The Building Porosity method uses a fixed conveyance porosity value of 0.5 for cells in building footprints for all test cases. This value represents the amount of cell space available to be occupied by the water, as well as the intercell discharge volume, and is similar to the approach adopted by Guinot (2012) and Soares-Frazao et al. (2008). The value of 0.5 is chosen as an average blockage effect. A range of values were tested but result variations for this value were negligible and are not considered further in this analysis. This reflects similar findings to Schubert and Sanders (2012).

The friction value is often the most considered parameter in calibrating models, due to the ease of implementing and its use to overcome deficiencies in the flood model. To test this, 13 sets of spatially distributed values based on OS Landline vector data are used. Low friction surfaces such as roads and pathways are given values of 0.008 to 0.02 at 0.001 intervals. Remaining land surfaces are defined as higher friction surfaces and are assigned values from 0.015 to 0.075 at 0.005 intervals. This creates a 13-level input as outlined in Table 1. The values are the same as used in a previous benchmarking case Hunter et al (2008). These values represent the theoretical range of possible Manning's *n* friction values as defined by the surface types in Chow (1988). The larger parameter space, compared to the other inputs is justified, as the values represent the range of values that could be justified in a modelling exercise.

2.3 Case Studies and Data

A typical problem for uncertainty analysis is the issue of test bias – where the specific test may unduly favour a particular input (Hall et al 2005, Hunter et al 2006, Apel et al 2009,). Two test cases with contrasting hydrodynamic characteristics are used to address this. The first test case is a culvert overtopping event that lead to a rapid (<1 hour), surface water flood event in the Greenfield region of Glasgow. This test case has been used in previous studies (Hunter et al 2008, Aronica et al 2012, Pender et al 2012), but remains a highly useful test for hydraulic models. This is due to the nature of the topography and the flow type (shallow and fast), which provide a stringent examination of the numeric model. The event inflow is introduced at the culvert location, to the east of the model domain (red circle, Figure 1), which leads to the formation of a main flow path in an east-west direction. A second flow path then forms leading to the south of the model domain, where water pools throughout the model run. Numerous smaller flow paths then emerge between the two main flow routes. The event is characterised by transcritical flows as a result of the short steep topography of the catchment, and the speed of the outflow. The peak discharge inflow for this event is estimated at 15m^3/s, with a total duration of 15 minutes from start to finish.

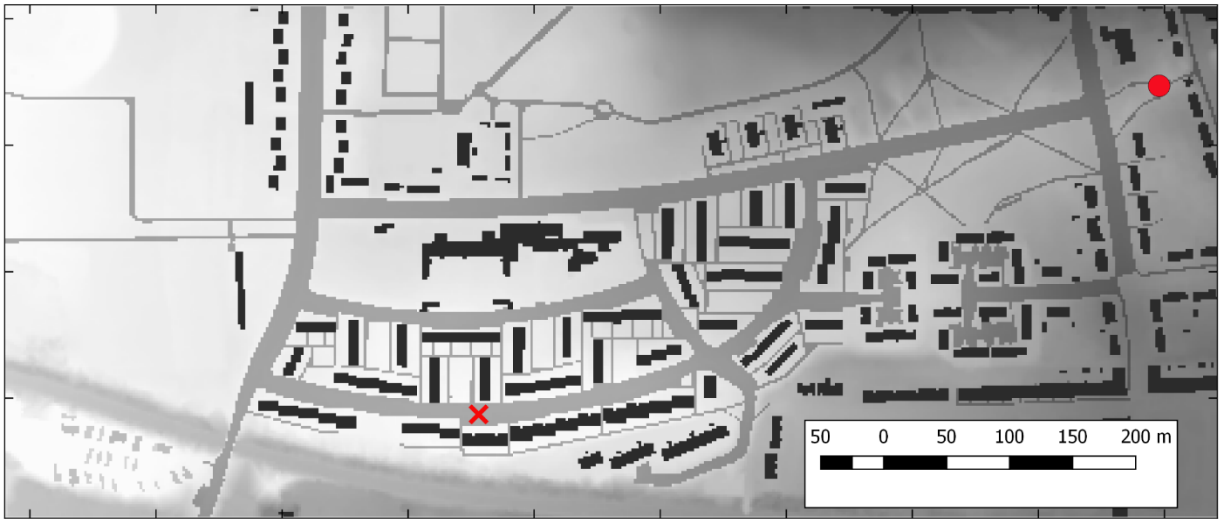


Figure 1: Model domain for Glasgow test case, where the red circle represents the culvert input and the red-cross is the model assessment location

The input ensemble defined in Table 1, created 3010 model runs. In the absence of observed data, a high-resolution benchmark model was used to act as a proxy to observed flood depth and extent data. The benchmark was created by resampling the original 0.5m LiDAR DEM to 0.1cm, with Manning's *n* values of 0.015 used for areas classified in OS Landranger data as roads and pavements, and 0.045 for the rest of the domain. Resampling the DTM is not used to improve the precision of the terrain data but to increase the number of calculation points in the model and improve the precision in the numeric model. This approach of using high resolution benchmarks to validate flood models has been used in previous studies (Fewtrell et al 2011). Here, the benchmark model is not being used to validate the

models, but to provide a dataset with which to evaluate the ensemble model runs with standard flood modelling evaluation methods.

The second test case is a river overtopping event, which occurred in Mexborough, South Yorkshire in July 2007. This was part of the larger UK Summer 2012 flood event that lead to 6 deaths and over £6 billion of damages. The model region is predominately rural with few modifications to the floodplain or the channel. The main flood mechanism is the overtopping of the largely unmodified river banks caused by the gradual increase in water levels in the channel. This test case provides a useful counterpoint to the test case in Glasgow, which was dominated by transcritical and low-depth, high-velocity flows. For this test case the main hydraulic properties and flood mechanisms are gradual varying and are characterised by low velocities. It can be assumed that the level of physical representation in the model will be less significant in impacting model results. Theoretically, the LISFLOOD-FP floodplain solvers should contain a level of physical representation to be able to replicate the main flood processes, and the requirement of conserving momentum in the model is less important in replicating this event.

A series of post-event wrack marks were surveyed with RTK GPS, which provide point information about flood depths and extents (Neal et al 2009). The position of the wrack marks was used to derive a water surface for the event (as seen in Figure 2) to compare the model run output.

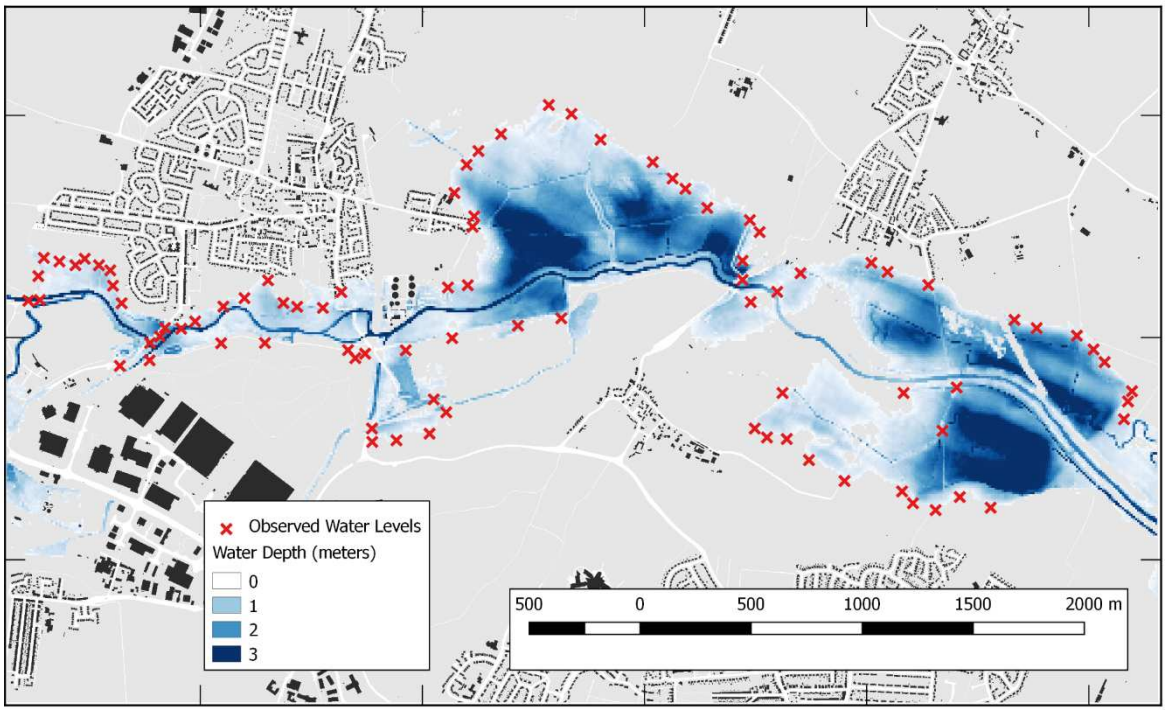


Figure 2: Model extent for the Mexborough test case, with the surveyed wrack mark points (red crosses) and the calculated observed water surface

The base model setup uses river gauge data for the inflow boundary conditions and LiDAR terrain data, with the model domain defined around the limits of the wrack mark data. The inflow is taken from 15-minute discharge data from the Bolton-upon-Dearne weir gauge which is located to the west of the model domain. The channel is represented as a series of wet cells located along the river centre line, rather than an explicit representation of the channel in the model. The downstream boundary is represented as a normal depth boundary condition which was tested to ensure that the boundary did

not influence model results through drawdown or backwater effects. Due to the larger model domain and associated computational costs, the event is defined as being 15 hours, rather than the 7 days of the actual event, but this is sufficient to capture the peak of the event.

The input sample space for this test case is adjusted slightly from the Glasgow test case, with fewer friction surfaced sampled. This leads to a total of 2020 model realisations, 505 model runs for each numerical solver.

2.4 Model Evaluation

For each test case, two model evaluation techniques were used. Common to both test cases is the use of the $F^{(2)}$ binary flood extent comparison method, where modelled extent is compared to observed extent (Aronica et al., 2002). The function ranges from 0 to 1, where a value of 1 represents a perfect fit between observed extent and modelled extent. This method for evaluating model output has been criticised for failing to distinguish between models where the terrain constrains the limits of flooding (Prestininzi et al., 2011) and for being only a small component of the complete evaluation of risk (Stephens et al., 2012). It still provides an indication of model performance, as well as providing a model evaluation technique that incorporates some of the spatial aspect of flooding. For the Glasgow test case, the second model evaluation is made by comparing water depths at the location of the red-cross point in the model domain. The Nash-Sutcliffe efficiency function is used to evaluate the differences in flow between the benchmark model and the model run. As with the $F^{(2)}$ approach, this has been criticised in the past for favouring peak values (Pappenberger et al 2004). For this test case, the peak value is of interest and this function is used for that purpose. For Mexborough, the second model function for the test is a comparison of water depth across the model domain using a Root Mean Square error measure (RMSE) as used by Stephens et al. (2012). The skill of replicating observed depths as opposed to extents alone can provide more insight into model ability (Apel et al 2009, Bates 2010, Stephens et al 2012). For this test case, both the modelled depths and extents were compared to the observed depths and extents using a RMSE measure and the $F^{(2)}$ function.

The test cases were analysed with a set of common approaches. An uncertainty flood extent diagram (also referred to a probabilistic flood extent in Aronica et al 2002) was created for each of the LISFLOOD-FP modules. Using the maximum flood extent from each model run, a cell is classified as flooded or not flooded, and given a score of 1 or 0. The average cell value is then calculated from the ensemble of model results, with a value of 1 representing a cell that has flooded in all test cases and a value of 0 a cell that has not flooded in any test cases. This value is described as the flood frequency value to distinguish it from a probabilistic estimation of uncertainty values.

A hazard uncertainty diagram is also used, and is based on a similar principle to the uncertainty flood extent diagram (Aronica et al., 2012). The level of hazard for each cell is calculated at each timestep using Equation 5, where d is the depth of water in the cell at a timestep $and\ v$ is the associated velocity

$$H = d_t * (v_t + 0.5) \quad (5)$$

From each model run the maximum hazard value for a cell across the runtime is calculated. Using this output, a cell that has a maximum value that exceeds 0.7 is defined as hazardous and given a value of

1 as a model output. The average cell value across the model ensemble is then determined, where values of 1 are cells that have been defined as hazardous in all model realizations. These diagrams allow major flood flow paths to be identified in the model domain, as areas of higher, faster flow that defines major flow paths can be determined from the regions defined as hazardous in the diagram.

