



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/79279/>

Version: Accepted Version

---

**Article:**

Neal, J, Bates, P, Villanueva, I et al. (2012) How much physical complexity is needed to model flood inundation? *Hydrological Processes*, 26 (15). 2264 - 2282. ISSN: 0885-6087

<https://doi.org/10.1002/hyp.8339>

---

**Reuse**

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# How much physical complexity is needed to model flood inundation?

Jeffrey Neal<sup>1\*</sup>, Ignacio Villanueva<sup>2</sup>, Nigel Wright<sup>3</sup>, Thomas Willis<sup>3</sup>, Timothy Fewtrell<sup>4</sup> and Paul Bates<sup>1</sup>

<sup>1</sup>School of Geographical Sciences, University of Bristol, Bristol. BS8 1SS. UK,

<sup>2</sup>Ofiteco Ltd., Avenida de Portugal, 81. 28071, Madrid. Spain.

<sup>3</sup>School of Civil Engineering, University of Leeds, Leeds. LS2 9JT. UK

<sup>4</sup>Willis Research Network, Willis Re, Willis Building, 51 Lime Street, London. EC3M 7DQ. UK

\* Corresponding author. Tel.: +44 (0)117 92 88290; fax: +44 (0)117 928 7878

E-mail address: [j.neal@bristol.ac.uk](mailto:j.neal@bristol.ac.uk)

## Abstract

Two-dimensional flood inundation models are widely used tools for flood hazard mapping and an essential component of statutory flood risk management guidelines in many countries. Yet we still don't know how much physical complexity a flood inundation model needs for a given problem. Here, three two-dimensional explicit hydraulic models, that can be broadly defined as simulating diffusive, inertial or shallow water waves, have been benchmarked using test cases from a recent Environment Agency for England and Wales (EA) study, where results from industry models are also available. To ensure consistency the three models were written in the same code and share subroutines for all but the momentum (flow) and time stepping calculations. The diffusive type model required much longer simulation times than the other models, whilst the inertia model was the quickest. For flows that vary gradually in time, differences in simulated velocities and depths due to physical complexity were within 10% of the simulations from a range of industry models. Therefore, for flows that vary gradually in time it appears unnecessary to solve the full two-dimensional shallow water equations. As expected however, the simpler models were unable to simulate supercritical flows accurately. Finally, implications of the results for future model benchmarking studies are discussed in light of a number of subtle factors that were found to cause significant differences in simulations relative to the choice of model.

**Keywords:** Hydraulic modelling, Benchmarking, flood inundation, physical complexity, simple models

## 1 Introduction

Two-dimensional flood inundation models are widely used tools for flood hazard mapping and an essential component of statutory flood risk management guidelines in many countries. For both industry and research applications there are a wide variety of shallow water codes that account for varying degrees of physical complexity and offer subtly different solutions to a given problem. Understanding the potential differences between these codes for industry applications was a key driver of recent two-dimensional model benchmarking reports commissioned by the Environment Agency for England and Wales (Crowder et al., 2004; Néelz and Pender, 2010) to aid procurement decisions and maintain standards.

Previous model benchmarking studies have usually tracked the development of new numerical methods or the adoption of new techniques as the necessary data or computational resources become available. For example, the increasing use of two-dimensional models over one-dimensional

41 models during the past decade has been partly driven by developments in digital elevation modelling  
42 (DEM's), especially from airborne LiDAR data (Cobby et al., 2001). Thus, as the capability has  
43 developed it has been necessary to better understand the effects of moving to two-dimensions  
44 given different applications. Comparisons between one-dimensional, two-dimensional and coupled  
45 one-two dimensional river modelling approaches (e.g. Horritt and Bates (2002); Werner (2004) and  
46 Tayefi et al. (2007)) have highlighted conceptual problems with the one-dimensional approach  
47 applied to overbank flows when compared to the sometimes complex flow pathways simulated by  
48 two-dimensional models. Leopardi et al. (2002) includes a more extensive review of benchmarking  
49 studies on coupled 1D and 2D codes from the 1990's.

