Benchmarking 2D hydraulic models for urban flooding

Proceedings of the Institution of Civil Engineers
Water Management 161
February 2008 Issue WM1
Pages 13–30
doi: 10.1680/wama.2008.161.1.13
Paper 700003
Received 22/01/07
Accepted 28/09/07
Keywords:
floods & floodworks/hydraulics & hydrodynamics/mathematical modelling

Neil M. Hunter
Postdoctoral Researcher, University of Bristol, UK; now JBA Consulting, Skipton, UK

Paul D. Bates
Professor, School of Geographical Sciences, University of Bristol, UK

Sylvain Neelz
Postdoctoral Researcher, School of the Built Environment, Heriot-Watt University, Edinburgh, UK

Gareth Pender
Professor, School of the Built Environment, Heriot-Watt University, Edinburgh, UK

Ignacio Villanueva
Postdoctoral Researcher, School of Civil Engineering, University of Nottingham, UK

Nigel G. Wright
Professor, University of Nottingham, UK; now Professor, UNESCO IHE Foundation for Water Education, Delft, the Netherlands

Dongfang Liang
Lecturer, Department of Engineering, University of Cambridge, UK

Roger A. Falconer
Halcrow Professor, School of Engineering, Cardiff University, UK

Binliang Lin
Reader, School of Engineering, Cardiff University, UK

Simon Waller
Director, JBA Consulting, Skipton, UK

Amanda J. Crossley
Analyst, JBA Consulting, Skipton, UK

David C. Mason
Reader, NERC Environmental System Science Centre, University of Reading, UK

Benchmarking 2D hydraulic models for urban flooding

N. M. Hunter PhD, P. D. Bates PhD, S. Neelz PhD, G. Pender PhD, FICE, I. Villanueva PhD, N. G. Wright PhD, D. Liang PhD, R. A. Falconer DSc, FREng, FICE, B. Lin PhD, S. Waller CEng, MICE, A. J. Crossley PhD and D. C. Mason PhD

This paper describes benchmark testing of six two-dimensional (2D) hydraulic models (DIVAST, DIVAST-TVD, TUFLOW, JFLOW, TRENT and LISFLOOD-FP) in terms of their ability to simulate surface flows in a densely urbanised area. The models are applied to a 1·0 km × 0·4 km urban catchment within the city of Glasgow, Scotland, UK, and are used to simulate a flood event that occurred at this site on 30 July 2002. An identical numerical grid describing the underlying topography is constructed for each model, using a combination of airborne laser altimetry (LiDAR) fused with digital map data, and used to run a benchmark simulation. Two numerical experiments were then conducted to test the response of each model to topographic error and uncertainty over friction parameterisation. While all the models tested produce plausible results, subtle differences between particular groups of codes give considerable insight into both the practice and science of urban hydraulic modelling. In particular, the results show that the terrain data available from modern LiDAR systems are sufficiently accurate and resolved for simulating urban flows, but such data need to be fused with digital map data of building topology and land use to gain maximum benefit from the information contained therein. When such terrain data are available, uncertainty in friction parameters becomes a more dominant factor than topographic error for typical problems. The simulations also show that flows in urban environments are characterised by numerous transitions to supercritical flow and numerical shocks. However, the effects of these are localised and they do not appear to affect overall wave propagation. In contrast, inertia terms are shown to be important in this particular case, but the specific characteristics of the test site may mean that this does not hold more generally.

1. INTRODUCTION

Application of two-dimensional (2D) hydraulic models to rural floodplains is now relatively well understood as a result of numerous model applications over the last two decades (see for example, Refs 1–6). Benchmarking of 2D models applied to rural floodplains (e.g. Bates and De Roo[7] and Horritt and Bates[8]) has also yielded insights into the level of physical and topographic representation required to simulate flow characteristics at different scales, the impact of different numerical solution schemes on the results obtained and the physical realism of model parameters in different model types. A similar process has yet to be undertaken for urban floodplains as applications have, until very recently, been prevented by a lack of high-resolution data to characterise complex urban topography and topology, and insufficient model efficiency to attempt urban flooding simulations on meaningful scales. Resolving surface water movement through urban areas requires resolution of complex flow paths around buildings, representation of micro-scale topographic and blockage effects (e.g. kerbs and walls[9]) and numerical schemes capable of dealing with high-velocity flow at shallow depth. This requires model grids of the order of 1–5 m resolution to capture the relevant topographic features.[10] It is thus no surprise that hydraulic modellers have concentrated on the simpler rural case. However, in order to undertake comprehensive flood risk management, there is a need to develop a capability to model urban areas as this is where the majority of at-risk assets are located.

Over the past decade a number of studies have documented the application of 2D hydraulic models to complex urban problems, including numerical solutions of the full 2D shallow-water equations,[11–14] 2D diffusion wave models,[15–17] analytical approximations to the 2D diffusion wave equations using uniform flow formulae[18] and grid-based geomorphological routing models.[19] Moreover, the work presented in this paper builds on a rich heritage of 1D and GIS-based hydraulic modelling of urban areas (for a review see Smith[20]), often involving linked surface water–sewer flow models.

Despite this proliferation of 2D hydraulic models applied to urban areas, understanding of the relative merits of different approaches is still limited (the study of Leopardi *et al.*[21] is an exception). Single model applications to sparse validation data confirm that such applications are indeed possible and identify the methodological constraints that need to be resolved to make such simulations happen, but do not tell us anything about the trade-offs between physical realism and code efficiency or the relative merits of various numerical solution techniques. Moreover, to make such comparisons meaningful requires that codes are considered within a framework that accounts for realistic uncertainties over topographic error and the fact that almost all hydraulic models require calibration, which may compensate for

data and model errors. This is the primary focus of this paper. The performance of six 2D hydraulic models (TUFLOW, DIVAST, DIVAST-TVD, TRENT, JFLOW and LISFLOOD-FP) has been compared, at high resolution, to simulate surface flooding of a small urban catchment in Glasgow, Scotland, UK. The codes represent a wide spectrum of fundamental equations, numerical techniques and time step control methods. The codes are first compared in terms of their relative performance and response to typical topographic error for a single set of roughness parameters for this test site. A sensitivity analysis was then undertaken to determine whether differences between the models can be compensated for by calibration using physically realistic parameter values. In this way the aim is to determine whether uncertainty is reduced most effectively by refining the topographic representation, improving the model physical basis or by collecting detailed hydrometric data to better constrain the model calibration process. A further goal of this work is to determine whether the performance differences caused by the use of different numerical solution techniques are more or less important than uncertainties introduced by the model calibration process.

2. MODELLING APPROACHES

The models selected for use in this paper were chosen to represent most of the main classes of 2D codes that could be applied to urban hydraulic modelling problems. The models are

(*a*) implicit finite-difference solutions of the full 2D shallow-water equations (TUFLOW[22] and the original DIVAST code,[23] both of which use an alternating direction implicit (ADI) solver)
(*b*) explicit finite-difference solutions of the full 2D shallow-water equations (DIVAST-TVD,[24,25] which uses a total variation diminishing (TVD) solver)
(*c*) explicit finite-volume solutions of the full 2D shallow-water equations (TRENT,[26] which uses a Roe Riemann solver)
(*d*) explicit finite-difference solutions of the 2D diffusion wave equations (JFLOW[27])
(*e*) explicit analytical approximations to the 2D diffusion wave equations (LISFLOOD-FP[28]).

The codes are fully described in the references cited; Table 1 briefly summarises the characteristics pertinent to the present discussion. The benchmarking process begins with structured grid models only, as this allows easily control of grid resolution effects in the numerical experiments that follow. A consequence of this is that unstructured finite-element and finite-volume methods have been omitted from the comparison at this stage.

Essentially, each code represents a different trade-off between physical representation and potential computational cost based on the developers' assumptions about those flow features that become critical in particular situations. For example, some

Model	Equations	Turbulence closure	Numerical solution	Shock capturing?	Time step control (for explicit schemes)	Mass balance error for benchmark test case (volume error as percentage of inflow volume)
TUFLOW[22]	Full 2D shallow-water equations	Constant eddy viscosity (scaling coefficient $k_0 = 0{\cdot}2$)	Modified ADI implicit finite-difference scheme[22]	No	Constant time step	0·5
DIVAST[23]	Full 2D shallow-water equations	Dynamic mixing length model	ADI implicit finite-difference scheme	No	Unconditionally stable, constant time step (0·1 s)	0·16
DIVAST-TVD[24,25]	Full 2D shallow-water equations	No	TVD-MacCormack explicit finite-difference scheme[24,25]	Yes	Typically uses a conditionally stable, adaptive time step based on Courant number; however, a fixed time step of 0·1 s was used here to allow direct comparison with DIVAST	0·2
TRENT[26]	Full 2D shallow-water equations	No	First-order explicit Roe Riemann solver[35]	Yes	Conditionally stable adaptive time step based on Courant–Friedrichs–Levy (CFL) condition	1·13
JFLOW[27]	2D diffusion wave	No	Explicit finite-difference	No, shocks not represented by controlling equations	Conditionally stable adaptive time step based on CFL condition and flow reversal constraint	$5{\cdot}8 \times 10^{-9}$
LISFLOOD-FP[28]	Analytical approximation to 2D diffusion wave using uniform flow formulae decoupled in x and y directions	No	None	No, shocks not represented by controlling equations	Unconditionally stable adaptive time step based on CFL condition and von Neumann stability analysis for a diffusion system	$0{\cdot}12 \times 10^{-3}$

Table 1. Main characteristics of the six codes used in the benchmark test

(TRENT and DIVAST-TVD) use more complex numerical solution schemes to allow shock capturing, while others (LISFLOOD-FP, JFLOW) sidestep this problem by solving equations that cannot represent such hydraulic transients (and indeed actively smooth them out). Some are implicit (TUFLOW, DIVAST), while others are explicit (JFLOW, LISFLOOD-FP, TRENT and DIVAST-TVD) and therefore require careful time step control to maintain stability. Given that very little is known about the likely importance of specific flow mechanisms in urban settings or model efficiency in simulating these, all the above codes are potentially adequate for the problem in hand. In addition, by building replication into the research design of the numerical experiments, it should be possible to trace consistent behaviours exhibited by particular groups of codes to their specific common features.