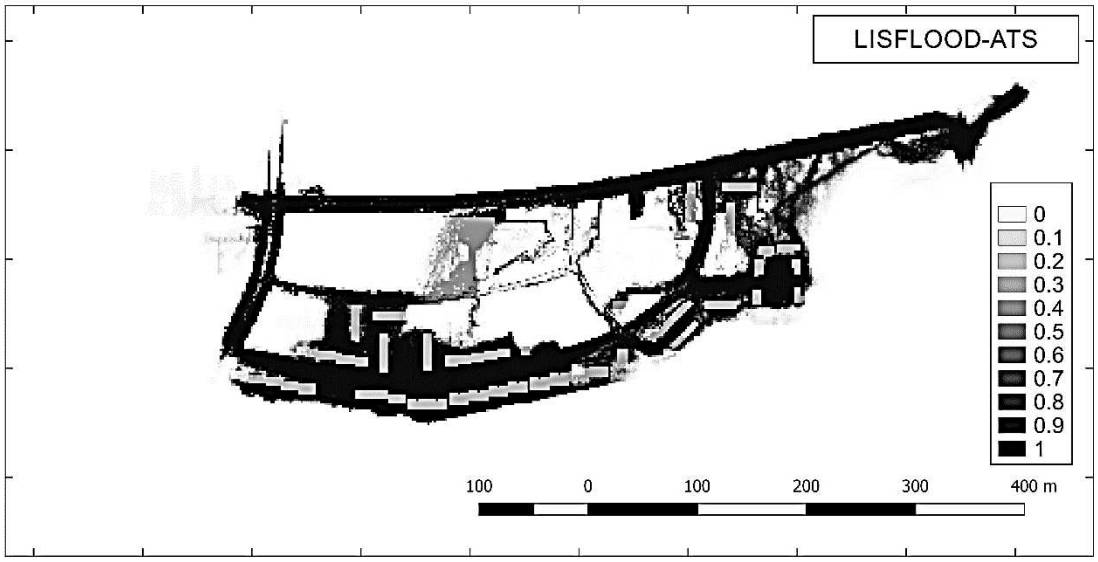
As well as investigating the main sources of uncertainty, it is also important to analyse second order input effects on model outputs, where combination of parameters may create a significant influence on model outputs. Typically, this could be done with variance-based sensitivity analysis approaches (Hall et al., 2009), although these are computationally expensive. In this paper, grouped interaction plots were used to display how parameters influence results and the impact with each other, and is described in detail in Section 4.

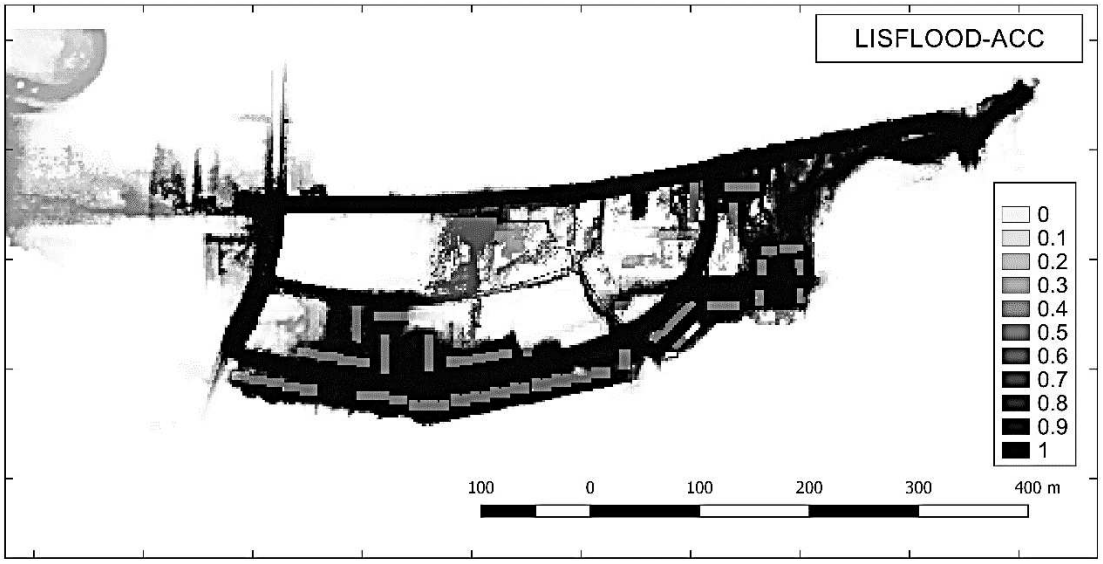
3. RESULTS

This results section is split into two sections: a comparison of the direct model outputs (where the direct model outputs include water depths and hazard ratings) and an evaluation of the model evaluations (such as the $F^{(2)}$ function). For both test cases the direct model outputs indicate that the model complexity has a strong control on model output. This is demonstrated in the uncertainty flood extent diagrams for each of the LISFLOOD-FP floodplain solver modules in the ensemble, displayed in Figure 3 and Figure 4. The range of flood frequency values for each module are summarised in Table 2;

Table 2: Summary of number of cells in each flood frequency bin for the Glasgow case

Model Type	0	0.1 -0.3	0.3-0.6	0.6-0.9	1
ATS	77899 *(79.7%)*	2645 *(2.7%)*	4126 *(4.2%)*	1781 *(1.8%)*	11317 *(11.67%)*
ACC	70874 *(72.5%)*	6945 *(7.1%)*	5116 *(5.2%)*	1488 *(1.5%)*	13345 *(13.6%)*
Rusanov	60842 (*62.2%)*	8946 *(9.2%)*	7409 *(7.6%)*	3705 *(3.8%)*	16866 *(17.3%)*
Roe	69775 (71.4%)	4482 *(4.5%)*	6487 *(6.6%)*	2785 *(2%)*	14239 *(14.5%)*


LISFLOOD-ATS
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100 0 100 200 300 400 m


LISFLOOD-ACC
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100 0 100 200 300 400 m

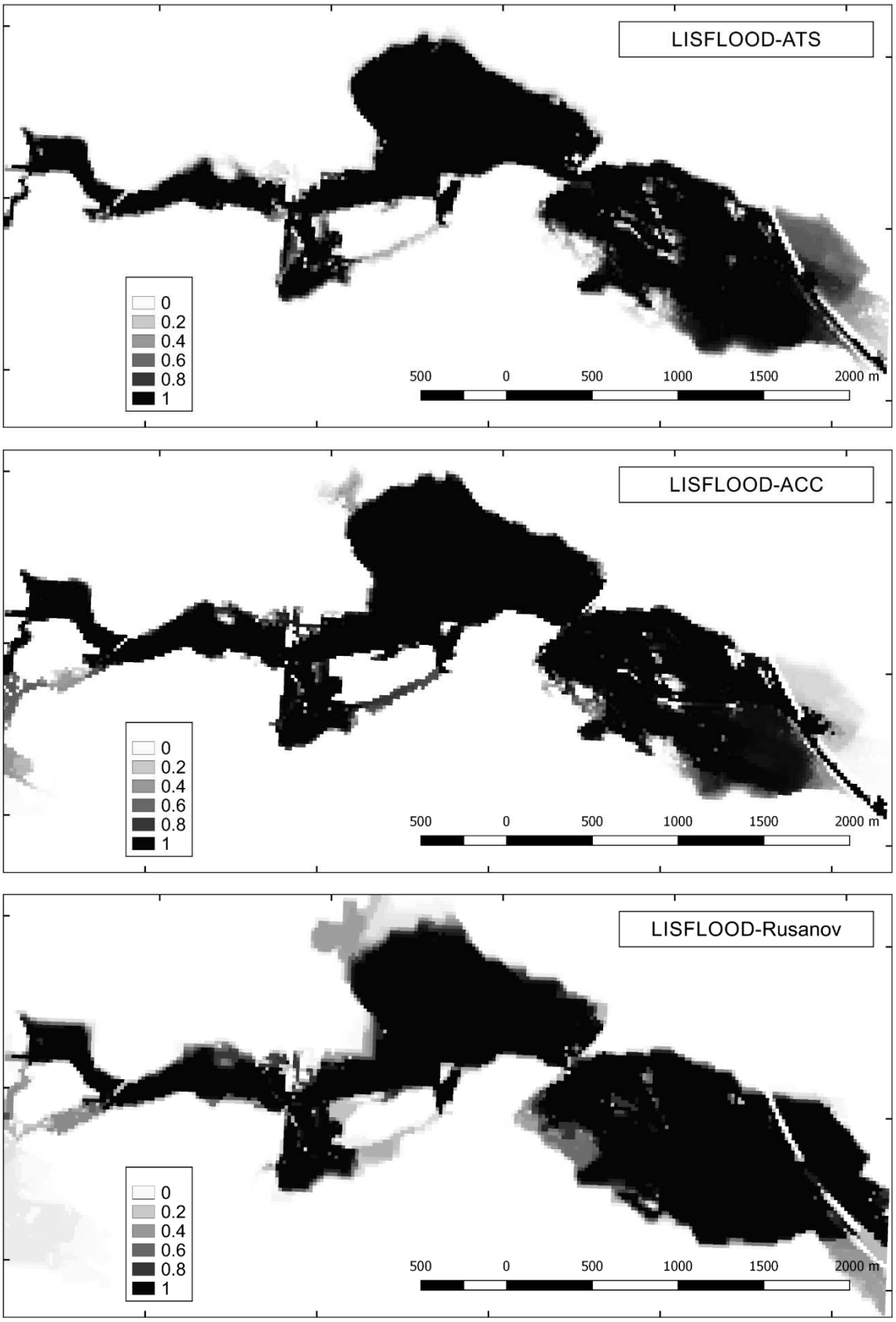

LISFLOOD-Rusanov
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
100 0 100 200 300 400 m

Figure 3: Uncertainty flood extent plots for the LISFLOOD-FP modules for Glasgow

The output is similar to results seen in the benchmarking paper of Hunter et al (2008). The simplified 2-term ATS solution is unable to overcome a significant topographical blockage at the end of the main flow pathway due to the lack of conservation of momentum over complex terrain (Neal et al 2011). In comparison, the full SWE models of LISFLOOD-Roe and Rusanov, are able to overcome this blockage and produce wider flood extents on the left-hand side of the model domain. This is perhaps surprising, considering that the inflow hydrograph is varied by 20% in the parameters space that may have provided the ATS solution with a greater volume of water to overcome the blockages in the topography. This demonstrates that uncertainty in the boundary conditions may not account for variations between numerical models. The ACC in comparison, can overcome this blockage, although the difference between this and the full SWE, both in terms of extent and flood frequency value, indicates that the exclusion of the advective term is still significant in conserving flow momentum in this flow path. The modules that use the full SWE by comparison, regularly flood this section, although distinctions occur between the flood extents for these solvers. For example, the Rusanov solver produces a number of models runs that flood the western extent of the flood map, as can be seen in the high flood frequency values in Figure 3, but is combined with a number of cells in this region classified as uncertain (values between 0.1 and 0.9 in Table 2). The total of these cells is as large as the number of cells predicted as flooded in all simulations, 20.5% vs 17.3%. By comparison the Roe and ACC contain a more even ratio of uncertain cells vs flooded in all model runs (13.9% vs 13.6% and 14.1% vs 14.6%). For both solvers, the area of the model domain that contributes the greatest number of uncertain cells is the western extent. In comparison the ATS solver shows a level of insensitive to the input parameters. This solver has a lower number of uncertain vs flooded cells (8.7% vs 11.6%), and fewer regions of uncertainty.

The differences between the solvers in this test cases suggest that the precise value of Manning's n, cell size and building representation are relatively unimportant in comparison to terrain data and the numeric model.

There are smaller variations in uncertainty flood extents of the four LISFLOOD-FP modules for the Mexborough test case (Figure 4).