50 Benchmarking studies will often take newly developed or simplified models and compare them to  
51 more established or complex models. Such work is usually motivated by the computational cost of  
52 many two-dimensional model codes, which still restricts the use of hydraulic models within  
53 Monte Carlo frameworks, despite continued advances in computer hardware (Neal et al., 2010;  
54 Lamb et al., 2010). The significant cost associated with each simulation has maintained interest in  
55 techniques that can approximate simulations from full two-dimensional shallow water models with  
56 less computation. Recent examples include porosity based methods for representing sub-grid scale  
57 features in coarse resolution models (Guinot and Soares-Fraza, 2006; Yu and Lane, 2006; McMillan  
58 and Brasington, 2007,) methods without momentum such as volume spreading (Hall et al., 2003;  
59 Gouldby et al., 2008), models that consider inertia and diffusion but ignore advection (Aronica et al.,  
60 1998; Bates et al., 2010), diffusive models (Prestininzi et al., 2009) and emulators (Hall et al., 2011).  
61 Bates and De Roo (2000) and Horritt and Bates, (2001) compared a storage cell approximation of a  
62 diffusion wave with an unstructured finite element model of a rural river and floodplain. Differences  
63 were noted between the models, particularly regarding the ability of the storage cell model to  
64 predict wave speed, which was later improved upon by Hunter et al. (2005) through the  
65 implementation of an adaptive time-step constraint. However, the models considered by Bates and  
66 De Roo (2000) and Horritt and Bates (2001) simulated similar inundation extents in that differences  
67 were less than the expected errors in the remotely sensed data used to evaluate the models at that  
68 time. Lack of observation data turned out to be a common problem when moving from purely  
69 comparing model simulations to evaluating model accuracy for spatially distributed real world  
70 events at and above the reach scale (e.g. Horritt et al., 2000; Mignot et al., 2006; Werner et al.,  
71 2005; Neal et al., 2009).

72 Other studies have looked at benchmarking alternative two-dimensional shallow water models (e.g.  
73 Horritt, 2007), where recent work has focused on urban settings because the risks are typically  
74 greater than in rural areas and the availability of DEM data perceived as fit for purpose has been  
75 increasing (Fewtrell, 2011). Hunter et al. (2008) compared three full shallow water codes (Syme,  
76 1991; Villanueva and Wright, 2006, Liang et al., 2006) and two diffusive codes (Bradbrook et al.,  
77 2004; Hunter et al., 2005) for an urban test site in Glasgow, UK. They found differences in the depth  
78 and extent dynamics given the range of physical process representations and numerical solvers  
79 tested, although the significance of these given uncertainty in factors such as inflow discharges and  
80 surface friction is an ongoing debate within the community. This test case was subsequently used to  
81 evaluate mesh generation techniques (Schubert et al., 2008), grid resolution effects (Fewtrell et al.,  
82 2008), methods of parallelising models (Neal et al., 2010) and uncertainty in the magnitude of flow  
83 for given rainfall return periods (Aronica et al., submitted).

84 Néelz and Pender (2010) benchmarked the majority of industry codes used for flood risk modelling  
85 in the UK by the Environment Agency and commercial consultants. The industry codes (including:  
86 ISIS2D, SOBEK, TUFLOW, MIKEFLOOD, InfoWorks2D, Flowroute & JFLOW-GPU) were required to  
87 simulate velocity and depth dynamics across the range test cases listed in Table 1, which were  
88 designed to cover most statutory flood risk modelling requirements in the UK. In this paper, three  
89 physical process representations of floodplain flow, described in the next section, will be  
90 benchmarked using four of the test cases from this study identified in Table 1.