3. STUDY SITE AND DATA AVAILABILITY

The models were applied to a small urban catchment within the city of Glasgow, Scotland, UK, which has flooded in the recent past. The domain to be modelled comprises a rectangle 1·0 km × 0·4 km, with dense urban development either side of two main streets and a topologically complex network of minor roads. The area is a mix of some steep sections of road and local depressions where water may pond. To characterise the topography and topology for this study site, a 1 m resolution airborne laser altimetry (LiDAR) survey was undertaken by Infoterra Ltd. This was then fused with Ordnance Survey (OS) Mastermap® digital map data that defined building locations, the road network and land-use type as vector layers. The LiDAR data acquired for this study had already been filtered by Infoterra to remove vegetation and building features, using their standard processing algorithms, to leave a 'bare earth' digital elevation model (DEM) with horizontal and vertical accuracies less than 50 cm and 15 cm root mean square error (RMSE) respectively. For hydraulic modelling the 'bare earth' LiDAR data were aggregated to 2 m and buildings, kerbs and roads were all reinserted, based on their locations in the digital map layer. To reinsert the buildings, all cells in the 2 m DEM that fell inside a building vector defined by the OS Mastermap® data were raised in elevation by either 12 m (for apartment blocks) or 6 m (for small houses) to represent building height. The precise height value is unimportant here as flow depths during all simulations were always <1 m and the purpose of this processing step is to allow buildings to be represented as 'islands' that water must flow around. Roads and kerbs were inserted in a similar way, assuming a uniform kerb height of 10 cm and with the road camber based on analysis of the 1 m LiDAR and observations at the study site. The result is a high-resolution DEM (hereafter termed the benchmark DEM) with realistic representation of urban morphologic features. While the benchmark DEM does not represent actual 'ground truth', it does, unlike raw LiDAR data, represent urban features as smoothly varying surfaces and can be degraded in a controlled way to study the impact of terrain data errors on model predictions. Fig.1 shows the road and building layout at this study site with the surface height (z) from the benchmark DEM shown as a grey scale overlain by water depths (h) from the end of a typical model simulation to give an indication of the flow complexity and areas of ponding.

Flooding at this site is caused, at least in part, by a small (∼1 m wide) stream that enters near the north-east corner of the domain and almost immediately enters a culvert (located at point X0 in Fig. 1) that runs under the entire site. Flooding has been observed to occur here as a result of flow exceeding the capacity of the culvert and spilling into the street network at point X0. The catchment area upstream of X0 is small (<5 km^2) and the stream responds very rapidly to heavy rainfall, with typical flooding events being less than 1 h duration. Once the capacity of the culvert is exceeded, the model simulation shown in Fig. 1 clearly suggests that flow diverges, with water moving along the two main east–west oriented streets before converging and ponding in low-lying areas in the southern part of the domain at location X3. This result is fully consistent with eyewitness accounts and contemporaneous photographs of flooding at this site. Sewer overflow flooding elsewhere in the domain may also contribute to

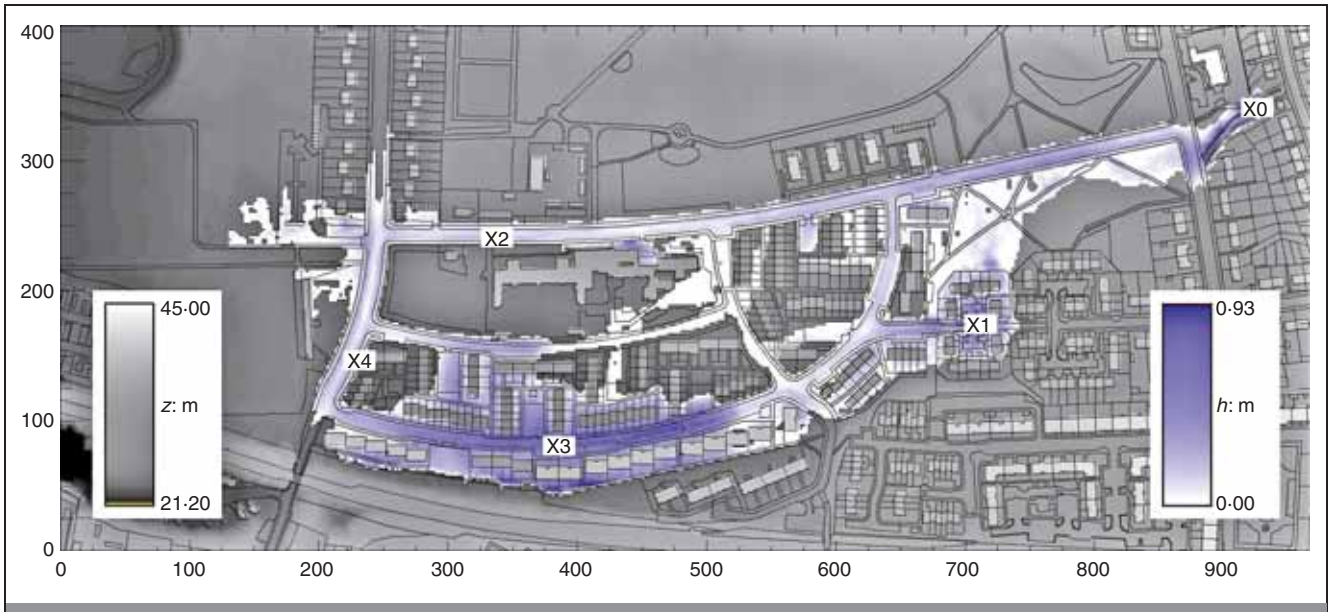


Fig. 1. Map of the study site showing building and road topology derived from Ordnance Survey Mastermap® data with the surface height (z) from the benchmark DEM shown as a grey scale overlain by water depths (h) from a typical model simulation. All map plots are in Cartesian coordinates where east–west is oriented along the x axis and north–south along the y axis. Dimensions in m

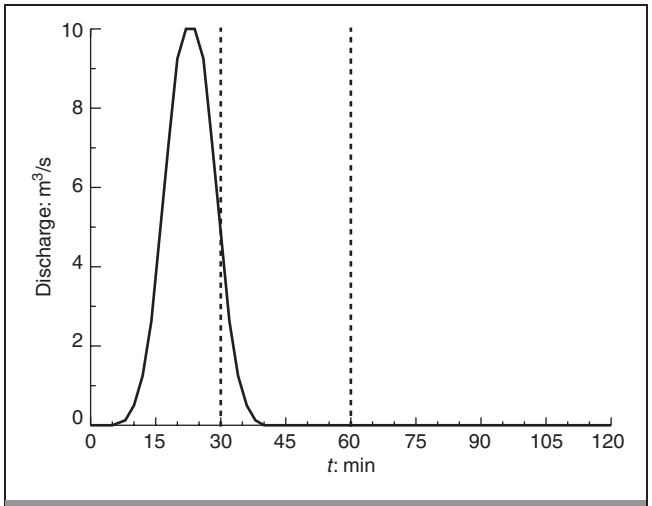


Fig. 2. Event hydrograph simulated in the benchmark tests. The vertical dashed lines at 30 and 60 min represent instances for which model results are presented in later figures

the overall flooding problem, but this is not the focus of this paper where emphasis concentrates on surface water flooding caused by the surcharging culvert only. The domain complexity and requirement to simulate both high-velocity low-depth flow on steep slopes, of low roughness and depression ponding is therefore a rigorous test of hydraulic model ability and is typical of many urban flooding situations.

The flow event simulated in this analysis is based on a real flood that occurred at this site on 30 July 2002. The inflow boundary condition for each model consists of the hydrograph shown in Fig. 2, which is imposed as a point source internal to the model domain at location X0. This hydrograph represents the water volume overflowing the culvert and reflects the authors' best interpretation of available eyewitness accounts and historical photography. Key features typical of many urban floods are the very rapid hydrograph rise and fall, and the ability for small catchments to generate relatively high peak flow rates as a result of the high percentage of impervious surfaces. The flood event lasts <60 min, but each simulation was continued for 120 min to allow water to come to rest and pond in depressions. This then allows a sensible value for mass conservation to be calculated for each code. After 120 min the hydraulic part of the event is effectively over and water levels have ceased to change significantly, although a considerable volume of water remains in the model domain. Observations at this site show that this ponded water takes some considerable time to drain from the catchment through the stormwater drainage system. All external boundaries for each model were closed with zero mass flux. This is a reasonable assumption as, in this application, any flow that does manage to reach the boundary is only 1–2 cm deep and flowing as a shallow sheet. Mass flux across the external boundary is therefore negligible and can be safely ignored for the sake of better consistency between models.

4. METHODOLOGY

Each model was initially set up to use the 2 m resolution benchmark DEM and the point source hydrograph shown in Fig. 2. Given the variety of spatial discretisations used, a first problem was to ensure that assimilation of the benchmark DEM into each model led to an identical representation of the site terrain. This is not straightforward as, even though the analysis is limited to structured grids, the selected models define the terrain at different locations (centre, side mid-points, etc.) on each grid cell. The task is also complicated by the fact that a number of models use intermediate data formats prior to assimilating the DEM data or have different spatial interpolation schemes. A correct solution to this problem was eventually found, which was tested by checking that each model had an identical surface level at a number of points in the domain. Identical spatially distributed friction coefficients were specified for each model that discriminated between two land-use classes—n_{veg} for all vegetated areas and n_{roads} for tarmac areas—determined from the OS Mastermap® data. Lastly, the time step was specified individually for each model. Table 1 shows the wide variety of approaches to time step control taken by the models used in this benchmark testing to ensure stability, computational efficiency or an oscillation-free solution (or trade-off between these) for particular numerical solution schemes. Given that the purpose of this research was to produce a rigorous benchmark test, it would have unfairly advantaged particular codes to have been more proscriptive about the time step to be used.

As noted in section 1, the benchmark analysis consisted of two parts. In part 1, the codes were compared in terms of their relative performance and response to typical topographic error for a single set of roughness parameters for this test site. In part 2, a sensitivity analysis study was undertaken to determine whether these differences between models can be compensated for by calibration, using physically realistic parameter values.

For the part 1 analysis, a second DEM (hereafter termed the LiDAR DEM) was generated by degrading the benchmark DEM through the addition of horizontal and vertical errors typical of airborne LiDAR data. Total LiDAR positioning errors (horizontal error of $\sim$50 cm RMSE, vertical error of $\sim$15 cm RMSE) divide into systematic and random components, with the former typically being dominant. Of these we here assume that the systematic error would be the same between different DEMs and can also be compensated for absolutely in any modelling study with a uniform offset derived from a contemporaneous ground truth campaign. We therefore degraded the benchmark DEM using typical values for the LiDAR random error component only. Based on values reported in the literature,[29,30] these were estimated as 5 cm RMSE for the vertical error and 15 cm for the horizontal error.

For the purpose of this analysis it is assumed that the benchmark DEM is an error-free representation of the site terrain and this is used as a control to determine the response of each model to typical errors in topographic data. To achieve this, the event shown in Fig. 2 was simulated for models constructed with the benchmark and LiDAR DEMs using a single set of friction coefficients ($n_{veg} = 0{\cdot}05$ and $n_{roads} = 0{\cdot}015$). The total mass balance error over the whole simulation is reported for each model in Table 1 and was found to be a maximum of 1·13% of the inflow hydrograph volume. Even if this mass error were to occur instantaneously in a single time step, when spread over the whole flooded area at peak inundation this is equivalent to a height error of only 3·5 mm. This is insufficient to generate significant differences between models for the purposes of this comparison.