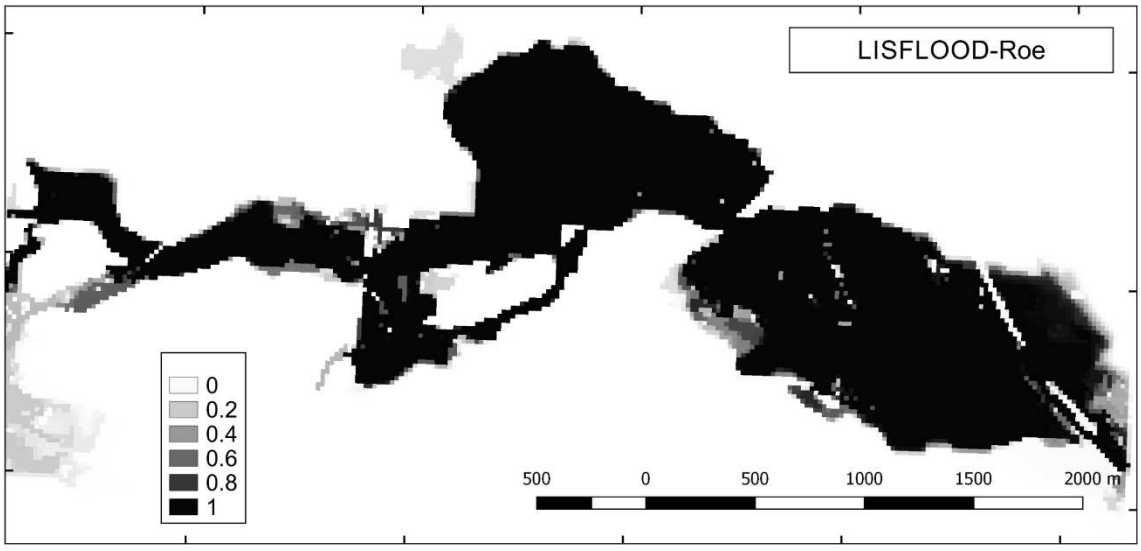


Figure 4: Uncertainty flood extent plots for the LISFLOOD-FP modules for Mexborough

Table 3: Summary of number of cells in each flood frequency bin for the Mexborough case

Model Type	0	0.1 -0.3	0.3-0.6	0.6-0.9	1
ATS	23727 *(74.5%)*	784 *(2.5%)*	760 *(2.4%)*	715 *(2.2%)*	5871 *(18.4%)*
ACC	23851 *(74.9%)*	800 *(2.5%)*	505 *(1.6%)*	406 *(1.3%)*	6295 *(19.8%)*
Rusanov	21051 *(66.1%)*	1719 *(5.4%)*	1309 *(4.1%)*	645 *(2.0%)*	7133 *(22.4%)*
Roe	22486 *(70.6%)*	878 *(2.8%)*	719 *(2.3%)*	473 *(1.5%)*	7301 *(22.9%)*

Each of the solvers produces a similar extent across the model input space, with high levels of flood frequency values, as can also be seen in Figure 4 and Table 3. This is due to the constraining nature of the floodplain and the scale of the event. However, small distinctions between these plots provide insight into the influence of the level of physical representation. Overall, the ATS and ACC solver results contain few regions of low flood frequency occurrence, with most model runs producing similar extents and with low ratio of uncertain cells vs flooded cells (7.1% vs 18.4% for ATS and 5.4% vs 19.8% for ACC, Table 3). For both solvers the area with high levels of uncertainty is located in a region to the east of model domain, which is located behind a gap in the northern river bank. By comparison, the full SWE modules produce high flood frequency values in this region. This difference is due to the ACC module producing a slower moving flood wave, which fails to inundate this region in the simulation time and the ATS solver failing to overtop the bank due to the lack of suitable physical processes in the model. The simulation time for the event has caused an unintentional bias against the ACC module.

The model runs that have flooded this region have input parameters with values in the extremes, for example low friction values, and the upper limit of the hydrograph. This highlights a common issue in calibrating model that the combination of inputs in the limit of acceptable values may be considered physically unrealistic, and using this models going forward from a calibration process would reduce confidence in the models' ability to predict future events.

The two full SWE modules also produce similar results to the Glasgow test case, in that a higher percentage of flooded cells are observed in comparison to the simplified approaches (22.4% of all cells for Rusanov, 22.9% for Roe), although the Rusanov solver has a higher percentage of uncertain cells (11%). In this test case, there are a number of regions in the model domain with moderate levels of uncertainty that provide insight into the difference in the outputs of the models (Figure 4). The first region is near the eastern boundary of the domain, where secondary channels allow water from the main channel to inundate a region of the model through the backwater effect. In this region the ATS fails to flood any cells, unlike the other 3 modules. The ACC solver produces a small area of flooding with an average flood frequency cell value of 0.4. The Roe and Rusanov solver produce a wider flood extent but with different levels of frequency (0.2 on average for Rusanov, 0.4 for the Roe solver). This level of flooding is consistent with the observed wrack marks (Figure 2). The infrequency of flooding in this region across the different modules demonstrates the difference not only between simplified and full physics approaches in solving local scale flow patterns, but also show a high level of uncertainty relating to the choice of parameters.

A second region of interest is a storm drain located in the centre of the model domain which also shows variation between the modules. The ATS, ACC and Roe solver produce a similar extent with increasing frequency value as the level of physical representation increases. This region is protected from the main river flow by a large embankment. As with the first area of interest, flooding of this channel occurs due to backwater effects and overtopping of a bank. The ATS solver flood this region only for model runs which include higher inflow levels, and allow the module to overtop a small bank. For both the Roe and ACC solver this region consists of high flood frequency values indicating consistent flooding of this section, which is again consistent with the observed flood extent in this region in Figure 2. By comparison the Rusanov produces a wider flood extent, but with lower frequency values, in part due to instabilities that occur in the module in this region. These small sections show that small variations in modelled extents may occur due to local scale hydraulic effects even when the main flood mechanisms for an event are captured by a range of approaches to solving floodplain flow.

Further insight into the significance of the numerical model can be determined with the hazard uncertainty figures (Figure 5 and Figure 6). Both test cases show differences between the simplified approaches and the fully physics based approaches. For the Glasgow model, the main flow paths can be seen in the outputs for all the solvers, which show the main east–west flow path and the secondary north–south flow path with high hazard frequency values. For the ATS module however, there are more cells with values of between 0.3 and 0.6, representing areas of uncertainty. This is particularly noticeable to the western edge of the model domain, where velocity reduces significantly. The other modules demonstrate consistency in replicating the flow paths (this can be interpreted from the high frequency hazardous values along the main flow path), but with more areas of uncertainty than had been seen in the flood frequency diagrams. The pattern of uncertainty is also discontinuous as a result of variations in the topography causing small differences of depth and velocity from cell to cell, which leads to different hazard index values at the cell level. This effect, and the extensive regions of

uncertainty for this figure, indicate that all inputs in the model ensemble may impact combined velocity and depth outputs from the model runs.

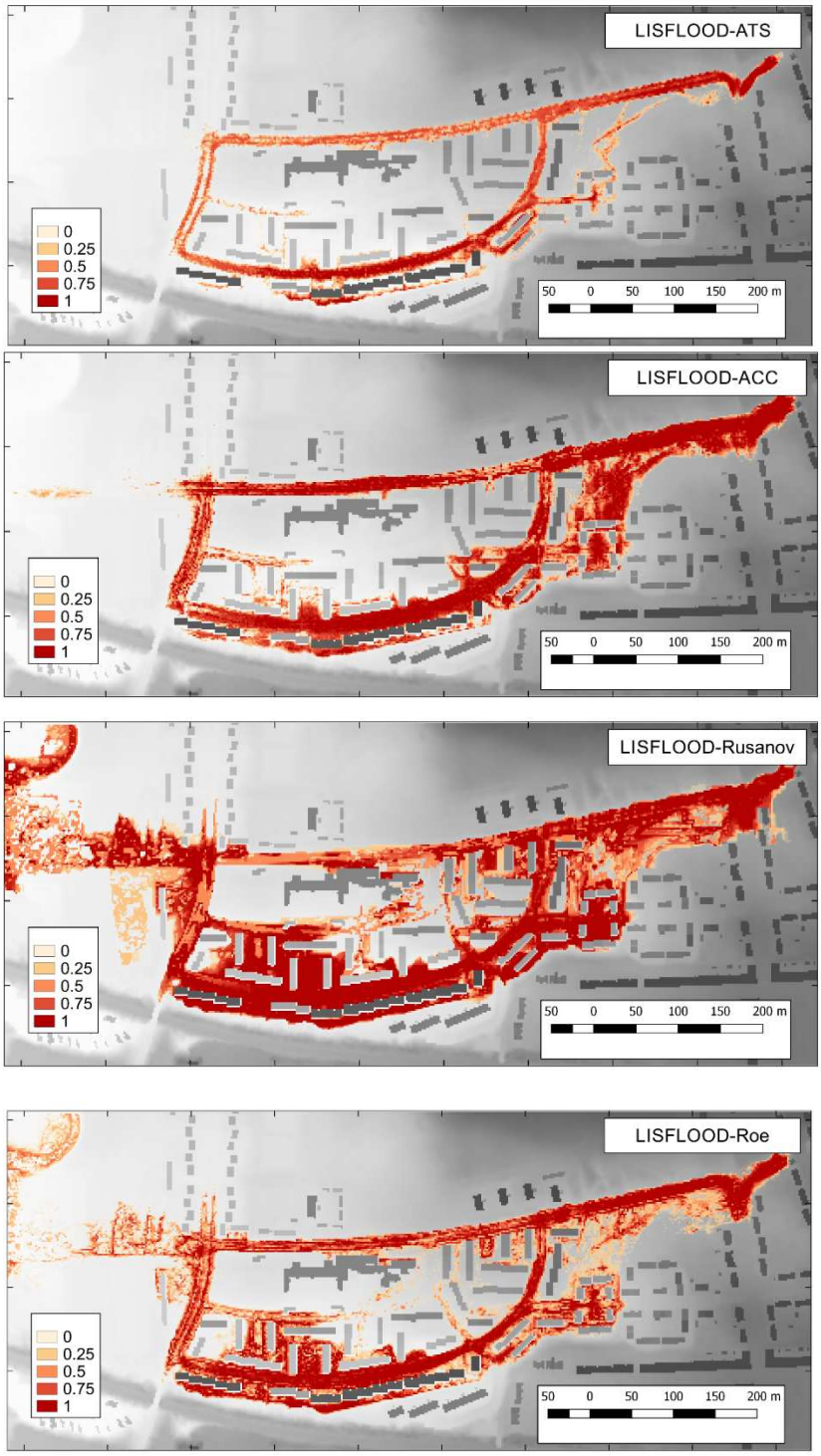
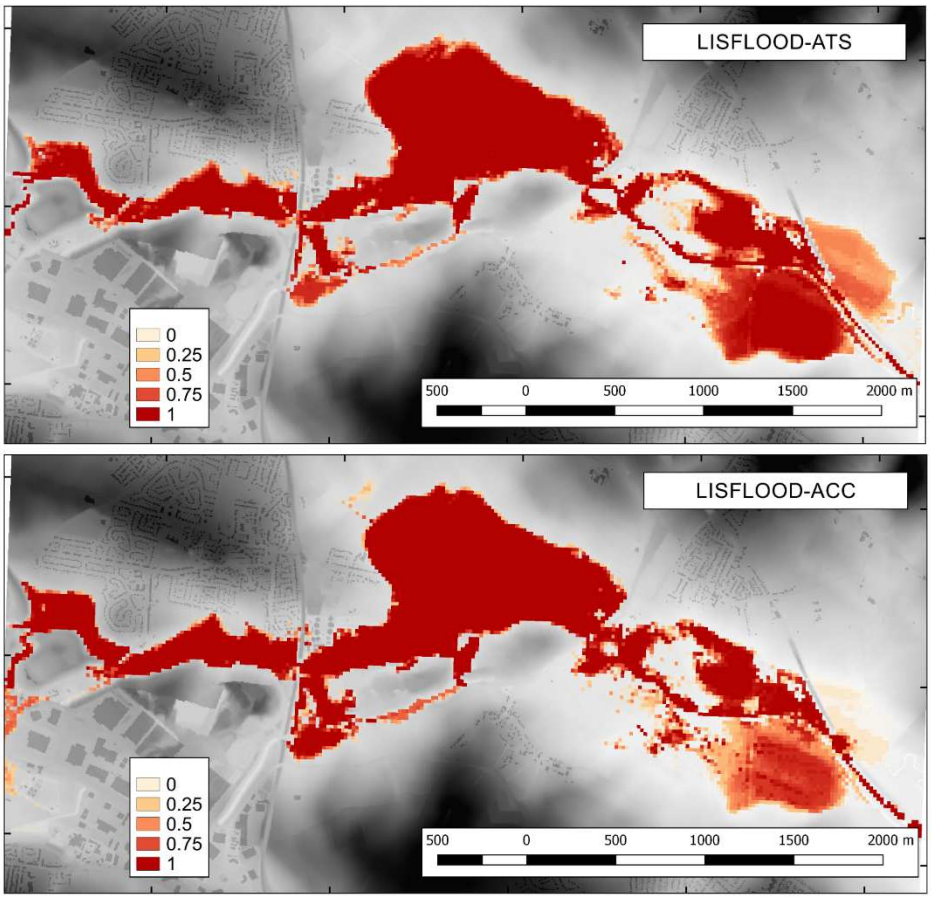


Figure 5: Hazard Uncertainty plots for Glasgow

The hazard uncertainty plots for the Mexborough test case show less variation than the Glasgow test case (Figure 6), but there are some distinctions between the simplified and full physics approaches. The ATS and ACC modules produces a narrower, high frequency hazard extent, whereas the full SWE models create a similar hazard extent to their uncertain flood extents, with high values of hazard frequency throughout the domain. As with the uncertain flood extent figures, one of the main differences

between the modules and the flood extents of Figure 4, relate to the smaller channels that contribute to flooding to the west of the model domain. This leads to either high levels of uncertainty, or as in the case for the ATS module, no cells having a hazard level over 0.7. As the flood wave moves across the domain, there are further differences, with the ATS and ACC solvers producing progressively lower hazard values in the eastern side of the domain in comparison to the full SWE approaches. This is related to differences in the velocities of the water at this point in the model domain which is not apparent in the analysis of the flood frequency figures. The full SWE approaches contain few regions of uncertainty. The areas of uncertainty are similar to those identified in Figure 4. This similarity, which can also been seen in Figure 5, suggests that these module are able to cope with a range of uncertain parameters and can reproduce critical flood processes, such as transcritical flow along main flow paths, regardless of the input. Comparing across these plots it would appear the location of the maximum flood extent contains greater regions of uncertainty than the hazard uncertainty plots.