91 One of the key issues when comparing industry codes is the difficulty in achieving suitable  
92 consistency between test case implementations. Without this there is significant uncertainty in the  
93 cause of simulation differences, meaning discrepancies between results cannot be attributed to a  
94 narrow enough range of factors to allow useful conclusions to be drawn. Differences in how  
95 modellers interpret the same test case are the easiest to avoid by using a single code where the  
96 state variables of each 'model' (elevation, inflow etc) can be taken from a shared environment. This  
97 also means model parameters will be sourced and manipulated in a consistent manner (e.g. in the  
98 model used here the roughness across a cell edge is a linear interpolation of the roughness  
99 attributed to each neighbouring grid cell). This degree of constancy can be assumed between models  
100 in different codes but is not easily verified without extensive analysis of the source code, which for  
101 commercial confidentiality reasons may not be available. Treatment of wetting and drying, wet to  
102 dry edges, friction and source terms, inflows and normal depth boundaries may also differ subtly  
103 between codes, both in terms of approach and parameterisation. All of these factors may alter  
104 simulation results before any consideration of numerical scheme and physical complexity is taken  
105 into account, adding significant uncertainty to any discussion.

106 To address this issue the models in this study were written within a single code ensuring a level of  
107 constancy between model state variables and parameters that was not possible in the studies  
108 presented above. The models used were the full shallow water model LISFLOOD-Roe based on the  
109 TRENT model (Villanueva and Wright, 2006), an inertial wave model LISFLOOD-ACC (Bates et al.,  
110 2010) that represents a simplified shallow water wave and a diffusion wave model LISFLOOD-ATS  
111 (Hunter et al., 2005). These all form part of the single LISFLOOD-FP code. Results from the industry  
112 models in Néelz and Pender (2010) have been used to provide context to the LISFLOOD-FP  
113 simulations, especially in regard to the magnitude of difference between the simpler and full shallow  
114 water models. A key interest not covered by previous studies will be the ability of the simpler  
115 models to simulate velocity and therefore flood hazard, along with the sensitivity to some often  
116 hidden coding and parameterisation decisions. Results are discussed under two sections. The first  
117 deals with the inter-comparison of the three models and their response to each test, from which,  
118 conclusions regarding the necessary physical complexity for each test are drawn. The second section  
119 discusses the implications of these results for the Environment Agency model benchmarking study  
120 and implications for model benchmarking best practice.

## 121 **2 Models**

122 This benchmarking exercise focuses on three two-dimensional hydraulic models within the  
123 LISFLOOD-FP code. Each model is a module within the code that is activated by a key word in the  
124 model parameter file prior to simulation. Once initialised, all models utilise the same input files and  
125 data structures, along with many shared subroutines. In fact, as the continuity equation is the same  
126 for all models they only differ in respect of the flow equation and time stepping. The models were

127 chosen to cover a typical range of physical complexities based on the shallow water equations or  
 128 simplifications of them. The three models (LISFLOOD-Roe, LISFLOOD-ACC and LISFLOOD-ATS) are  
 129 summarised in Table 2 and described in the following section.

130 LISFLOOD-Roe is the two-dimensional shallow water model from Villanueva and Wright (2006), thus  
 131 it calculates the flow according to the complete Saint Venant formulation. The method is based on  
 132 the Godunov approach and uses an approximate Riemann solver by Roe (Roe, 1981). The explicit  
 133 discretisation is first-order in space on a raster grid. It solves the full shallow water equations with a  
 134 shock capturing scheme. LISFLOOD-Roe uses a point-wise friction based on the Manning's equation,  
 135 while the domain boundary/internal boundary (wall) uses the ghost cell approach. The stability of  
 136 this approach is approximated by the CFL condition for shallow water models, which is shown in  
 137 Table 2. As the complete model formulation is quite lengthy and relatively well known it is not  
 138 reproduced here.