The part 1 analysis also showed that models based on the 2D diffusion wave incurred a higher computational overhead than those based on the full shallow-water equations for the 2 m resolution grid used. All the shallow-water model simulations took approximately 1 h to complete (given differences between compilers and computer architectures used for the tests), while the diffusion wave codes took several times longer. This is perhaps surprising as the diffusion wave models solve simpler flow equations or even, in the case of LISFLOOD-FP, do not use a numerical approach at all. This reverses the typical trend found for larger grid sizes where the computational advantages of using explicit diffusion wave codes are significant. This is explained by the analysis of Hunter *et al.*[28] who show that the optimal time step to maintain stability in an explicit diffusion wave hydraulic model is a quadratic function of grid size. Hence for the 2 m grid used here, very small time steps are required to ensure stability with JFLOW and LISFLOOD-FP and this increases the computational burden. This increase in runtime does not occur for the explicit TRENT and DIVAST-TVD codes, suggesting that computational cost can be significantly reduced for grids of this scale by including inertial terms in the controlling equations. Without these terms, inter-cell fluxes on the floodplain can become unrealistically large and lead to 'checkerboard'-like oscillations unless careful time step control is employed. This is a particular problem in areas of deep slow-moving flow with a low free surface gradient.[28] Here we might expect the inclusion of inertial effects to significantly reduce flow velocities and inter-cell fluxes, thus allowing higher time steps to be employed. This explains the increase in computational cost for the explicit diffusion wave models. Research is now ongoing to address this issue with diffusion wave models applied to very fine grids; this will be reported in a future paper. Despite this increased cost, an advantage of the formulations used in JFLOW and LISFLOOD-FP is that they are inherently mass conservative. This is reflected in the fact that mass balance errors for these codes are orders of magnitude lower than for the more complex numerical schemes used in the full shallow-water models. While we argue above that these larger mass balance errors are unimportant for this application, this may not always be the case.

For the part 2 analysis, an ensemble of 13 simulations was run for each model on the LiDAR DEM using different pairs of roughness coefficients chosen from a wide but physically plausible range to mimic the effect of typical calibration procedures. Parameters n_{veg} and n_{roads} were simultaneously varied: n_{veg} in the range 0·015 (bare earth) to 0·075 (dense tall grass and shrubs) in steps of 0·005 and n_{roads} in the range 0·008–0·02 in steps of 0·001. The pairs of roughness coefficients used and their associated simulation numbers are shown in Table 2. While this is not a large number of simulations compared to previous Monte Carlo analysis of hydraulic model sensitivity to friction specification (see, for example, Aronica *et al.*[31]) and does not account for any parameter independence or non-linearity, it should be sufficient to determine the range and distribution of model response. In turn this will allow determination of whether calibration can compensate for any differences between models with different physics or numerical solutions.

Simulation No.	n_{veg}	n_{roads}
1	0·015	0·008
2	0·020	0·009
3	0·025	0·010
4	0·030	0·011
5	0·035	0·012
6	0·040	0·013
7	0·045	0·014
8	0·050	0·015
9	0·055	0·016
10	0·060	0·017
11	0·065	0·018
12	0·070	0·019
13	0·075	0·020

Table 2. Friction coefficient values used for the simulations in the part 2 analysis

5. RESULTS AND DISCUSSION

For the part 1 analysis, the time series of water heights predicted by each model for each DEM at the four points denoted X1, X2, X3 and X4 in Fig. 1 were compared. These data are shown in Fig. 3. Points X1–X4 were chosen from a much larger set to represent the spread of typical hydraulic conditions occurring in the domain and to simplify presentation of the results. Hence, point X1 represents an area where water rapidly accumulates at the start of the simulation and then slowly releases as the simulation proceeds. Point X2 represents a zone of shallow, high-velocity flow in the middle of a road section that receives water from a single direction (east) and over which the complete flood wave passes during the simulation. Point X3 represents an area of permanent ponding by the end of the simulation, while point X4 represents an area of convergent flow on the road network as it receives water from both north and south directions. Each of these situations thus provides a stringent test for the hydraulic models and allows a rigorous comparison of performance.

Several points emerge from consideration of these data. First, although the results show remarkable similarity, there are some consistent differences between the models tested. For example, the LISFLOOD-FP code always predicts the earliest arrival of the flood wave at each point, while the TRENT code always predicts it arriving last. These timing differences are only of the order of 3–4 min but, given the short duration of this event (dynamic effects have largely ceased after ~60 min into the simulation), this may be significant. Other codes fall between these extremes and show a more varied pattern with, for example, the JFLOW code behaving more like LISFLOOD-FP at point X1 and more like TRENT at point X4. In addition, a further general pattern is that the slower the predicted wave propagation speed the lower the predicted water depth at any given time during the simulation. An exception to this is the DIVAST code, which typically predicts the flood wave arriving relatively late, but then rises to a high peak depth before declining to a low final value. While these results imply significant mass balance differences between the codes, Table 1 shows that this is not the case. Moreover, these behaviours do not seem to follow any obvious pattern related to particular features shared by groups of codes, and may be due to the way in which either the friction term, the transition to and treatment of supercritical flow or viscosity is handled by each code.

Differences in maximum simulated water depth (h) for the benchmark DEM are minimal for point X1, up to ~ 8 cm at X3 and up to ~10 cm at X2 and X4. This is of the same order as the typical random component of LiDAR data vertical error (~5 cm RMSE),

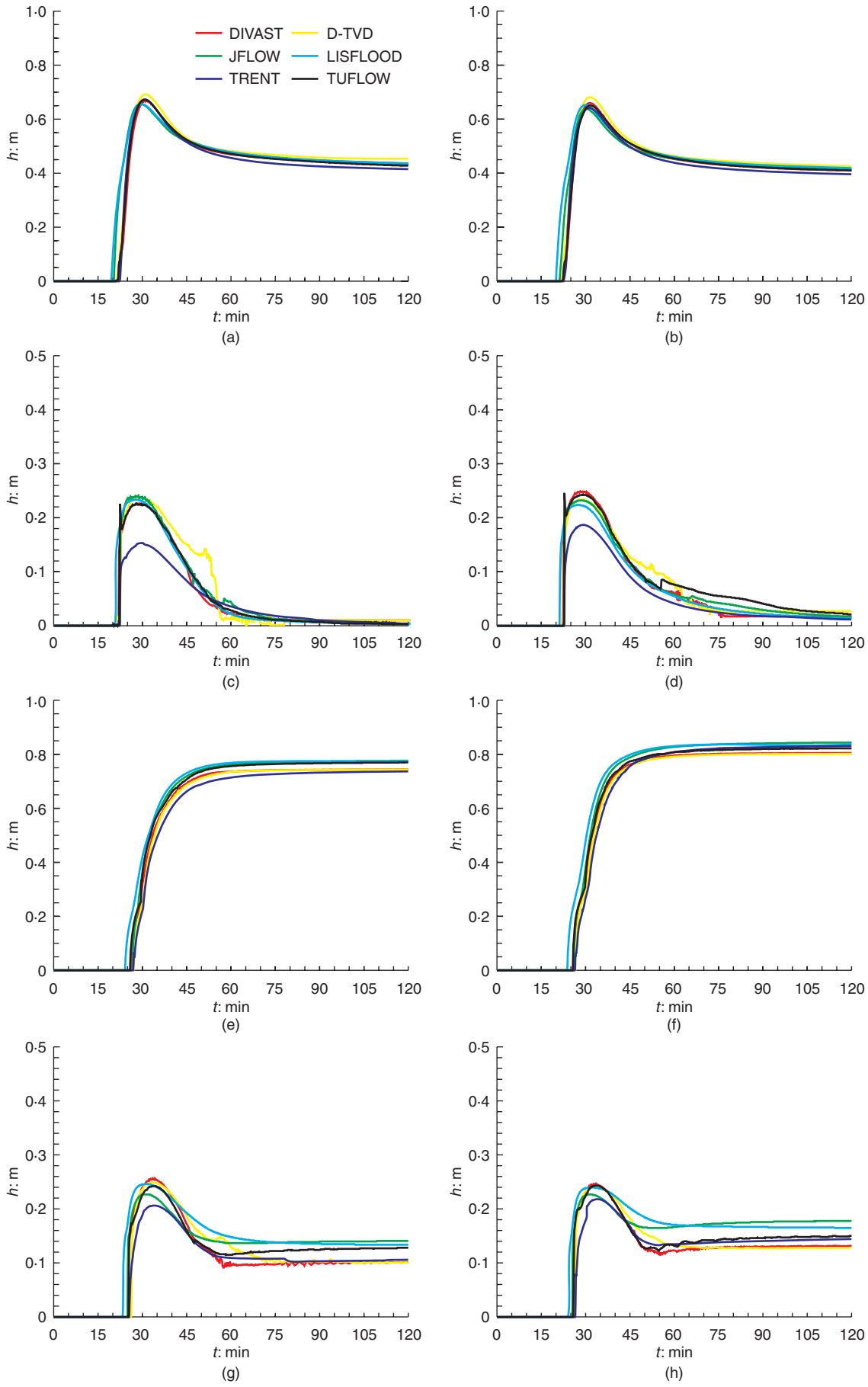


Fig. 3. Time series of water surface elevation simulated by each model at points X1, X2, X3 and X4 marked on Fig. 1 for the benchmark and hypothetical LiDAR DEMs: (a) benchmark DEM: point X1; (b) LiDAR DEM: point X1; (c) benchmark DEM: point X2; (d) LiDAR DEM: point X2; (e) benchmark DEM: point X3; (f) LiDAR DEM: point X3; (g) benchmark DEM: point X4; and (h) LiDAR DEM: point X4

which implies that in practice the differences between the models for this test case are not significant and that there is little to be gained by switching between different formulations. For the LiDAR DEM the absolute ground surface (z) errors compared to the benchmark DEM at points X1–X4 are −0·02, −0·02, 0·03 and 0·0 cm respectively, and these differences are reflected in the different water depths predicted at X1–X4 in these simulations. If anything, the range of results obtained is slightly narrower when using the LiDAR DEM ((b), (d), (f) and (h) in Fig. 3) and this implies, perhaps unsurprisingly, that degrading terrain data with random noise of the same order as that contained in LiDAR data does not affect the simulated flood wave propagation. LiDAR data therefore seem to be a good choice for urban flood modelling, even at 2 m horizontal resolution.

It is also clear that certain numerical solution schemes are prone to small (<1 cm) oscillations at shallow flow depths, especially on the hydrograph falling limb. These can be observed for the DIVAST code at points X2 and X4 and TUFLOW at point X4 on the LiDAR DEM after 50 min of simulation time. In addition, for TUFLOW, noticeable spikes in predicted water depth occur at point X2: one at 20 min into the simulation for the benchmark DEM and two at 20 and 55 min into the simulation for the LiDAR DEM. However, it is not clear whether these are physical effects or the result of a numerical instability. The presence of oscillations on the falling limb can be explained by the occurrence of grid scale oscillations, which arise due to the numerical treatment of advective acceleration terms. This effect can be overcome either by using upwind difference schemes (potentially leading to high levels of artificial diffusion) or by using higher-order accurate schemes (which can be computationally expensive and more complex in terms of treating boundaries). Either solution may be problematic for modelling flows in urban environments. Despite this, both TUFLOW and DIVAST capture the general pattern of flow variation at these locations without the computational expense of a more complex numerical technique. The JFLOW and LISFLOOD-FP codes, which are based on the diffusion wave equation and do not include the advective accelerations that lead to grid scale oscillations, produce solutions rather similar to the full shock capturing code. This implies that, while the numerical shocks may be difficult to simulate, they may not be important to overall flow development at these points.