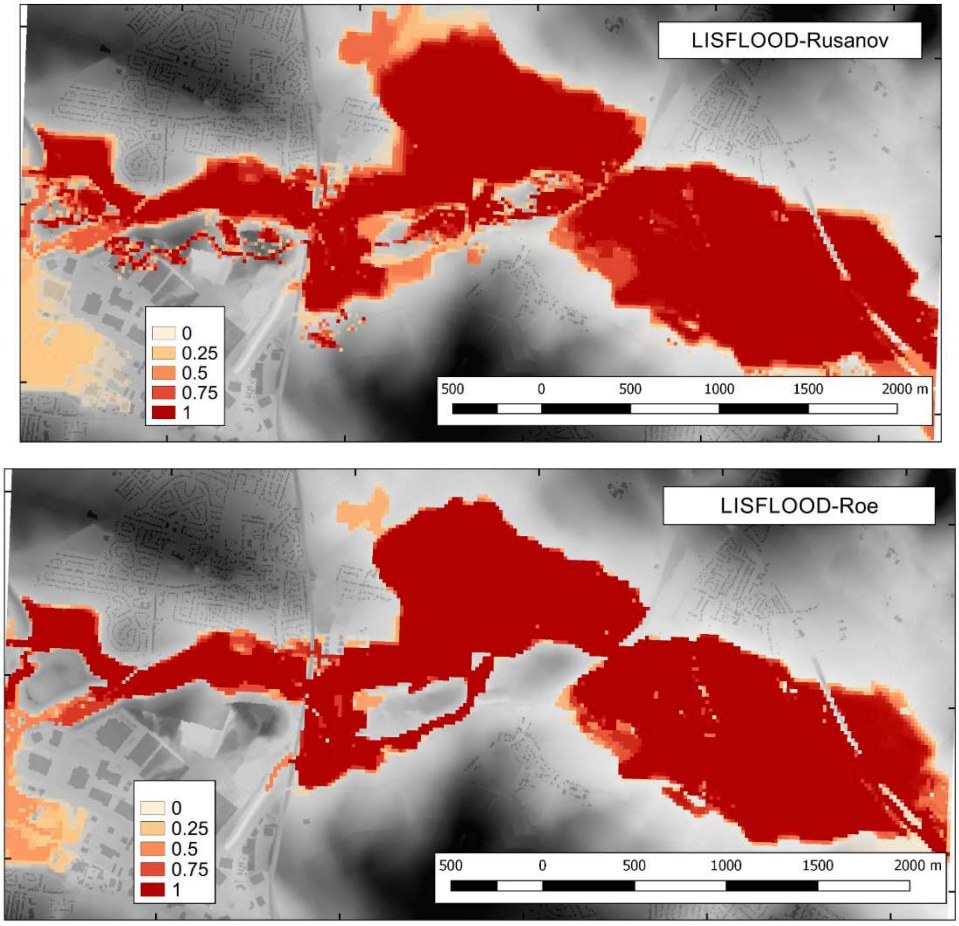


Figure 6: Hazard Uncertainty plots for Mexborough

Whilst the direct model outputs provide insight into how each module represents particular flood mechanisms across the model domain, the objective functions elaborate on this further. A comparison of the spread of $F^{(2)}$ results across the 4 different modules and for both test cases is made using box and whisker plots in Figure 7and Figure 8.

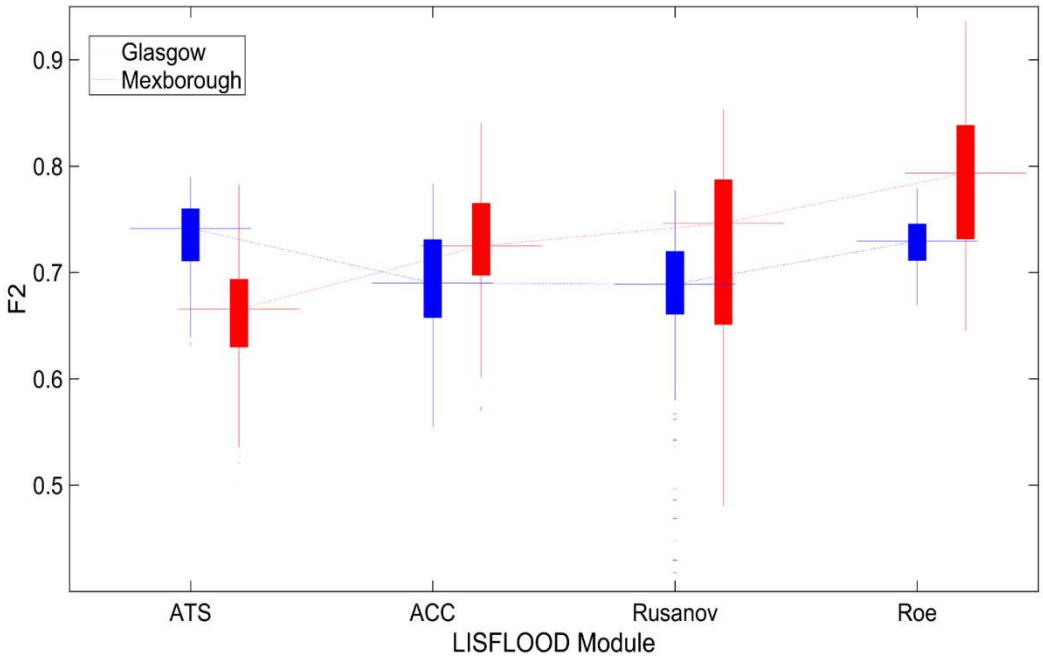


Figure 7: Box and whisker plots for analysis of $F^{(2)}$ for each LISFLOOD module for both Glasgow test case (red boxes) and Mexborough (blue boxes)

Figure 7 demonstrates how the significance of the physical model changes across the two test cases. Overall, the Glasgow test case demonstrates a monotonic relationship between the complexity of the floodplain flow solver and an increasing mean $F^{(2)}$ value (red boxes), whilst the Mexborough model results display both a narrower range of model results for each module, with a similar range of results (around 0.6 to 0.8 for all models). This can be seen by comparing the line of median model results in Figure 7. For the Glasgow case, this is perhaps not surprising considering the use of a benchmark model to compare results but still shows the influence of physics approaches in this test. By comparing the two lines, it can be seen that the ranges of results for both test cases varies between the LISFLOOD modules for each of the test cases. The extent of the boxes (which represent the range between the 25^{th} and 75^{th} percentiles), is larger for the Glasgow test case, which shows that the modules are more sensitive to the input ensemble when the hydraulic characteristics are more complex. This figure demonstrates that the simplified physics models have some ability to replicate the full physics model results for part of the parameter space. In the Mexborough test case, the highest performing model across the input ensemble is an ATS driven model ($F^{(2)}$= 0.78), although the difference between the best performing realisation from each module is less than 0.01 in the $F^{(2)}$ function. For the Glasgow test case, this becomes less clear, but the range is reasonable in comparison to the full SWE modules and indicates that the diffusion wave approach still replicates a significant part of the baseline modelled flood extent test. Comparing the simplified approaches, the ACC results show an overall increase in the $F^{(2)}$ value for Glasgow compared to the ATS and demonstrates that a slight increase in model complexity has an impact on model results. However, for the Mexborough test case, this module has the widest range of results and the lowest median score. This is in part due to the sensitivity to the cell size where at the higher cell size of 40m, the overall $F^{(2)}$ value drops and suggests this module has a level of sensitivity to the range of input parameters. For the Glasgow test case, the full SWE models, produce

both higher median values and higher peak 95 percentile points than the simplified solutions. There is a distinction between the two SWE modules in both test cases, with the Rusanov numerical solver producing a greater range of results than the Roe solver. Further, this method also has lower percentile values which exceed the lowest $F^{(2)}$ values of the simplified models in all test cases. Both of these points show that the Rusanov solver not only has the highest level of variation in the model outputs across the input ensemble, it also indicates that the numerical structure of the computational code is also significant factor in modelling uncertainty. For Glasgow, the range of results for the Rusanov solver is large (0.65 to 0.95), but is fairly insensitive to the range of inputs in the Mexborough test case (0.66 to 0.75). For the Roe solver, this appears to suggest that as flood mechanisms increase in complexity, the Roe solver will produce more varied results. This in part due to differing responses in the module to the parameter inputs and the flood mechanisms, where local hydraulic processes such as flow over bumps are more sensitive to the input parameter, as seen in Neal et al (2011). The Rusanov solver is more sensitive to these inputs by comparison which demonstrates the significance of the numerical model as a source of uncertainty.

The other objective functions used to evaluate the model results (Nash-Sutcliffe and Root Mean Square Error of depths), are also summarised with boxplots in Figure 8. Whilst the two methods are not directly comparable, the comparison of the two results demonstrate that the second evaluation technique responds differently to the model inputs. The Mexborough results appear to be similar for all the module types for both the RMSE and $F^{(2)}$ evaluation functions, whereas the Glasgow results now show different characteristics between module types. For the Nash-Sutcliffe results, the variation between the different LISFLOOD modules is now no longer monotonic, and a significant amount of variation between each solver exist. The ATS solver produces the highest peak median and 75th and 25th quantiles (0.88-0.94), whilst both the ACC solver and the Roe solver produce a similar range of model results, around 0.75-0.92 (although the Roe solver produces a higher median value of 0.87, compared to 0.82). The narrow range of results for the ATS module indicates insensitivity to other parameters, whilst the wide range of results for the Rusanov solver show a high level of sensitivity, which confirms the conclusions of the analysis of $F^{(2)}$ and Figure 7. The Roe solver demonstrates a degree of robustness against the range of parameter inputs and produces smaller range of possible depths than the Rusanov solver. This in part due to amount of numerical diffusion to the solution that can reduce the peak values of depth and discharge. This is particularly noticeable when comparing modelled depth outputs for spot locations, and further demonstrates the importance of the numerical solver as a potential source of uncertainty.

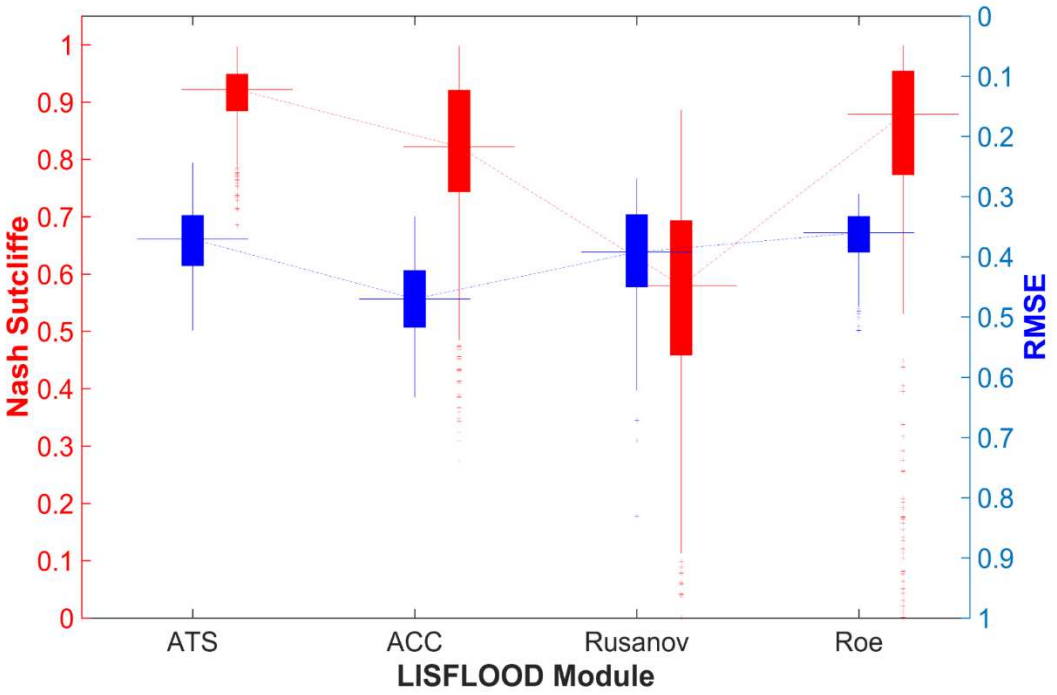


Figure 8: Box and whisker plots for analysis of Nash-Sutcliffe coefficient (Glasgow, red) and RMSE for Mexborough (blue) for each LISFLOOD module.

The RMSE measure displays less variation between both the LISFLOOD modules and between the box plots of Figure 7. The range of results is between 0.3 and 0.7, indicating that all simulations produce reasonable approximations to the observed depths. As with the $F^{(2)}$ evaluation, the ranges of results appear to be very similar but further details of how each solver operates can be demonstrated. The ATS solver produces the highest level of model performance (RMSE ~ 0.24), but produces a wider range of model results with a higher mean value than the Roe solver. By contrast, the ACC produces a similar range of results, but with the lowest model performance of the modules. This occurs because the ACC solver produces a higher peak water depth as the flood wave moves across the model domain in comparison with the other solvers. This difference between the two simplified approaches highlights that broad assumptions about how simplified approaches may perform for given hydraulic conditions may not always be straight forward. The full SWE based models also display similar responses to the input ensemble as has been seen in the analysis of the $F^{(2)}$ function. The Rusanov solver demonstrates similar behaviour of greater sensitivity to parameter input, and the Roe solver appears more robust to the parameters. This further emphasise the importance of the numerical model in the consideration of uncertainty in flood models.

These insights into module performance are further confirmed by analysing how a model performs with respect to multiple evaluation techniques (Dung et al 2011). Figure 9 is a comparison space of the function results for each model run from the model ensemble for both tests. Each dot represents a model run with the corresponding $F^{(2)}$ value on the y-axis and the second function on the x-axis. The LISFLOOD-FP module used is indicated by the shape of the dot. The Pareto front, representing the region of the figure occupied by models with high values of performance function results, is formed at the top right of the axis space.

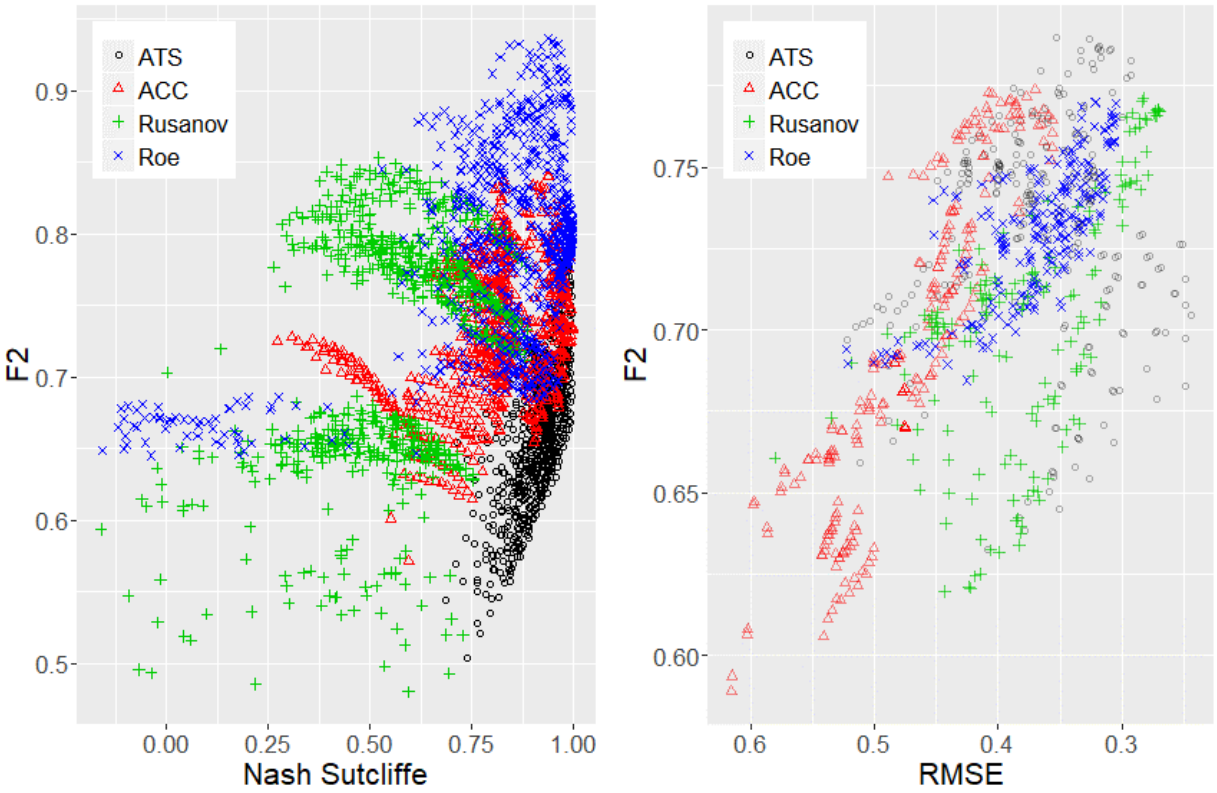


Figure 9: Combined results for $F^{(2)}$(y-axis) and second function (x-axis), for Glasgow and Mexborough test case (left and right) where each dot represents a model run. The pareto front of best performing models occurs in the top right section of the graph

These plots demonstrate that only a few simulations across the ensemble can provide good results for both objective functions. By displaying the model type as well in Figure 9, trends related to numerical complexity can be identified from the test cases. For the Glasgow test case, the region of best performing models is occupied by models computed with the Roe solver, whereas the Mexborough test case contains all model types. In the Glasgow test case results, the sensitivity of the full SWE models to the model parameters is evident by the wide spread of results. The simplified approaches occupy a narrow range, indicating the opposite. This indicates that although full SWE-based approaches can provide a high level of performance in this test case, they are more sensitive to uncertain parameters, and require detailed calibration. Of the simplified approaches, the ACC solver though demonstrates a greater sensitivity to the other inputs, than the ATS solver.

Analysis of the results from the Mexborough test case (left hand side, Figure 9) shows that a number of individual models can achieve a high goodness of fit level as well as a low RMSE value. For this test case, the Pareto front is not dominated by any individual LISFLOOD module, as the Roe solver does for Glasgow. By contrast with Glasgow results, the Rusanov code and the ATS code have a number of model runs that occupy the Pareto region, although overall these two codes produce model results that occupy a wide region of this plot. This shows that these codes are capable of replicating observed depths and extents to a high level of accuracy when suitable parameters are chosen, but only when a detailed calibration process has been conducted to select these parameters. Further to this, the analysis of these results suggest that the diffusion wave approach can replicate observed data where the flood

event is dominated by sub critical and low energy flows. This contrasts with the position of the ACC models in this figure, where most of the model runs appear in the centre of the figure, with a wide distribution across the x-axis. This again highlights the difference between simplified approaches in overall model performance. In comparison, the consistency of evaluation values from models using the Roe solver is apparent by the smaller area occupied by these runs. Whilst the code does not produce an outright best performing model, this demonstrates that the code is more resilient to the range of model inputs, and thus indicates that the model is robust to uncertainty.

Further insights into modelling uncertainty can be seen when comparing all the parameters from the input ensemble. Interaction plots provide a means to visually compare the influence of each parameter on the overall model function results. These plots are comprised of an array of plots, each of which compare the mean values of model results for 2 parameters, where the inputs being compared are denoted by the column and row headings. For the interaction plots used below (Figure 10, Figure 11 and Figure 12) the plot in the first column, second row is the comparison of LISFLOOD-FP module (represented along the x-axis) and the level of building representation (represented by the line). The position of the line is determined from the mean value of the model results (which is the dependent variable) where these two factors are shared. For example the mean value of models that use the ATS solver and the BP building representation inputs in Figure 10, is 0.68, whereas models that use ACC and BP building representation is 0.71. The overall significance of the input represented on the x-axis in the first order can be determined by the gradient of the line across the x axis, whilst the significance of the input represented by the y axis can be determined by the spread of the lines across the individual plot. The level of interaction of the two parameters can be determined from any crossings of the lines and the variation of two lines along the x axis, which indicate that the factor on the y-axis influences the factors on the horizontal axis in different ways. By looking broadly at an interaction plot, the significance of an input can be determined from the overall gradients of a line in either its corresponding row or column. These plots can be used to determine the effects of one parameter on another, and across the input ensemble. The plot provides a method to efficiently compare multiple model inputs, without relying on multiple plots of other evaluation techniques. Whilst the use of comparing mean values in this way may limit the insight into overall levels of interaction that a more computationally costly sensitivity analysis approach may determine, the significance of a parameter can still be understood from this approach.

Figure 10 and Figure 11 provide analysis of combined parameter performance for Glasgow and Mexborough for the $F^{(2)}$ function.

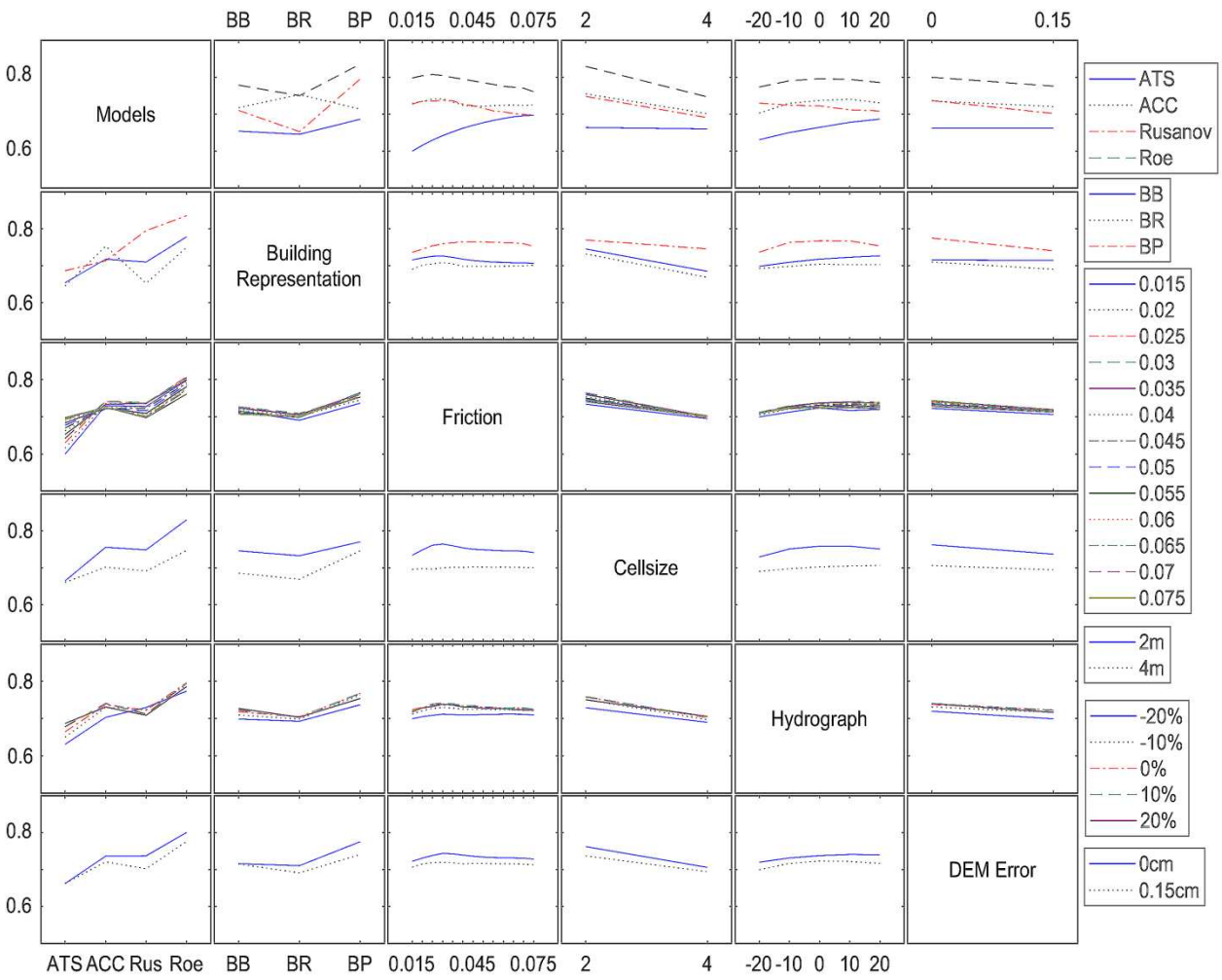


Figure 10: Interaction plot for each input factor comparing mean value of $F^{(2)}$ per level of factor vs other levels of factors for the Glasgow Test Case (where $F^{(2)}$ is the value on the y axis of all plots and the x axis in each column is the levels for the named parameter)