An alternative way of visualising these results is to look at the maximum inundation extent vector predicted by each model. These are shown in Fig. 4 and suggest that the differences between models implied by Fig. 3 translate into numerous localised differences in extent predictions, especially in areas of low slope with very shallow flow ($h < 5$ cm). Where this occurs, such as in the central portion of the domain around coordinates [500, 200] and in the north-west corner, all codes predict rather fragmented arrangements of wet and dry areas. This is likely to result from a combination of local micro-topography and the particular characteristics of the numerical solver used. In particular, small numerical oscillations may have a large relative effect at very shallow flow depths resulting in a more fragmented pattern of inundation. However, these differences are rather minor in absolute terms and occur in areas with water depths less than the typical LiDAR data random vertical error. It could therefore be argued that any predictions of very shallow flow are unreliable and can be ignored unless they can be shown to affect the flow propagation.

In general, TRENT predicts a wider extent of inundation than the other codes, although maximum differences in shoreline location are only of the order of 30–40 m (e.g. at coordinates [820, 250]). While small, these differences explain the implied mass balance discrepancy between the codes suggested by the at-a-point time series shown in Fig. 3, whereby for the TRENT code a similar volume of water is spread over a wider area. Similar, but less readily discernible, effects happen with the other codes and this in turn means that the water surface slopes simulated by the codes are subtly different, and are both non-zero and different from the underlying topographic slope. This also supports the conclusion that minor differences in the dynamic behaviour of the codes may be due to the way that friction, supercritical flow transitions or viscosity effects are handled.

Lastly, both codes based on the diffusion wave equation (JFLOW and LISFLOOD-FP) show the westward propagation of the flood wave terminating in an almost identical pattern around coordinates [150, 250]. The full shallow-water codes predict that inundation continues into the region $x < 150$ where it crosses a flat and non-built-up area before, in certain cases, reconnecting with the road network in the top north-west corner of the domain. The most likely explanation for this is that connectivity between the main flow domain and the area $x < 150$ is a result of inertial flow effects that are only represented in the full shallow-water models. Essentially it appears that flow 'shoots' along the steep and straight east–west oriented road in the upper portion of the catchment. By the time flow reaches $x = 150$ m it has sufficient momentum to overtop bounding topography and continue on into the flat area beyond. Examination of water depths at $x = [150, 250]$ shows the diffusion wave codes predict values up to 5 cm lower than the full shallow-water models and it is clear from Fig. 1 that there are no obvious major topographic features present. It therefore appears that inertial effects only increase water depths by a few centimetres but that this is sufficient to allow extension of the flood wave over a micro-topographic obstruction in this case.

For the part 2 analysis, a mini-ensemble of simulations was computed for each code using the pairs of friction coefficients given in Table 2. For each model ensemble, the cumulative distribution of water depths predicted at points X1–X4 was calculated at each time step (Fig. 5). In Fig. 5 the upper and lower black lines represent the minimum and maximum water depths predicted by any ensemble member at a particular time step, while the inner shaded area represents the 25–75% range of the cumulative distribution. These results are plotted spatially in Figs 6, 7 and 8. Figs 6 and 7 show the magnitude of the range of water depth (h) values predicted by each model in each grid cell at 30 and 60 min into the simulation respectively; Fig. 8 shows the magnitude of the range of maximum predicted value of h in each grid cell over the whole simulation.

Figure 5 shows that the difference in water depths predicted by each model at points X1–X4 for a wide (but physically realistic) range of friction values is at least as great as the differences between models shown in Fig. 3. If we consider the range of friction values to be typical of that used in model calibration, this suggests that the minor differences between models can easily be subsumed within parameter optimisation, particularly given uncertainties in terrain and boundary condition data. Fig. 5 also more clearly shows the oscillatory behaviour of certain codes as

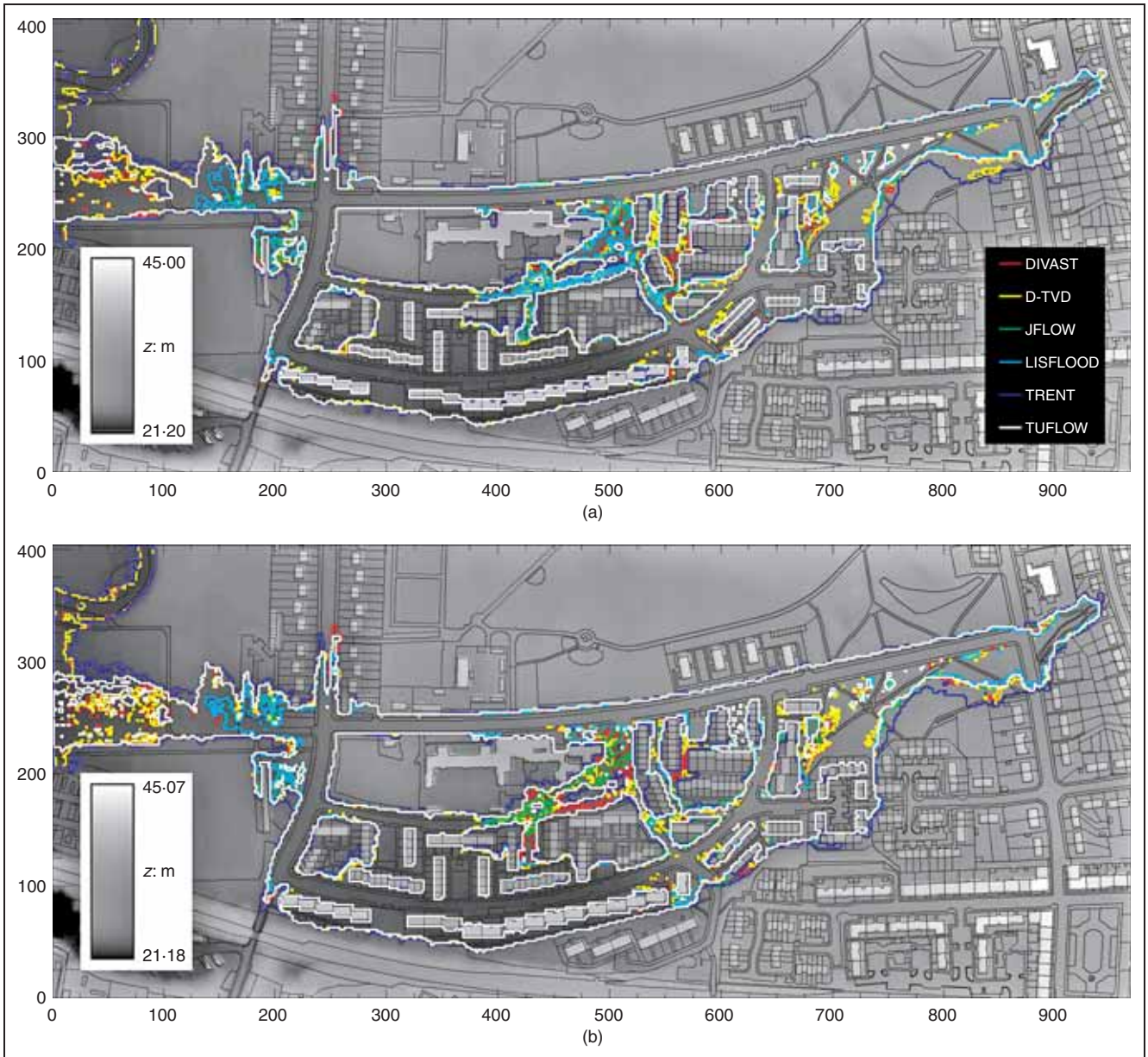


Fig. 4. Vectors of maximum inundation extent (the zero water depth contour) predicted by each model for (a) the benchmark DEM and (b) the LiDAR DEMs

noted earlier. This is probably because the lower friction value simulations result in shallower water depths and higher velocities, which enhance such effects. Oscillations are most pronounced in the full shallow-water codes that do not include shock capturing (TUFLOW and DIVAST), but minor oscillations can also be noted in the DIVAST-TVD and JFLOW results. TRENT and LISFLOOD-FP produce the smoothest results, but this can only be achieved by LISFLOOD-FP at increased computational cost compared with TRENT. Fig. 5 also demonstrates the impact of the different approaches to time step control in JFLOW and LISFLOOD-FP. Both these codes discretise the floodplain as a series of raster storage cells and then use an explicit solution to calculate the inter-cell fluxes. Both require an adaptive time step to avoid the development of numerical oscillations and mass balance errors, but JFLOW implements a conditionally stable scheme based on the Courant–Friedrichs–Levy (CFL) condition coupled with a flow reversal constraint where the potential for oscillations is minimised but cannot be eliminated. LISFLOOD-FP uses an unconditionally stable scheme based on the classic von Neumann stability analysis for a diffusion system, which guarantees an oscillation-free solution but at the expense of a ~4% increase in computational overhead compared with JFLOW. While both models produce very similar results, small oscillations can be noted at point X2 for the low-friction value simulations with JFLOW. However, as noted above, the oscillations do not seem to affect the overall wave propagation predicted by any of the codes and for similar test cases this should not be a factor in model selection.

Figures 6 and 7 show both the magnitude of the range in possible water depth values in each grid cell predicted by each model at a given time step and also how differences in the simulated wave propagation speed lead to different dynamic behaviour in each model. These differences are quite marked and, as previously noted, it seems likely that subtle differences in how the friction or other energy loss laws are coded in each model accounts for many of these phase differences. For example, Fig. 7 shows that in the TUFLOW model much of the flooded area has completely drained