A number of conclusions can be drawn from Figure 10. The gradient of the lines in first column shows that the difference in mean for each of the LISFLOOD-FP modules is greater than the difference in mean for other inputs, as can be seen by inspection of column 1 in both plots. Here, the range of the mean results from each module is from 0.65 for ATS up to 0.8 for the Roe solver and this pattern is replicated in each of the rows of the first column, which shows that this factor will be more influential in affecting model outcome than the other parameters. The lower influence of other factors can be determined by the flatter gradients in the other columns. This shows that despite a change in level for a parameter, the mean result is essentially the same. For example, the 3rd column which represents the friction value, produces small variation across the x-axis for all the different model inputs – the range of the mean values is typically from 0.68 to 0.73. This implies that even for edge case friction values, the overall output of extent will be similar. The low significance of the friction parameter can be explained in part due to the constraining urban topography, which limits the extent to which water can flow across the model and thus limits the influence of this factor. Whilst direct first order impact of other parameters is low, analysis of the other inputs also displays some variations and interactions. The building representation method shows some interaction with module type (first row and second column). A small

upward inflection due to a change in the mean values can be seen when comparing the blue ATS line, the red Rusanov line, and the grey Roe line in this plot, as the building representation method changes from Building Resistance (BR) to the Building Porosity (BP) method. The ACC solver by contrast creates a higher mean value when using the building resistance method and a lower value when used with the BP method, leading to a downward inflection (grey dotted line). This change, although small is indicative of higher order interaction, and suggests the ACC solver response varies non-linearly to this parameter. The grid cell size also has an element of influence when compared to the LISFLOOD-module. A convergence of overall results is apparent as this factor increases, across all the input factors as well as the choice of numeric model. (This can be determined from the converging lines in column 4 row 1, and through the other plots in column 4). This suggests that with increasing cell size, the significance of the model complexity reduces, but at the cost of confidence and accuracy of model results. This response to cell size suggests that the 'rule of thumb' for determining the grid resolution for a model is appropriate. The convergence of results for the model type, and at a lower mean value show that micro scale hydraulic processes are not as well replicated at the higher grid resolution, and the overall model is less representative of the underlying processes of the event.

The DEM error and the hydrographic input appear to be less significant as a primary source of uncertainty which can be determined from the low gradient variations of the columns 5 and 6. The importance of the friction parameter has been noted as being of low significance in primary effects, but there is influence in the higher order effects. The plot in column 3 row 1, shows that at low friction values, there is a wide range of model results for the 4 LISFLOOD modules (represented by the lines). Moving along the x-axis and with increasing friction value, there is a convergence of model results between the levels of physical representation, until the lines converge at the end of the x-axis, representing the highest level of friction values. It can be inferred that at higher friction values, the significance of the model structure uncertainty will reduce. This plot also underlines that although model complexity is significant, interactions with other parameters influence the level of this significance in this test case. This suggests that whilst the values adopted in an urban model will be far less critical in creating uncertain input, than the initial choice of numerical model, they will still exert control on model results.

By comparison there is relatively narrow range of $F^{(2)}$ values in the model results of the Mexborough test case. In order to demonstrate this, the Y axis of Figure 11, representing the mean $F^{(2)}$ value is scaled to the same range as Figure 10, to demonstrate the lower range of results.

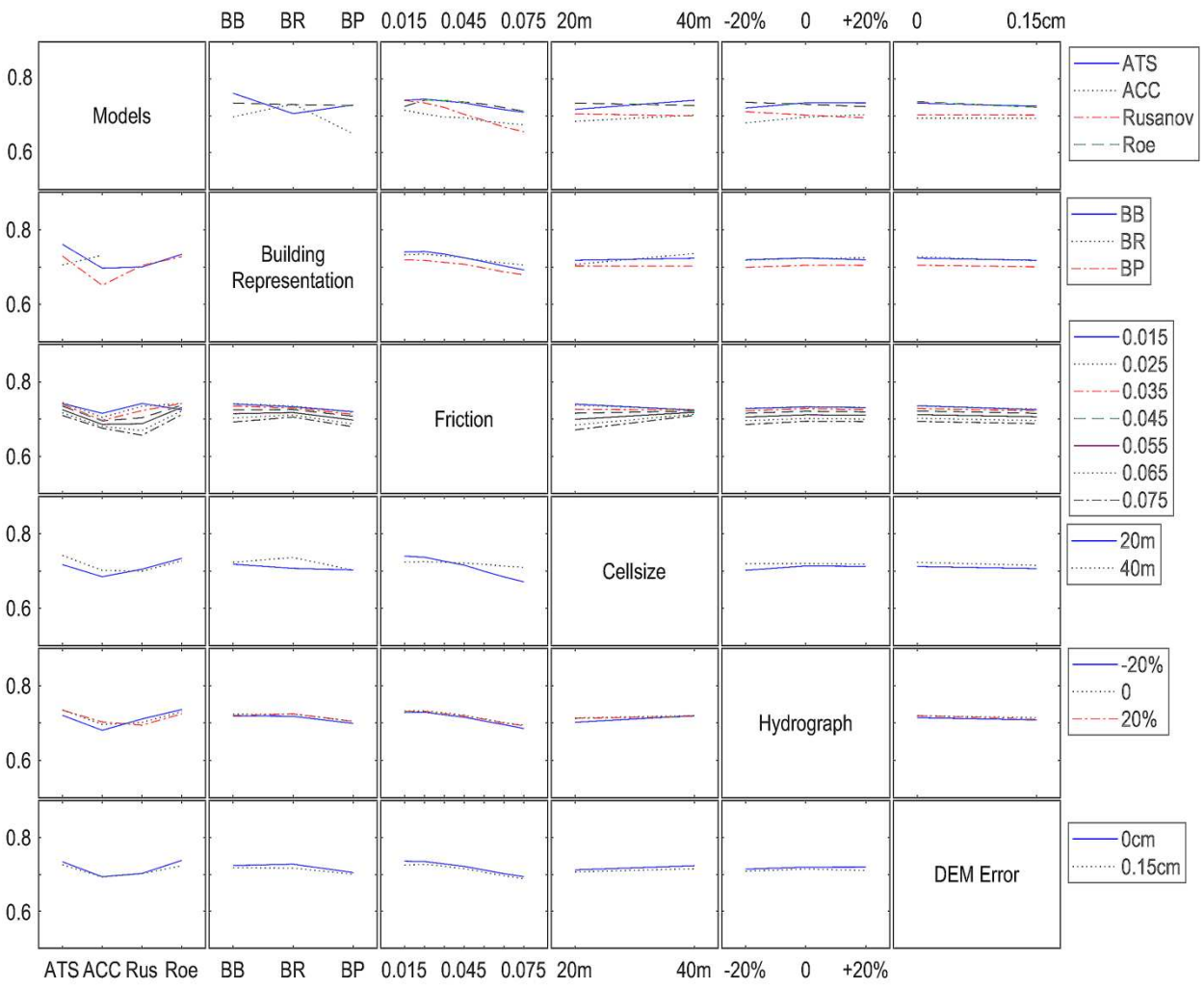


Figure 11: Interaction plot for each input factor comparing mean value of $F^{(2)}$ per level of factor other levels of factors for the Mexborough Test Case

As is seen in Figure 10, model type and the friction value have a significant control on the overall model results. The gradient of the lines representing the model types (column 1), shows a decrease in the mean value from ATS to ACC, then rise with increased representation of the full SWE models, which reflects findings of the earlier analysis. In comparison with Glasgow though, other parameters exert greater control on model results. For the friction value input (column 3) there is a clear decrease in $F^{(2)}$ value with increasing Manning's n, which indicates more primary influence on model results. A large part of this influence is caused by the sensitivity of the Rusanov solver to this value, which can be seen in the plot of column 1, row 3, where a large variation in results occurs for Rusanov solver position on the X-axis. Although buildings occupy a smaller area of the model domain, the use of a Building Porosity model appears to have a negative impact on model result, as can be seen across the plots of Row 2, where the line representing this level is consistently below the other building representation methods. This is due in part to buildings located around the floodplain nearest to the model inflow location. This is the opposite pattern that is seen in the Figure 10 and suggests that the building representation method is an important influence on results, regardless of the underlying hydraulic conditions.

As with the Glasgow test case, higher order parameter interactions appear to be relatively limited, with few crossings of lines in any of the columns. An exception is between cell size and friction value (column 4, row 3), which shows a converging of model performance when both factors increase. As has been seen in Figure 10, there is an improvement in the model results for the simplified approaches and the higher grid resolution (column 1, row 4). Both points demonstrate that the increase in cell size appears to reduce the significance of other factors in altering model output.

As the variations of the $F^{(2)}$ results are small, further insight to the results of the Mexborough test case is made by analyzing the other model function, RMSE (Figure 12). This figure shows that the LISFLOOD-FP module is the most significant factor in affecting results, as can be seen in the gradients of the line in column 1. This significance is a result of the ACC module, which skews the results as the other 3 modules appear to produce roughly the same mean value module. This can be seen in the top row, where the mean ACC results, represented by the dotted line are far higher than the other 3 modules. The effects of other parameters on module performance can also be seen. The Roe solver is more intuitively responsive to friction parameters (column 3, row 1), where the dashed grey line increases in RMSE values at the edges, and decreases in the center, where the Manning's n value is 0.035, which would typically be used as a value for representing grass area, which dominates this model domain. In comparison the ATS solver (solid blue line) produces a more linear relationship. Further, the Roe solver (column 1, row 6) also demonstrates a small variation with respect to DEM error, which is noted by a small difference in the position of the two lines. Unlike the previous two figures, other factors are have a similar level of influence, with the gradients of the columns and rows associated with the inflow uncertainty, friction value and cell size are greater than had been seen in Figure 10.