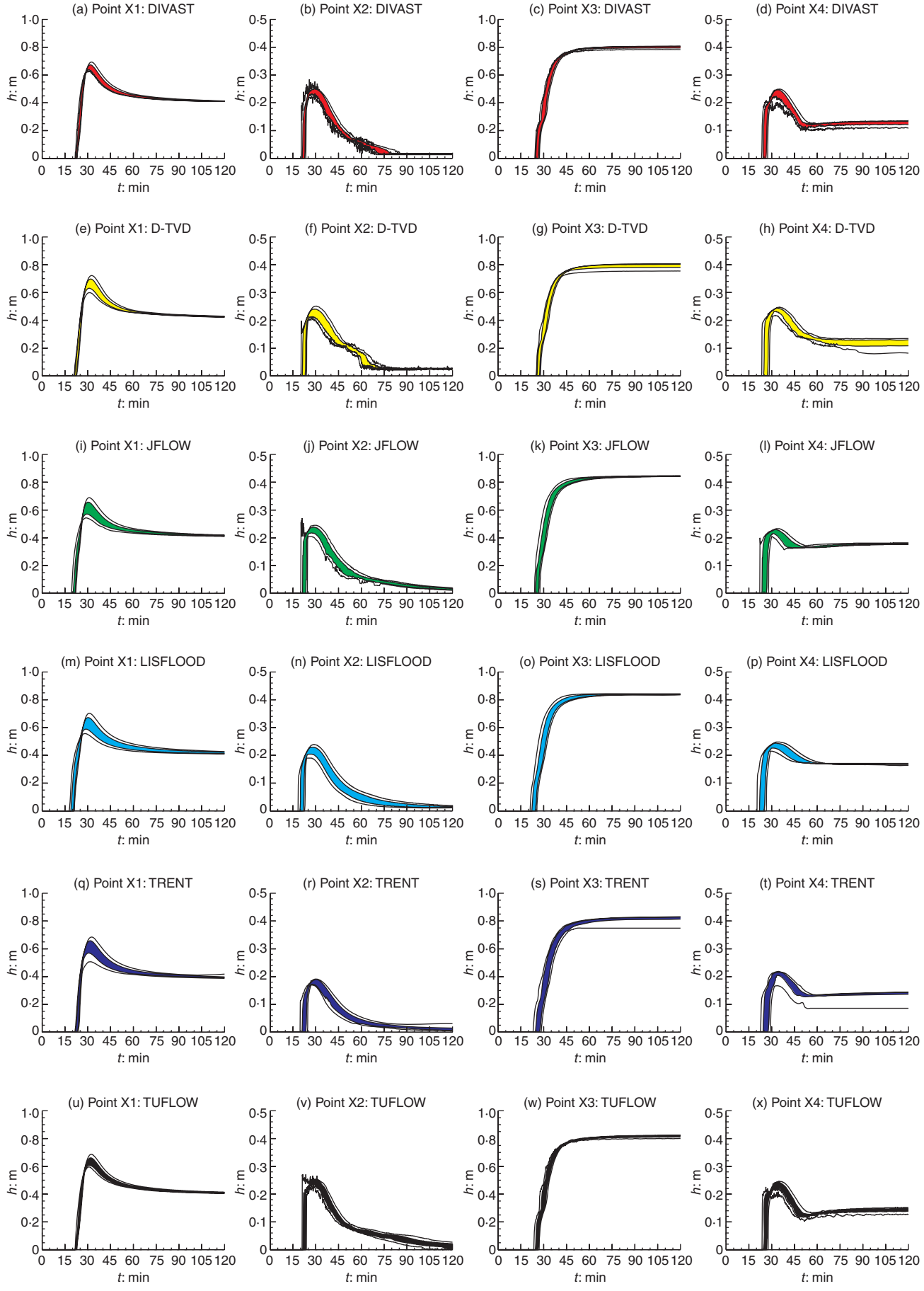


Fig. 5. Cumulative distribution functions of water depth predicted by each model at points X1–X4 on Fig. 1 for the ensemble of varying friction coefficient simulations shown in Table 2. The black lines represent the minimum and maximum water depths predicted over the ensemble, while the shaded area represents the 25–75% range of the cumulative distribution

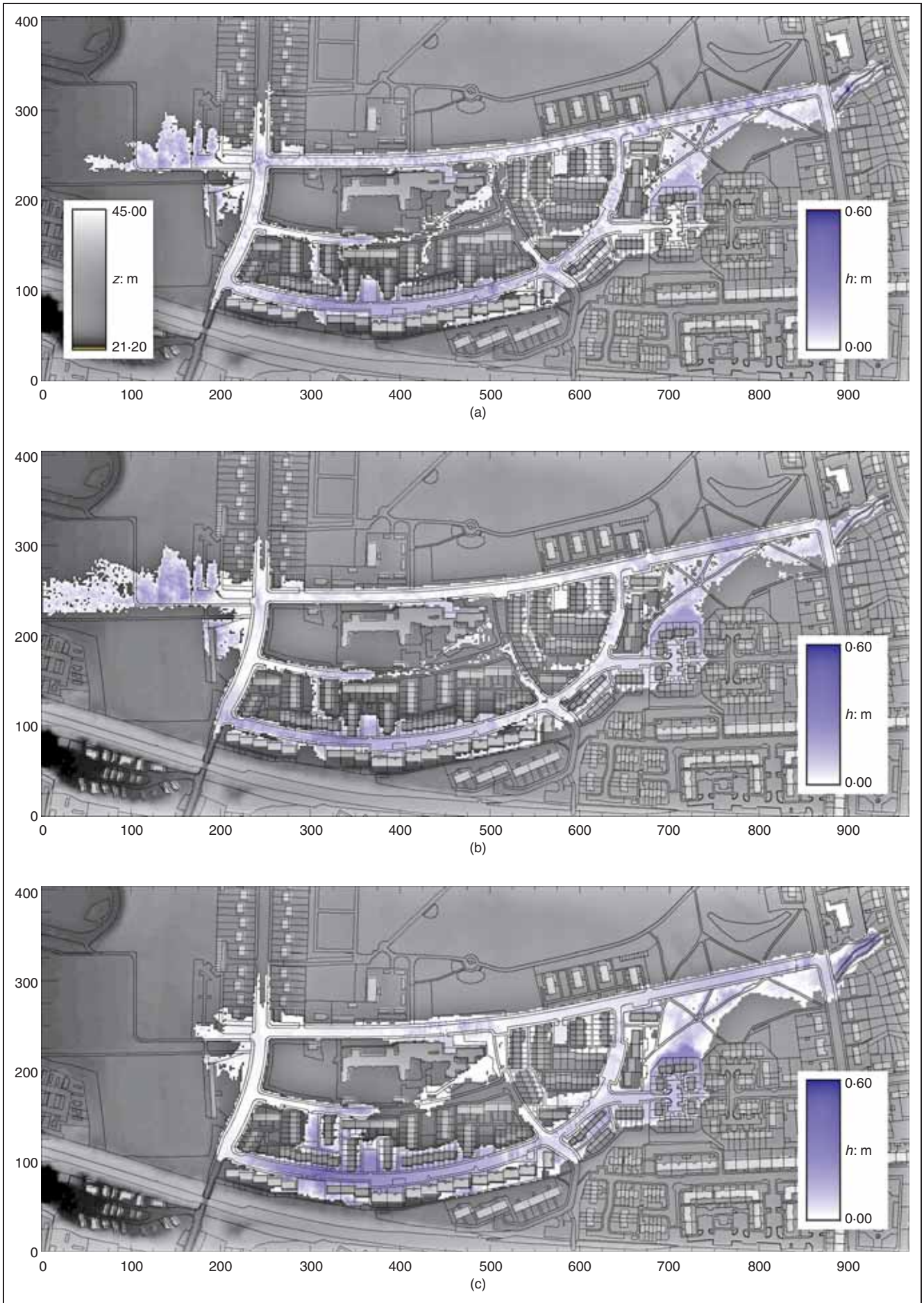
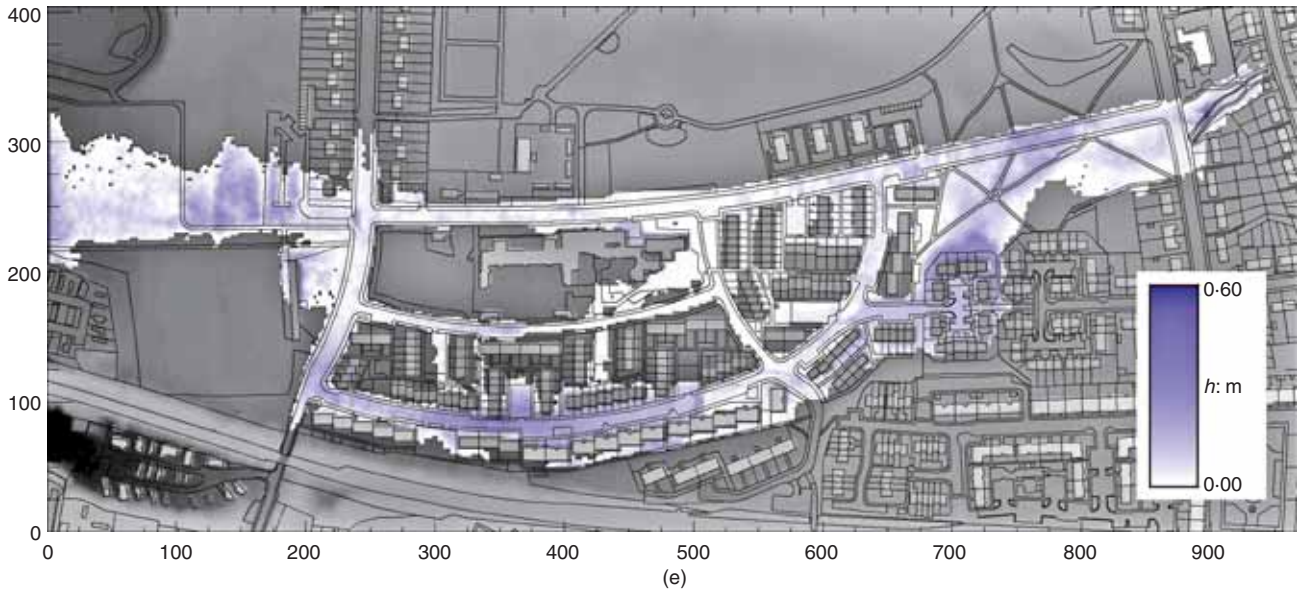


Fig. 6. Magnitude of range in water depth (h) predicted by each model at 30 minutes into the event for the ensemble of varying friction coefficient simulations shown in Table 2: (a) DIVAST; (b) DIVAST-TD; (c) JFLOW; (d) LISFLOOD; (e) TRENT; and (f) TUFLOW



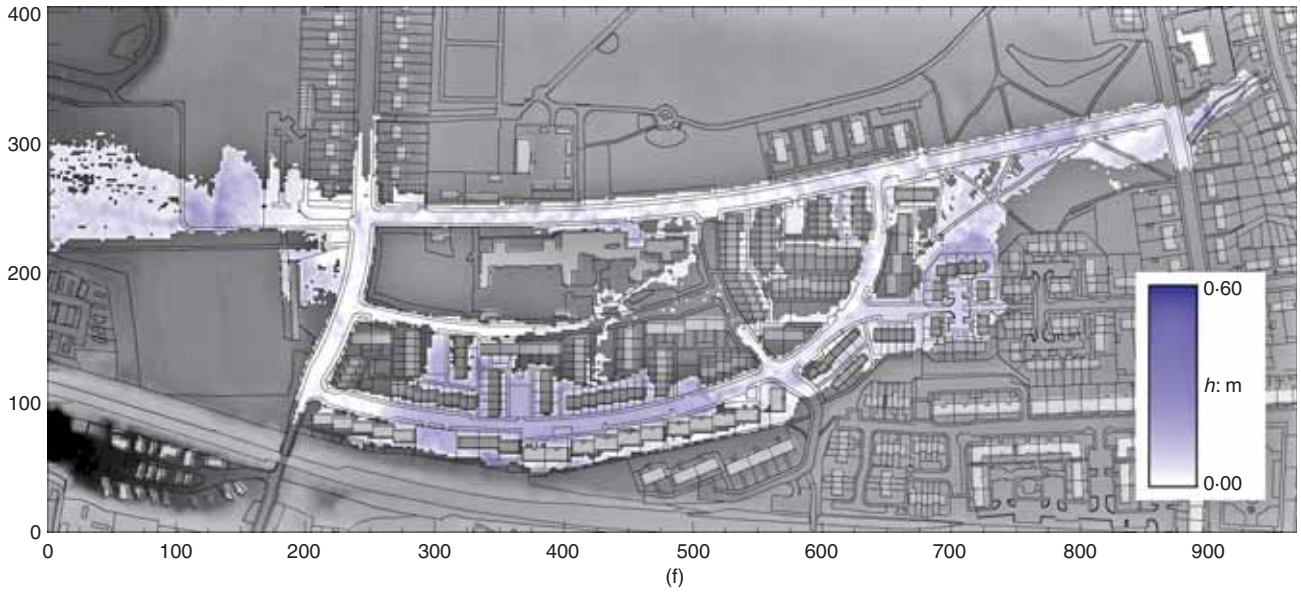


Fig. 6. Continued

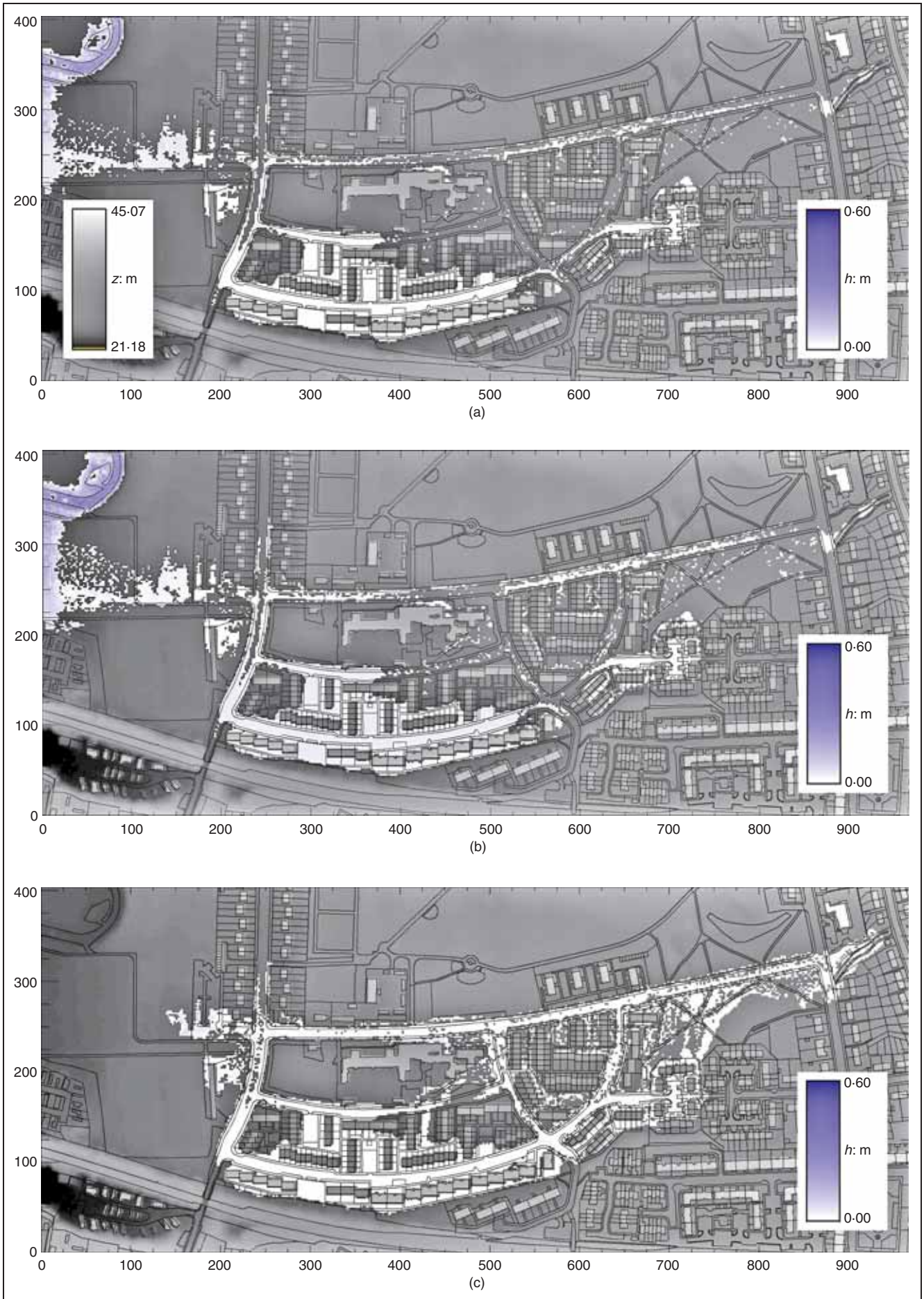
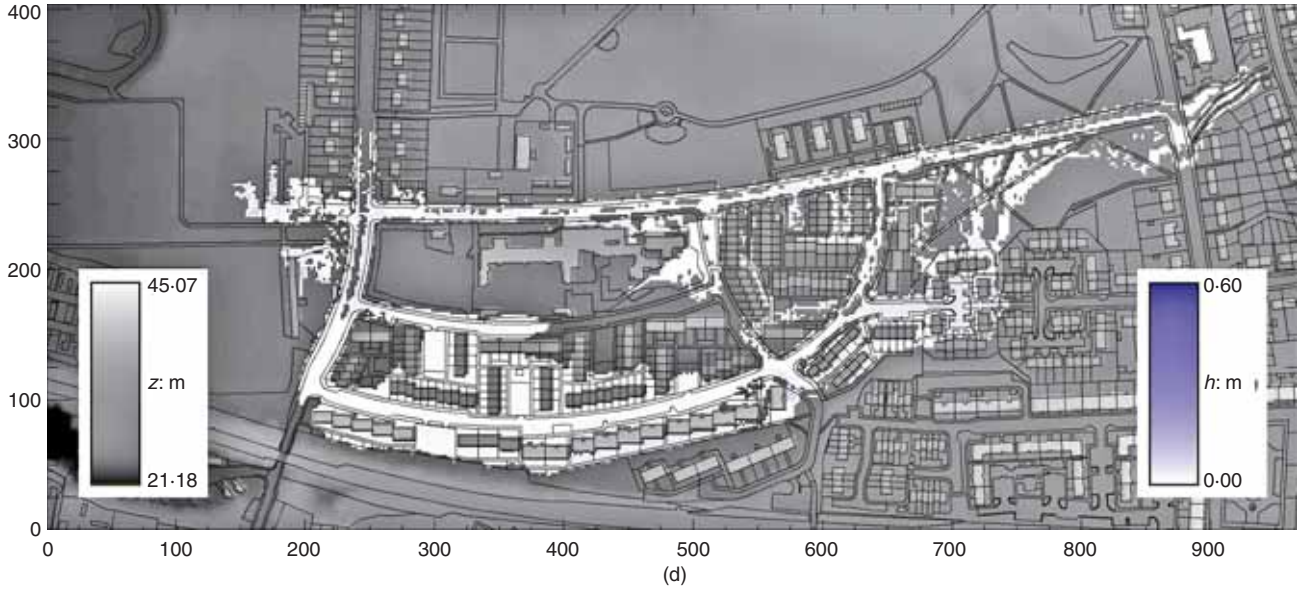
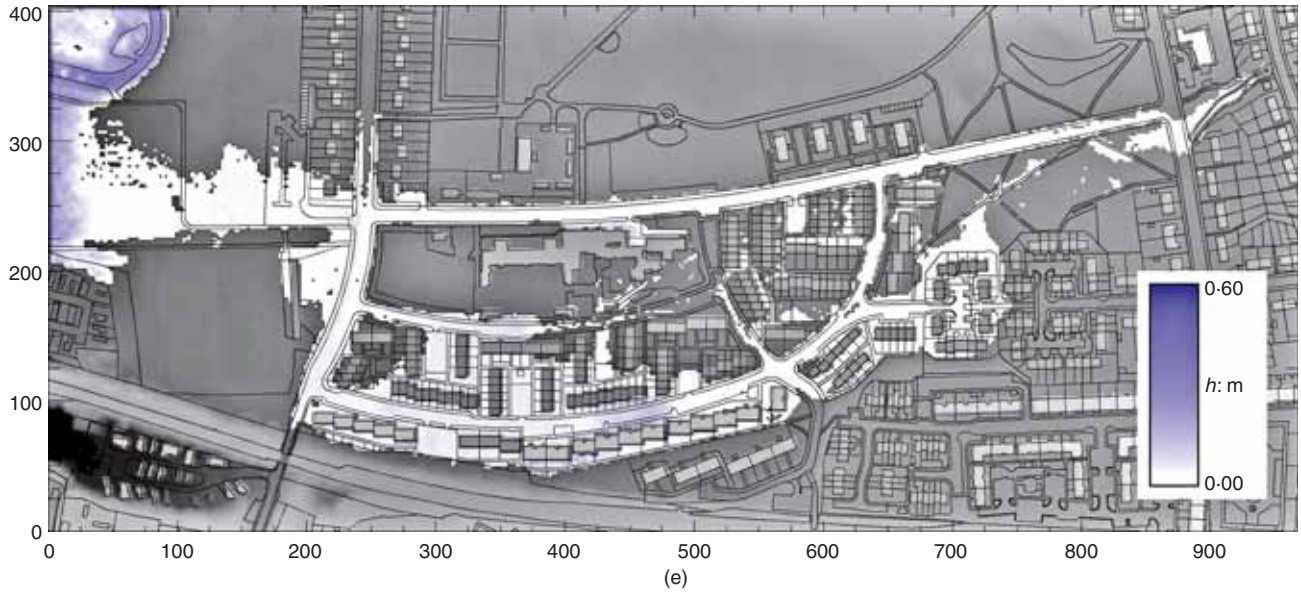


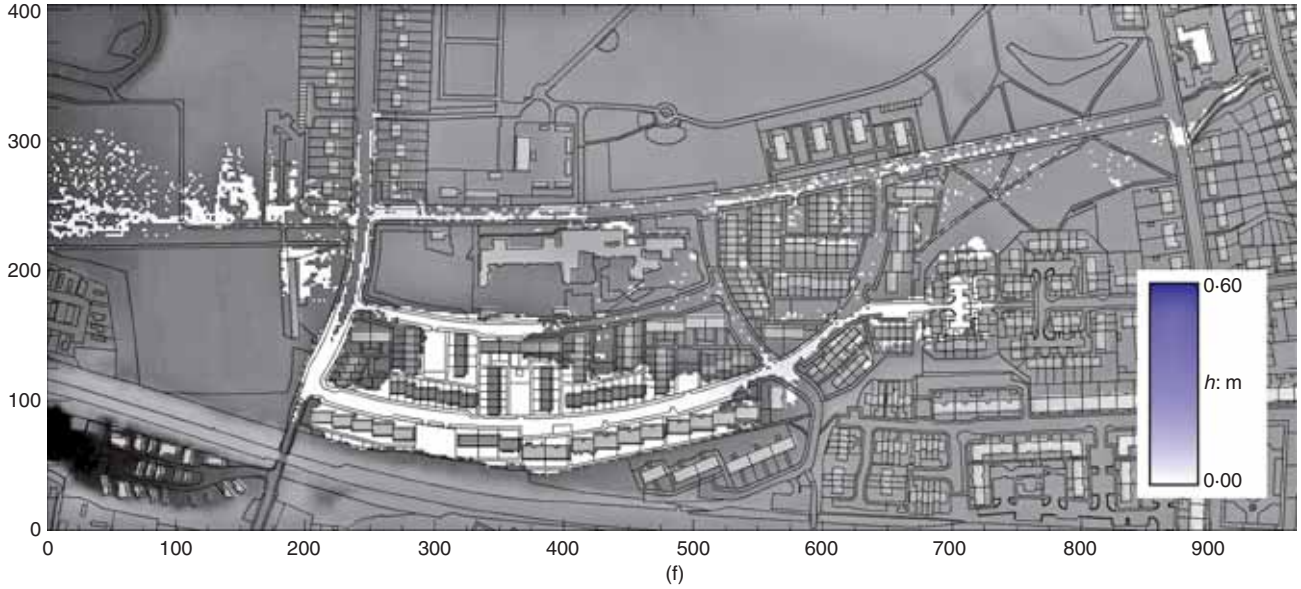
Fig. 7. Magnitude of range in water depth (h) predicted by each model at 60 minutes into the event for the ensemble of varying friction coefficient simulations shown in Table 2: (a) DIVAST; (b) DIVAST-TD; (c) JFLOW; (d) LISFLOOD; (e) TRENT; and (f) TUFLOW