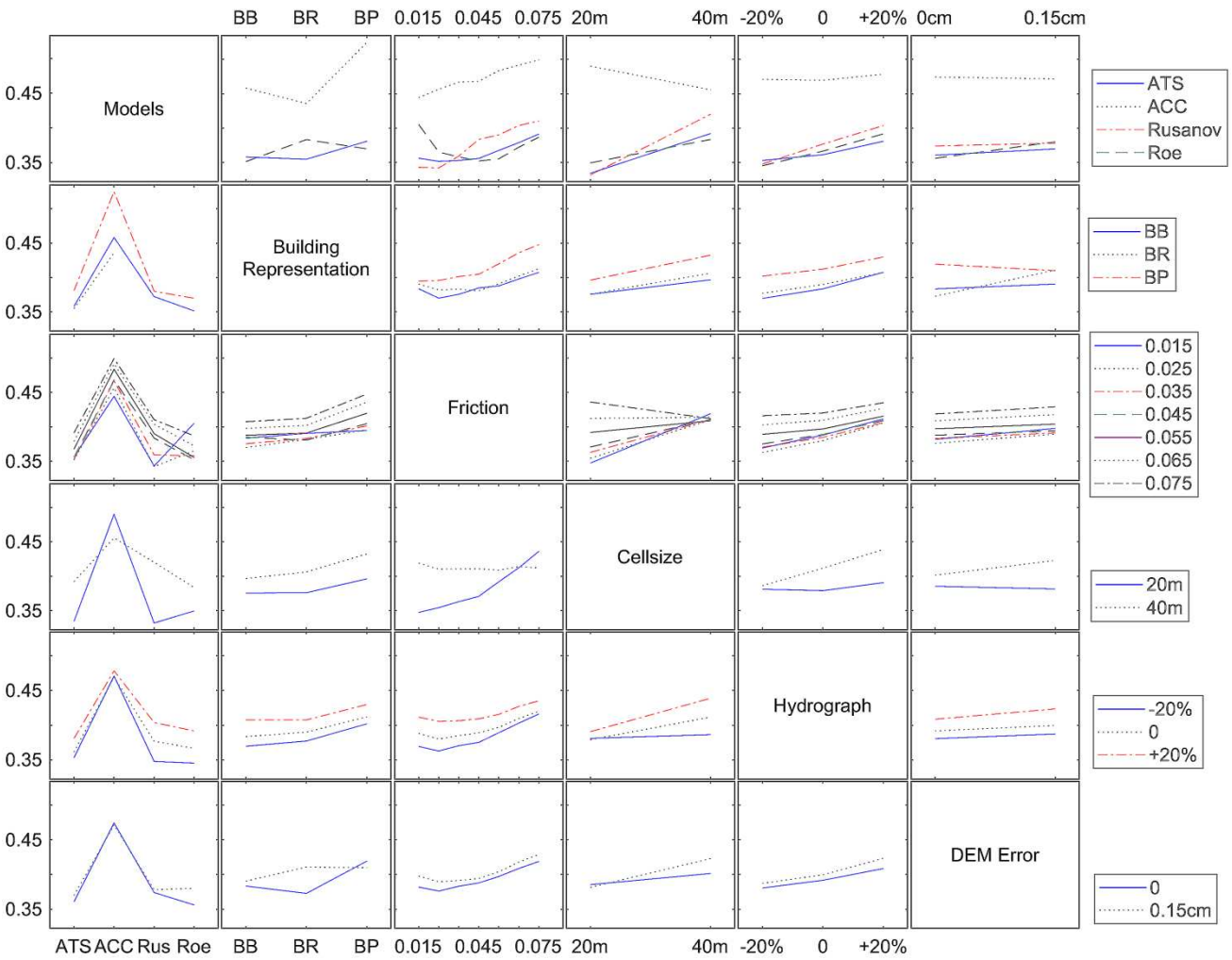


Figure 12: Interaction plot for each input factor comparing mean value of RMSE per level off factor other levels of factors for the Mexborough Test Case

The value of friction choice displays a strong gradient of increasing RMSE with increasing value (Column 3). As with the analysis of $F^{(2)}$ an interaction occurs between increasing friction and increasing cell size, where the higher cell size creates a similar mean level for all friction values (as can be seen in the plot in column 4, row 3). The significance of this is that with increasing cell size the sensitivity of the friction value decreases, and therefore requires less effort to determining an appropriate value. Cell size also has a similar effect on the LISFLOOD modules, and shows a similar pattern to Figure 10 (column 4, row 1). This higher RMSE value is the result of increased water depths on the floodplain. This greater depth with increased cell size suggests that coarsening or removing topographic features alters a controlling factor in floodplain flow mechanisms. This can also be seen when comparing the hydrographic input, which also shows a similar pattern of increasing RMSE with higher inflow boundary conditions (column 5). The significance of the inflow as a primary control on model results is less than the cell size or model type, which can be seen by the range of results across the x-axis of column 5, which is less than either of these factors. There is however, a level of interaction with the level of cell size (column 4, row 5). At the higher grid resolution, the results are similar, but as the grid size increased, there is a divergence in results between the different levels of inflow, which along with the comparison of Figure 10 and Figure 11, show the importance of cell size in model results. The impacts

between these factors are difficult to determine when only considering the difference in $F^{(2)}$, which highlights the advantage of the multiple function approach. Overall, though the relatively limited range of values across all the results for Mexborough, and at a relatively high level for both objective functions indicates that most parameters sets appear to be representative of the observed flood conditions, and despite the uncertainty in the inputs, produces similar model outputs.

4. DISCUSSION

The test cases presented here indicate that the model complexity has a significant influence on most of the model functions and model results. For the Glasgow test case, the level of physical representation controlled most of the variation in model results and was reflected in both evaluation techniques and direct model outputs. The impact appears to be monotonic in this test case when evaluating extent, where levels of model complexity are matched by alterations in the output and evaluation techniques. Variations between the modules in replicating main pathways, observed extent and timings of depths are clear across the evaluations of the flood domain. Whilst the results are potentially biased by the use of the high-resolution benchmark approach, this confirms previous findings about the significance of the level of physical representation to accurately model inundation in urban environments (Hunter et al 2008, Neal et al 2011). Further, by comparing these results to the work of Hunter et al (2008), the significance of the model has been demonstrated when considered as part of the input ensemble. However, the significance of the numerical model appears to be lower in the rural test case of Mexborough, where the main flood mechanisms and flow paths vary gently with time, and sub critical conditions dominate the event. In this test case, whilst variations between the different LISFLOOD-FP modules are smaller, the control of the floodplain topography ensures that all models across the ensemble produce similar results. What is also clear in this test case in comparison to the urban test case, is a relative insensitivity of the LISFLOOD modules to input parameters. This would indicate that not only is the choice of numerical model less significant in affecting the model outputs, but that a certain amount of insensitivity to the parameter inputs exists. This is similar to the conclusions made by Pender and Neelz (2010) which identified that for larger domains with regular flood plain topography, models with varying levels of complexity would produce similar final water depths and extents. By comparing multiple inputs, this research has further illustrated this point.

Although the Mexborough test case had fewer difference between model runs than had been observed in the Glasgow test case, these dissimilarities still provided insight into the importance of physical representation in the model. The variations noted in Figure 4, which are caused by backwater effects and conservation of momentum over complex terrain, affect flood extents along small channels. These variations demonstrate the impact of the modules on macro scale hydraulic problems. These variations have not been significant enough to alter the overall results however. What this demonstrates is that analysis methods that focus on general model comparisons will not fully display all local variations. Higher detail in the analysis methods (such as used by Pappenberger et al 2008) should be used to highlight these variations. At this larger model domain scale however, the influence of other parameters in affecting model results become apparent.

There are some clear differences between the simplified and full physics solutions in both the cases presented here. For the urban test case, the difference in flooded extent and depths between the four modules is clear to see. Whilst the simplified approaches are capable of replicating aspects of the observed depths the range of results, that this is possible only with detailed calibration data (Bates et al., 2010). It can also be seen that the difference cannot be reduced through a calibration process, or be related to uncertainties associated with inflow volume, but through the inability of the simplified approaches to preserve momentum through critical areas of the model domains. This difference is not always clear in the Mexborough test case. By using other methods to evaluate the model, such as the hazard extent likelihood plots, the impacts of using simplified approaches can be clearly seen.

The results also show that the approach to solving the shallow water equations is important, not only in terms of model performance, but also in terms of robustness to uncertain parameters. The reduced complexity in the LISFLOOD-Rusanov solver can produce similar output to LISFLOOD-Roe, and in replicating critical flow paths through the model domain. However, this code produces significantly larger areas of uncertain flood inundation frequency and a wider range of model results from the model ensemble, indicating susceptibility to uncertain inputs. This is notable in both test cases. Examples of this can be seen in the uncertainty flood extents, where the Rusanov solver produces wider flood extents than the Roe solver, but with lower flood frequencies values in these regions. This is further confirmed in the range of model evaluation results. The potential savings in terms of computational cost is therefore offset by the requirement for robust calibration, and as with the simplified solutions to the governing equations, demonstrates that reduced complexity in the numerical methods should be used with caution. Further to this, the sensitivity of the Roe solver to the range of inputs in the Glasgow test case, suggests that the model could be tuned within the range of parameters to represent different hydraulic processes. This is a useful feature of a model and increases the potential application of the code to a range of problems.

Whilst the impact of the model type is clear, a number of trends are noticeable across the test cases which highlight the significance of the other model inputs. The method of building representation is critical in the urban environment test present here. This input in particular influences the accuracy of flow path representation, which can impact both extent and the modelled hazard level in key locations. As with the variation of results caused by the model complexity, the impact is often localised in the model domain with variations in the water depths and extents occurring near the location of buildings. The overall impact is often small, compared to other factors but can impact the level of hazard associated with individual buildings. By comparison, the choice of cell size, appears to have a broader, global effect on model output. The impact of this input is often not directly noticeable, but by examining the interaction with other parameters it can be seen that as grid resolution is increased, model performance converges, including the LISFLOOD-FP modules. This indicates that underlying topography representation plays a significant control on model performance. This is a similar conclusion as noted by Savage et al (2016), which also shows a similar pattern of decreasing model performance and sensitivity with increasing cell size. However, it is noted in that paper that by using local measures of model variations rather than global measures, the cell size also plays an important part in the model

results during the drying phase of an event. Identifying these local variations represents a potential further area of research to consider when using the range of inputs presented in this paper

The inflow boundary conditions are also shown to be of lower significance than model type, building representation and cell size when considering the comparison of extent measures only, but this impact changes when comparing the depth across the floodplain. This is not surprising considering the issues of using $F^{(2)}$, where variations in depth do not create similar variations in extent (Stephens et al 2012), but this furthers findings by Savage et al (2016) that shows the significance of this factor.

By comparison the error associated with the DEM shows little signs of significant contribution to variations in model output. However, the method to assess this factor may be a contributing factor, where the error is randomly assigned between the maximum value and zero. Further information on this may allow the DEM to be degraded in a non-random way, to define spatial patterns between error, surface features and measured elevation. The results here compare with the conclusions drawn by Tsubaki and Kawahara (2013), who demonstrate that in urban areas, and in complex topography the variance related to variations in the DTM surface are significantly less than variations in rural areas. By employing a similar approach to representing the DTM error, and by cross comparing it with other parameters the overall significance of this in impacting model results is shown to be low.

The model evaluation techniques also provide further insight to the significance of each of the parameters. In each test case the use of extent measure $F^{(2)}$, provides a useful comparison point to previous research, and provides an indication of the response of module performance, particularly in urban areas. Methods that go beyond the approach of binary comparison of extent are increasingly popular (Stephens et al 2012) and provide different properties of the observed data with which to test the model outputs. In terms of evaluating uncertainty in modelling, the use of alternative evaluation functions has demonstrated two points. First, that the method of evaluation has a distinct impact on what factors are identified as critical. For example, the use of extent indicates the importance of friction for most test cases, and less critical is the value of inflow. However, this importance changes where depth becomes a key element in the model evaluation technique, such as Nash-Sutcliffe in the Glasgow test case. Secondly, that combined approaches can significantly overcome this issue, and can be used to select fewer better performing models. The combined multi objective function analysis can also refine the number of best performing models down to a smaller selection (as in Dung et al 2011). The use of two different second evaluation techniques reflects the difference in the nature of the function and the event - a RMSE measure would be less representative, for short, rapid events like Glasgow. But future studies could use the same evaluation functions to provide a comprehensive analysis between multiple test cases.

Further to this, using quantitative sensitivity analysis techniques with these multiple evaluation functions will further develop the relationship between factors and model outputs. It should be noted as well that choice of location in evaluating the models with the Nash-Sutcliffe coefficient in Glasgow may also influence the results and conclusions presented. This would not typically be a problem in most modelling studies, but in evaluating models, full consideration must be made to consider what the influence of certain approaches will be on results (Hunter et al., 2006).

Ultimately, to further understand the modelling results presented here, detailed sensitivity analysis that compares global differences in variance will be required (Hall et al., 2005). Whilst the approach here provides qualitative conclusions, there is a need to quantify the difference between levels of physical representation in an uncertainty context. By undertaking a Global Sensitivity Analysis (GSA) approach with the inputs used here, the variations in the model results can be further explored, which will further improve model development and understanding. The advantage of this is shown in the work by Savage et al (2016), but also some of the potential computational issues are also shown in that study. Here the emphasis has been on including as wide a range of potential inputs as possible, at the cost of quantitative analysis.

5. CONCLUSIONS

This paper has investigated the major sources of uncertainty in flood modelling using a systematic approach to structure the distribution of key modelling inputs. Over the two test cases, it appears the level of model complexity has the greatest control on model output. The significance of this reduces though, for the Mexborough test case, where both the simplified physical approaches and full physics models are able to replicate the observed data. This test case is characterised by gradual varying flood processes. This confirms, to an extent, the theoretical limits of the physical model, and reaffirms that for suitable cases, simplified models are still valid approaches to flood risk evaluation. In this test case, though, the majority of models from the ensemble are able to replicate the flood processes to a reasonable level. Whilst the other inputs are less significant, there is still evidence of parameter interaction and control on model outputs increasing cell size, leads to convergence of results between the LISFLOOD modules. This highlights that uncertainty is a multidimensional problem, and higher order interactions are significant in model results. The overall model results also demonstrate that the use of friction alone to vary outputs results can be overestimated particularly where spatially varied values are used.