(d)



(e)



(f)

Fig. 7. Continued

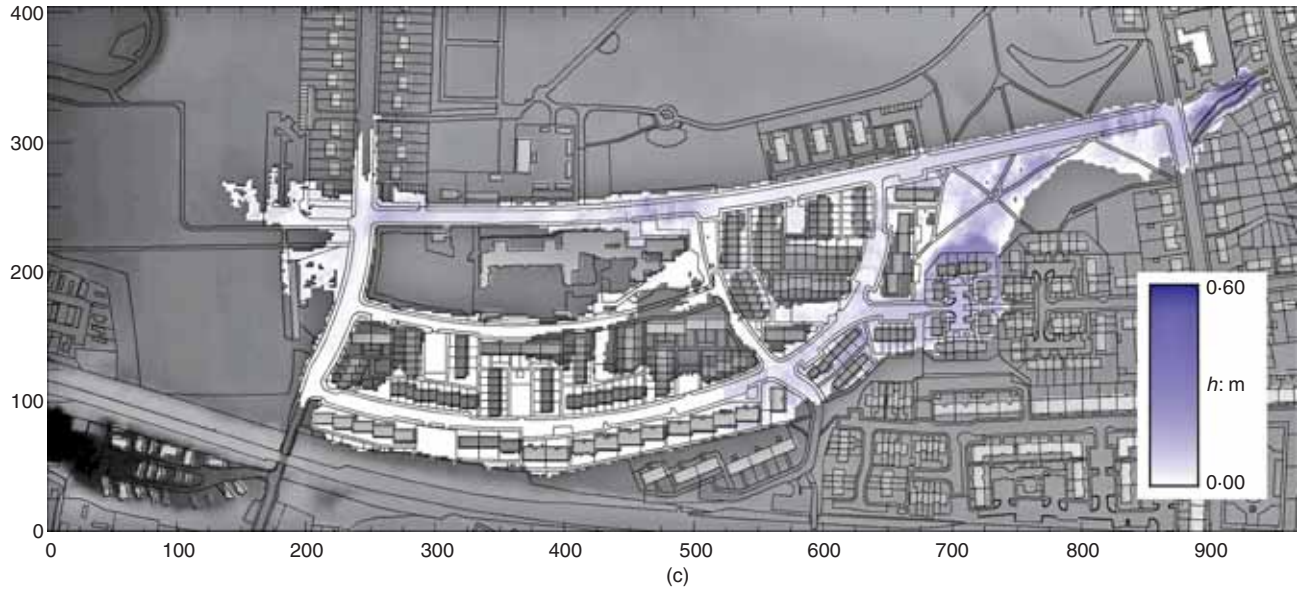
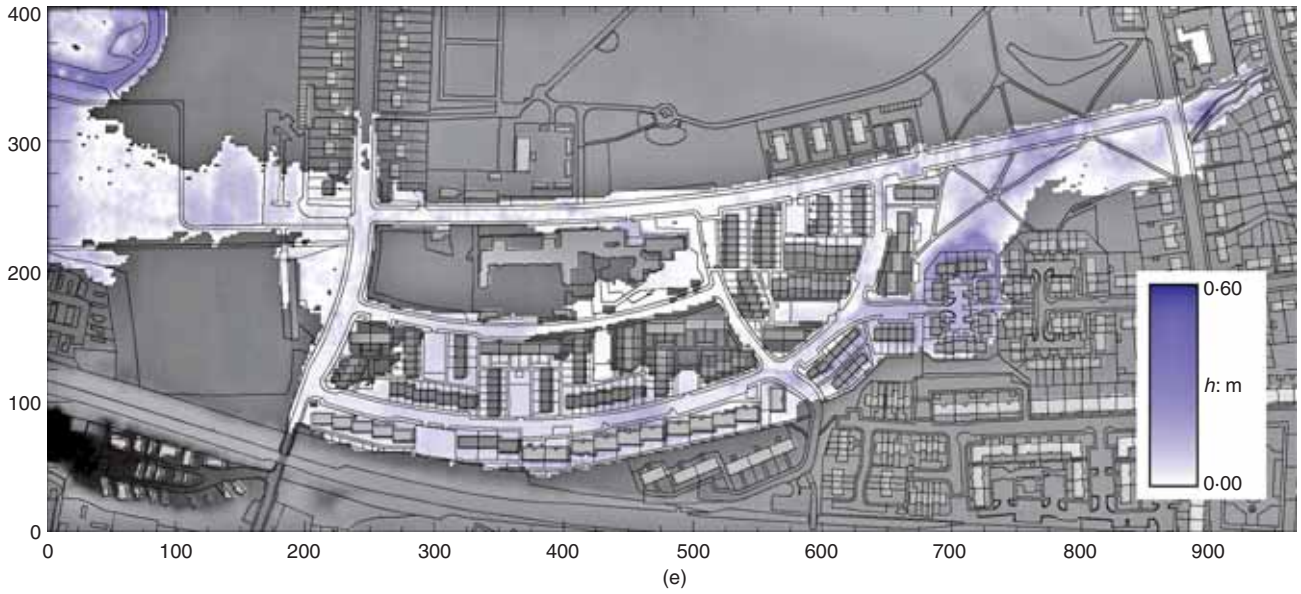
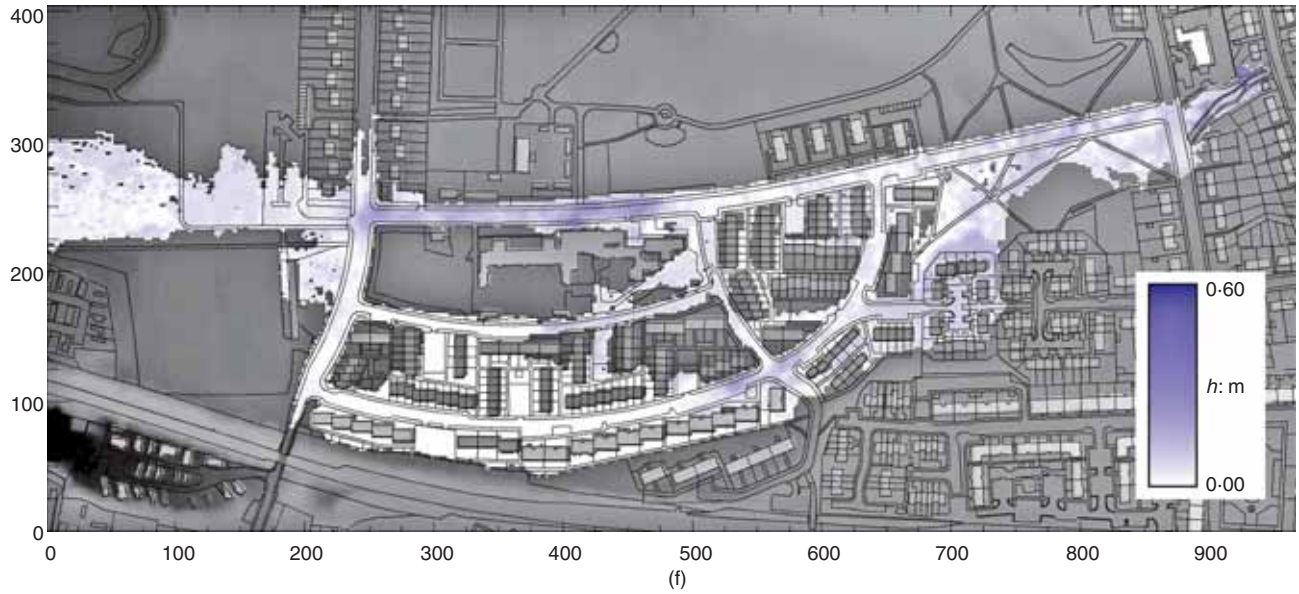


Fig. 8. Magnitude of range in maximum predicted water depth (h) in each model grid cell over the whole event for the ensemble of varying friction coefficient simulations shown in Table 2: (a) DIVAST; (b) DIVAST-TD; (c) JFLOW; (d) LISFLOOD; (e) TRENT; and (f) TUFLOW

(d)



(e)



(f)

Fig. 8. Continued

by 60 min into the simulation, whereas in TRENT large areas of the domain remain wet. Differences in predicted water depth in each pixel between the highest and lowest friction simulations are most marked at 30 min into the simulation and can be up to 60 cm depending on the particular time of arrival of the flood wave in a cell. By 60 min into the simulation, the magnitude of water depth range predicted in each grid cell for the different friction values is much lower and is generally <10 cm apart from in the north-west corner of the domain. Figs 6 and 7 also demonstrate the extension of the wave front into the area $x < 150$ by the full shallow-water codes in contrast to the diffusion wave codes.

Figure 8, while superficially similar to Figs 6 and 7, is actually quite different in that it shows the magnitude of the range in maximum water depths predicted in each model grid cell at any time during the simulation. Effectively, this is the uncertainty in predicted maximum water depth generated by forcing each model with a range of friction coefficients. This confirms the results of previous studies (e.g. Lane *et al.*[32] and Bates *et al.*[33])—that is, the response of a non-linear distributed model to uniform parameter changes is spatially complex, but also shows, for the first time, that this response is highly variable for similar but subtly different models. Thus, although each model is forced with the same spatially uniform changes in friction parameters, the patterns of changes in maximum predicted depth that result are markedly different. For example, Fig. 8 shows that the greatest uncertainty in maximum depth for the TRENT model (Fig. 8(e)) occurs around points X1 and X3, while for the TUFLOW model (Fig. 8(f)) this occurs around point X2. Similar differences occur between all other codes. This implies that calibrating 2D hydraulic models against point data using spatially lumped parameters is unlikely to result in spatially uniform changes in predicted flow quantities over the whole model domain. In part this is likely to be due to non-linear effects resulting from complex micro-topography, but may also be due to the numerical solver used and its potentially non-linear response to uniform forcing. In practice this will mean that, even though a model may replicate observed data at particular points, one cannot guarantee similar levels of performance throughout the whole model domain. This also implies that calibrating 2D hydraulic models is spatially complex and code-specific, and should be approached differently to the process of calibrating a 1D flood routing model. Typical point-based model calibration and validation data are therefore only a limited test of model performance and new measuring techniques are urgently required to better understand the behaviour of, and discriminate between, competing model formulations.

6. CONCLUSIONS

This paper has presented the most rigorous comparison to date of 2D hydraulic models applied to flows in urban areas. While all the models tested produce plausible results, subtle differences between particular groups of codes give considerable insight into both the practice and science of urban hydraulic modelling. In practical terms, it was found that terrain data available from modern LiDAR systems are sufficiently accurate and resolved for simulating urban flows, but such data need to be fused with digital map data of building topology and land use to gain maximum benefit from the information contained therein. This study also compared the relative impact of uncertainty over terrain data and model parameters on hydraulic model output. While the conclusions partly depend on the parameter and terrain error ranges fed into the analysis, it is clear that once fine spatial resolution, high-accuracy terrain data are available, uncertainty over model parameters becomes the more dominant of these factors. In this case the best way to reduce uncertainty in model predictions is to find better ways to estimate model parameters, particularly friction, or to constrain these via a robust calibration process. Remotely sensed and digital map data may provide ways of estimating physical components of the friction term directly (see Mason *et al.*[34] for an example for rural floodplains), however considerable uncertainty might still remain. In this case calibration becomes the only way of reducing uncertainty over model parameters, but to date few such data have been available for urban areas. Despite the frequent occurrence of urban floods, almost no field observations of urban flooding and no mechanisms for their routine monitoring or post-event reconstruction are available. With so little data available it becomes critically important for any urban flood modelling study to examine the impact of a physically plausible range of friction parameters on the results obtained, rather than just relying on single deterministic simulations.

In terms of the science of urban hydraulic modelling, it was found that flows in such environments are characterised by numerous transitions to supercritical flow and numerical shocks. However, the effect of these are localised and they do not appear to affect overall wave propagation. For flood risk studies it therefore does not appear necessary to utilise a shock capturing code unless an oscillation-free solution is important for some other reason. For this test case, inertial effects appear to be important in terms of inundation extent, although it is worth noting that topographic slopes in the domain are high compared to typical floodplain values; furthermore, the topology of the site (with long uninterrupted streets oriented in the main flow direction) may act to increase the importance of the inertial terms. This conclusion therefore may not readily hold for other urban areas. What is clear, however, is that small differences in predicted water elevation and micro-topographic barriers can combine in urban settings to give significant differences in predicted inundated area. The water elevation differences that generate these effects can be of the same order of magnitude as the height errors in available terrain data (i.e. $\sim$5 cm RMSE) indicating the nonlinear sensitivity of urban flooding to micro-topographic effects. However, for practical applications, uncertainty over inflow boundary conditions (at least $\pm$10% for even the best available data) is likely to induce water elevation differences over the domain that are much greater than 5 cm. Uncertainty over boundary forcing may therefore be more important than whether inertial effects are included or not. Lastly, this study shows that assumptions about which classes of code will be more computationally expensive in particular situations do not always hold. In particular, it was perhaps surprising to find that storage cell codes were more expensive than full shallow-water codes for the 2 m grid used here. A solution may be to modify explicit storage cell codes to include inertial terms (or simple approximations to these) that may allow the use of a larger stable time step, and hence quicker run times, as well as including elements of the flow physics that may be important in particular urban settings.

ACKNOWLEDGEMENTS

The research reported in this paper was conducted as part of the Flood Risk Management Research Consortium (FRMRC).

The FRMRC is supported by grant GR/S76304 from the Engineering and Physical Sciences Research Council, in partnership with the Natural Environment Research Council, the Defra/EA Joint Research Programme on Flood and Coastal Defence, the Scottish Executive and the Rivers Agency (Northern Ireland). This financial support is gratefully acknowledged. The authors are also grateful to Glasgow City Council and Infoterra for providing LiDAR data for Glasgow, to WBM for the use of their TUFLOW software and Ordnance Survey for providing Mastermap® data.

REFERENCES

1. Gee D. M., Anderson M. G. and Baird L. Large scale floodplain modelling. *Earth Surface Processes and Landforms*, 1990, **15**, No. 6, 512–523.
2. Bates P. D., Stewart M. D., Siggers G. B., Smith C. N., Hervouet J.-M. and Sellin R. H. J. Internal and external validation of a two-dimensional finite element model for river flood simulation. *Proceedings of the Institution of Civil Engineers, Water Maritime and Energy*, 1998, **130**, No. 3, 127–141.
3. Horritt M. S. Calibration of a two-dimensional finite element flood flow model using satellite radar imagery. *Water Resources Research*, 2000, **36**, No. 11, 3279–3291.
4. Nicholas A. P. and Mitchell C. A. Numerical simulation of overbank processes in topographically complex floodplain environments. *Hydrological Processes*, 2003, **17**, No. 4, 727–746.
5. Romanowicz R. and Beven K. Estimation of flood inundation probabilities as conditioned on event inundation maps. *Water Resources Research*, 2003, 39, No. 3, art. No. 1073. DOI: 10.1029/2001 WR00 1056.
6. Neelz S., Pender G., Villanueva I., Wilson M., Wright N. G., Bates P., Mason D. and Whitlow C. Using remotely sensed data to support flood modelling. *Proceedings of the Institution of Civil Engineers, Water Management*, 2006, **159**, No. 1, 35–43.
7. Bates P. D. and De Roo A. P. J. A simple raster-based model for floodplain inundation. *Journal of Hydrology*, 2000, **236**, No. 1–2, 54–77.
8. Horritt M. S. and Bates P. D. Evaluation of 1-D and 2-D numerical models for predicting river flood inundation. *Journal of Hydrology*, 2002, **268**, No. 1–4, 87–99.
9. Djokic D. and Maidment D. R. Terrain analysis for urban stormwater modelling. *Hydrological Processes*, 1991, **5**, No. 1, 115–124.
10. Mark O., Weesakul S., Apirumanekul C., Aroonnet S. B. and Djordjević S. Potential and limitations of 1D modelling of urban flooding. *Journal of Hydrology*, 2004, **299**, No. 3–4, 284–299.
11. Calenda G., Calvani L. and Mancini C. P. Simulation of the great flood of December 1870 in Rome. *Proceedings of the Institution of Civil Engineers, Water and Maritime Engineering*, 2003, **156**, No. 4, 305–312.
12. Aronica G. T. and Lanza J. G. Drainage efficiency in urban areas: a case study. *Hydrological Processes*, 2005, **19**, No. 5, 1105–1119.
13. Mignot E., Paquier A. and Haider S. Modelling floods in a dense urban area using 2D shallow water equations. *Journal of Hydrology*, 2006, **327**, No. 1, 186–199.
14. Guinot V. and Soares-Frazao S. Flux and source term discretization in two-dimensional shallow water models with porosity on unstructured grids. *International Journal for Numerical Methods in Fluids*, 2006, **50**, No. 3, 309–345.
15. Hsu M. H., Chen S. H. and Chang T. J. Inundation simulation for urban drainage basin with storm sewer system. *Journal of Hydrology*, 2000, **234**, No. 1–2, 21–37.
16. Yu D. and Lane S. N. Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 1: mesh resolution effects. *Hydrological Processes*, 2006, **20**, No. 7, 1541–1565.
17. Yu D. and Lane S. N. Urban fluvial flood modelling using a two-dimensional diffusion-wave treatment, part 2: development of a sub-grid-scale treatment. *Hydrological Processes*, 2006, **20**, No. 7, 1567–1583.
18. Schmitt T. G., Thomas M. and Ettrich N. Analysis and modelling of flooding in urban drainage systems. *Journal of Hydrology*, 2004, **299**, No. 3–4, 300–311.
19. Lhomme J., Bouvier C. and Perrin J.-L. Applying a GIS-based geomorphological routing model in urban catchments. *Journal of Hydrology*, 2004, **299**, No. 3–4, 203–216.
20. Smith M. B. Comment on 'Analysis and modelling of flooding in urban drainage systems'. *Journal of Hydrology*, 2006, **317**, No. 3–4, 355–363.
21. Leopardi A., Oliveri E. and Greco M. Two-dimensional modeling of floods to map risk-prone areas. *ASCE Journal of Water Resources Planning and Management*, 2002, **128**, No. 3, 168–178.
22. Syme W. J. *Dynamically Linked Two-dimensional/One-dimensional Hydrodynamic Modelling Program for Rivers, Estuaries & Coastal Waters*. MEngSc thesis, University of Queensland, Australia, 1991.
23. Falconer R. A. A water quality simulation study of a natural harbour. *American Society of Civil Engineers, Journal of Waterway, Port, Coastal and Ocean Engineering*, 1986, **112**, No. 1, 15–34.
24. Liang D., Falconer R. A. and Lin B. Comparison between TVD-MacCormack and ADI-type solvers of the shallow water equations. *Advances in Water Resources*, 2006, **29**, No. 12, 1833–1845.
25. Liang D., Lin B. and Falconer R. A. Simulation of rapidly varying flow using an efficient TVD-MacCormack scheme. *International Journal for Numerical Methods in Fluids*, 2007, **53**(5), No. 3, 811–826.
26. Villanueva I. and Wright N. G. Linking Riemann and storage cell models for flood prediction. *Proceedings of the Institution of Civil Engineers, Journal of Water Management*, 2006, **159**, No. 1, 27–33.
27. Bradbrook K. F., Lane S. N., Waller S. G. and Bates P. D. Two-dimensional diffusion wave modelling of flood inundation using a simplified channel representation. *International Journal of River Basin Management*, 2004, **2**, No. 3, 211–223.
28. Hunter N. M., Horritt M. S., Bates P. D., Wilson M. D. and Werner M. G. F. An adaptive time step solution for raster-based storage cell modelling of floodplain inundation. *Advances in Water Resources*, 2005, **28**, No. 9, 975–991.
29. Gomes-Pereira L. M. and Wicherson R. J. Suitability of laser data for deriving geographical data: a case study in the context of management of fluvial zones. *ISPRS Journal of Photogrammetry and Remote Sensing*, 1999, **54**, No. 2–3, 105–114.

30. Cobby D. M., Mason D. C. and Davenport I. J. Image processing of airborne scanning laser altimetry for improved river flood modelling. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2001, **56**, No. 2, 121–138.
31. Aronica G., Bates P. D. and Horritt M. S. Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrological Processes*, 2002, **16**, No. 10, 2001–2016.
32. Lane S. N., Richards K. S. and Chandler J. H. Distributed sensitivity analysis in modelling environmental systems. *Proceedings of the Royal Society, Series A*, 1994, **447**, No. 1929, 49–63.
33. Bates P. D., Horritt M. and Hervouet J.-M. Investigating two-dimensional finite element predictions of floodplain inundation using fractal generated topography. *Hydrological Processes*, 1998, **12**, No. 8, 1257–1277.
34. Mason D., Cobby D. M., Horritt M. S. and Bates P. D. Floodplain friction parameterization in two-dimensional river flood models using vegetation heights derived from airborne scanning laser altimetry. *Hydrological Processes*, 2003, **17**, No. 9, 1711–1732.
35. Toro E. F. *Shock-capturing Methods for Free-surface Shallow Flows*. Wiley, Chichester, 2001.