The use of multiple evaluations techniques across two test case has also helped to highlight responses and interactions between the inputs. These approaches help overcome some of the issue of bias that occur when uncertainty analysis is undertaken with single test case and model evaluation technique. Further to this where datasets to undertake model evaluation is either uncertain or not extensive, the use of multiple or combined evaluations can still provide insight into modelling uncertainties. Future uncertainty assessments should use multi data sets for assessment, where the data exists to allow this to occur.

This research highlights a key aspects of flood inundation modelling, in that model selection and parameter choice will be problem and data specific. To suggest that one level of physical representation of flow is the most appropriate does not hold true in all scenarios, and each modelling study should be validated in turn rather than rely on broad rules. As the collective understanding of uncertainty improves, it may require an increase in the detail of the modelling, such as drainage models, improved building representation in models and more detail in the parameters than single global friction values, in order to overcome current levels of uncertainty.

Acknowledgements

Thomas Willis was funded by the Flood Risk Management Research Consortium which was supported by grant number EP/F20511/1 from the EPSRC and the DEFRA/EA Joint Research Programme on Flood and Coastal Defence. The authors wish to thank the anonymous referees and Dr Mark Smith who provided comments and feedbacks that helped improve the quality of the manuscript. The authors also wish to thank for valuable insight Professor Paul Bates and Dr Jeffery Neal for use of and insights into the LISFLOOD-FP code, and Professor Rob Lamb for use and insight into the Mexborough dataset.

References

APEL, H., ARONICA, G. T., KREIBICH, H. & THIEKEN, A. H. 2009. Flood risk analyses-how detailed do we need to be? *Natural Hazards,* 49**,** 79-98.

ARONICA, G., BATES, P. D. & HORRITT, M. S. 2002. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes,* 16**,** 2001-2016.

ARONICA, G., HANKIN, B. & BEVEN, K. 1998. *Topographic sensitivity and parameter uncertainty in the predictions of a 2D inundation model,* Leiden, A a Balkema Publishers.

ARONICA, G. T., FRANZA, F., BATES, P. D. & NEAL, J. C. 2012. Probabilistic evaluation of flood hazard in urban areas using Monte Carlo simulation. *Hydrological Processes,* 26**,** 3962-3972.

BATES, P. D. & DE ROO, A. P. J. 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology,* 236**,** 54-77.

BATES, P. D., HORRITT, M. S. & FEWTRELL, T. J. 2010. A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *Journal of Hydrology,* 387**,** 33-45.

BEVEN, K. J. 2008. *Environmental Modelling: An Uncertian Future?*, Taylor & Francis.

BOSS, C. C.-L., GABRIELA; CATON ROB; CLOUSTON, DAVID; ETHERIDGE, STEPHEN; FOOTE, MATTHEW; GARCIA, GIOVANNI; LATCHMAN, SHANE; MAYNARD, TREVOR; MILLER, PAUL; MITCHELL, ANDREW; PAINTER, MICHAEL; REYNOLDS, IAIN; SCHMID, LARS; SIMIC, MILAN; SOUCH, CLAIRE; STEIMEN, SIBYLEE; THRAINSSON, HJORTUR; TSCHUDI 2011. Industry Good Practice for Catastrophe Modelling. *Association of British Insurers.*

CHOW, V. E., MAIDMENT, D.R., MAYS, L.M. 1988. *Applied Hydrology*, McGraw Hill Book Company.

COBBY, D. M., MASON, D. C. & DAVENPORT, I. J. 2001. Image processing of airborne scanning laser altimetry data for improved river flood modelling. *Isprs Journal of Photogrammetry and Remote Sensing,* 56**,** 121-138.

DE ALMEIDA, G. A. M., BATES, P., FREER, J. E. & SOUVIGNET, M. 2012. Improving the stability of a simple formulation of the shallow water equations for 2-D flood modeling. *Water Resources Research,* 48.

DOMENEGHETTI, A., CASTELLARIN, A. & BRATH, A. 2012. Assessing rating-curve uncertainty and its effects on hydraulic model calibration. *Hydrology and Earth System Sciences,* 16**,** 1191-1202.

DUNG, N. V., MERZ, B., BARDOSSY, A., THANG, T. D. & APEL, H. 2011. Multi-objective automatic calibration of hydrodynamic models utilizing inundation maps and gauge data. *Hydrology and Earth System Sciences,* 15**,** 1339-1354.

FEWTRELL, T. J., BATES, P. D., HORRITT, M. & HUNTER, N. M. 2008. Evaluating the effect of scale in flood inundation modelling in urban environments. *Hydrological Processes,* 22**,** 5107-5118.

FEWTRELL, T. J., DUNCAN, A., SAMPSON, C. C., NEAL, J. C. & BATES, P. D. 2011. Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data. *Physics and Chemistry of the Earth,* 36**,** 281-291.

GUINOT, V. 2012. Multiple porosity shallow water models for macroscopic modelling of urban floods. *Advances in Water Resources,* 37**,** 40-72.

GUPTA, H. V., CLARK, M. P., VRUGT, J. A., ABRAMOWITZ, G. & YE, M. 2012. Towards a comprehensive assessment of model structural adequacy. *Water Resources Research,* 48.

HALL, J. W., BOYCE, S. A., WANG, Y. L., DAWSON, R. J., TARANTOLA, S. & SALTELLI, A. 2009. Sensitivity Analysis for Hydraulic Models. *Journal of Hydraulic Engineering-Asce,* 135**,** 959-969.

HALL, J. W., TARANTOLA, S., BATES, P. D. & HORRITT, M. S. 2005. Distributed sensitivity analysis of flood inundation model calibration. *Journal of Hydraulic Engineering-Asce,* 131**,** 117-126.

HUNTER, N. M., BATES, P. D., HORRITT, M. S., DE ROO, P. J. & WERNER, M. G. F. 2005a. Utility of different data types for calibrating flood inundation models within a GLUE framework. *Hydrology and Earth System Sciences,* 9**,** 412-430.

HUNTER, N. M., BATES, P. D., HORRITT, M. S. & WILSON, M. D. 2006. Improved simulation of flood flows using storage cell models. *Proceedings of the Institution of Civil Engineers-Water Management,* 159**,** 9-18.

HUNTER, N. M., BATES, P. D., HORRITT, M. S. & WILSON, M. D. 2007. Simple spatially-distributed models for predicting flood inundation: A review. *Geomorphology,* 90**,** 208-225.

HUNTER, N. M., BATES, P. D., NEELZ, S., PENDER, G., VILLANUEVA, I., WRIGHT, N. G., LIANG, D., FALCONER, R. A., LIN, B., WALLER, S., CROSSLEY, A. J. & MASON, D. C. 2008. Benchmarking 2D hydraulic models for urban flooding. *Proceedings of the Institution of Civil Engineers-Water Management,* 161**,** 13-30.

HUNTER, N. M., HORRITT, M. S., BATES, P. D., WILSON, M. D. & WERNER, M. G. F. 2005b. An adaptive time step solution for raster-based storage cell modelling of floodplain inundation. *Advances in Water Resources,* 28**,** 975-991.

LIU, Z., MERWADE, V. & JAFARZADEGAN, K. 2019. Investigating the role of model structure and surface roughness in generating flood inundation extents using one- and two-dimensional hydraulic models. *Journal of Flood Risk Management,* 12**,** 19.

MORGAN, A., OLIVIER, D., NATHALIE, B., CLAIRE-MARIE, D. & PHILIPPE, G. 2016. High-Resolution Modelling With Bi-Dimensional Shallow Water Equations Based Codes - High-Resolution Topographic Data Use For Flood Hazard Assessment Over Urban And Industrial Environments. *In:* KIM, J. H., KIM, H. S., YOO, D. G., JUNG, D. & SONG, C. G. (eds.) *12th International Conference on Hydroinformatics.* Amsterdam: Elsevier Science Bv.

NEAL, J., VILLANUEVA, I., WRIGHT, N., WILLIS, T., FEWTRELL, T. & BATES, P. 2012. How much physical complexity is needed to model flood inundation? *Hydrological Processes,* 26**,** 2264-2282.

NEELZ, S. & PENDER, G. 2010. Benchmarking of 2D hydraulic modelling packages. *Science Report SC080035/SR2. Bristol Environment Agency*.

NEELZ, S. & PENDER, G. 2013. Benchmarking the latest generation of 2D hydraulic modelling packages. *Science Report SC120002/SR. Bristol Environment Agency*.

PAPPENBERGER, F., BEVEN, K., 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**,** 46-69.

PAPPENBERGER, F., BEVEN, K. J., RATTO, M. & MATGEN, P. 2008. Multi-method global sensitivity analysis of flood inundation models. *Advances in Water Resources,* 31**,** 1-14.

PAPPENBERGER, F., MATGEN, P., BEVEN, K. J., HENRY, J. B., PFISTER, L. & FRAIPONT DE, P. 2006. Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Advances in Water Resources,* 29**,** 1430-1449.

PRESTININZI, P., DI BALDASSARRE, G., SCHUMANN, G. & BATES, P. D. 2011. Selecting the appropriate hydraulic model structure using low-resolution satellite imagery. *Advances in Water Resources,* 34**,** 38-46.

ROE, P. L. 1981. APPROXIMATE RIEMANN SOLVERS, PARAMETER VECTORS, AND DIFFERENCE-SCHEMES. *Journal of Computational Physics,* 43**,** 357-372.

SANDERS, B. F., SCHUBERT, J. E. & GALLEGOS, H. A. 2008. Integral formulation of shallow-water equations with anisotropic porosity for urban flood modeling. *Journal of Hydrology,* 362**,** 19-38.

SAVAGE, J. T. S., PIANOSI, F., BATES, P., FREER, J. & WAGENER, T. 2016. Quantifying the importance of spatial resolution and other factors through global sensitivity analysis of a flood inundation model. *Water Resources Research,* 52**,** 9146-9163.

SCHUBERT, J. E. & SANDERS, B. F. 2012. Building treatments for urban flood inundation models and implications for predictive skill and modeling efficiency. *Advances in Water Resources,* 41**,** 49-64.

SCHUBERT, J. E., SANDERS, B. F., SMITH, M. J. & WRIGHT, N. G. 2008. Unstructured mesh generation and landcover-based resistance for hydrodynamic modeling of urban flooding. *Advances in Water Resources,* 31**,** 1603-1621.

SIMOES, F. J. M. 2011. Finite Volume Model for Two-Dimensional Shallow Environmental Flow. *Journal of Hydraulic Engineering-Asce,* 137**,** 173-182.

SOARES-FRAZAO, S., LHOMME, J., GUINOT, V. & ZECH, Y. 2008. Two-dimensional shallow-water model with porosity for urban flood modelling. *Journal of Hydraulic Research,* 46**,** 45-64.

STEPHENS, E. M., BATES, P. D., FREER, J. E. & MASON, D. C. 2012. The impact of uncertainty in satellite data on the assessment of flood inundation models. *Journal of Hydrology,* 414**,** 162-173.

TORO, E. F. 2001. *Shock Capturing Methods for Free-Surface Shallow Flows*, Wiley and Sons Ltd.

TSUBAKI, R. & KAWAHARA, Y. 2013. The uncertainty of local flow parameters during inundation flow over complex topographies with elevation errors. *Journal of Hydrology,* 486**,** 71-87.

VAN STEENBERGEN, N., RONSYN, J. & WILLEMS, P. 2012. A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software,* 33**,** 92-105.

VILLANUEVA, I. & WRIGHT, N. G. 2006. Linking Riemann and storage cell models for flood prediction. *Proceedings of the Institution of Civil Engineers-Water Management,* 159**,** 27-33.

WARMINK, J. J., VAN DER KLIS, H., BOOIJ, M. J. & HULSCHER, S. 2011. Identification and Quantification of Uncertainties in a Hydrodynamic River Model Using Expert Opinions. *Water Resources Management,* 25**,** 601-622.

WILDEMEERSCH, S., GODERNIAUX, P., ORBAN, P., BROUYERE, S. & DASSARGUES, A. 2014. Assessing the effects of spatial discretization on large-scale flow model performance and prediction uncertainty. *Journal of Hydrology,* 510**,** 10-25.

WILLEMS, P. 2012. Model uncertainty analysis by variance decomposition. *Physics and Chemistry of the Earth,* 42-44**,** 21-30.