139 LISFLOOD-ACC is a one-dimensional inertial model (e.g. advection is ignored), and hence is  
 140 decoupled in x and y directions for two-dimensional simulation over a raster grid. The method is  
 141 first-order in space and explicit in time, but uses a semi-implicit treatment for the friction term to aid  
 142 stability (See Bates et al., 2010). To calculate the flow between cells the equation derived by Bates et  
 143 al. (2010) is implemented:

$$Q^{t+\Delta t} = \frac{q^t - gh_{flow}^t \Delta t \frac{\Delta(h^t + z)}{\Delta x}}{\left(1 + gh_{flow}^t \Delta t n^2 |q^t| / (h_{flow}^t)^{10/3}\right)} \Delta x$$

144 where  $g$  is the acceleration due to gravity ( $\text{ms}^{-2}$ ),  $n$  is Manning's roughness coefficient ( $\text{sm}^{-1/3}$ ),  $h$  is  
 145 depth (m) and  $z$  is cell elevation (m) such that  $\Delta(h^t + z)$  is the difference in water surface elevation  
 146 between two cells (m),  $\Delta x$  is the cell resolution (m),  $Q$  is flow ( $\text{m}^3 \text{s}^{-1}$ ),  $q$  is water flux ( $\text{m}^2 \text{s}^{-1}$ ) and  
 147  $h_{flow}^t$  is the depth of flow between two cells (m) defined as the maximum water surface elevation in  
 148 neighbouring cells minus the maximum bed elevation in neighbouring cells. This model formulation  
 149 was used previously by Neal et al. (2011) under the name LISFLOOD-INT. The stability of this  
 150 approach is approximated by a modification to the CFL condition for shallow water models that  
 151 neglects the velocity component as shown in Table 2.  
 152

153 LISFLOOD-ATS is a one-dimensional approximation of a diffusion wave based on uniform flow  
 154 formula, which are decoupled in x and y directions (Bates and de Roo, 2000). Manning's equation, as  
 155 shown below, is implemented explicitly on a raster grid as described in detail by Hunter et al.  
 156 (2005a):

$$Q^{t+\Delta t} = \frac{(h_{flow}^t)^{5/3}}{n} \left( \frac{\Delta(h^t + z)}{\Delta x} \right)^{1/2} \Delta x$$

157  
 158 LISFLOOD-ATS should be very cheap to solve as it has the simplest physical representation, however  
 159 its stable time-step has been shown to be significantly smaller than that determined by the CFL  
 160 condition (Bates et al., 2010; Hunter et al., 2005) and is calculated by the equation in Table 2. This  
 161 equation includes the water surface slope which causes the time-step will go to zero with a flat

162 water surface, meaning a linearization of the time-step equation is needed at low slope to prevent  
163 the scheme stalling. The conditions under which this linearization is implemented are also  
164 summarised in Table 2.

165 As discharge between cells is calculated across each cell face, the continuity equation for the three  
166 models sums the flows across each face of every cell in the model and then multiplies by the time  
167 step to calculate a volume change, before dividing by the cell area to calculate a depth change for  
168 the cell.

169 Where available, simulation results will be presented from a range of industrial codes that were  
170 benchmarked on the test cases considered here as part of a recent Environment Agency  
171 benchmarking exercise (Néelz and Pender 2010). The EA benchmarking study included 14 models;  
172 however results from only six commercial programs will be presented here. The aim being to provide  
173 an industrial benchmark for the LISFLOOD-FP results rather than a comparison of all models. Of the  
174 six programs selected, four are full shallow water models broadly similar to LISFLOOD-Roe (TUFLOW  
175 (Syme, 1991), ISIS2D (Liang et al., 2006; 2007), Infoworks2D (Lhomme et al., 2010), SOBEK (Stelling,  
176 1998)) and two are diffusive type models that are similar to LISFLOOD-ATS (JFLOW-GPU (Bradbrook  
177 et al., 2004; Lamb et al., 2010) and FlowRoute). There is no current industry implementation of the  
178 LISFLOOD-ACC algorithm.

### 179 3 Results

180 These results summarise findings from four of the ten EA test cases listed in Table 1. The reasons for  
181 not implementing all tests are as follows:

- 182 • Tests 1&2 were ignored to save space and because later test were assumed to be more  
183 difficult and evaluate similar properties.
- 184 • Test 6a is a higher resolution (laboratory scale) and lower friction version of 6b. It was not  
185 practical to apply either of the simpler models to this test given the results from test 6b.
- 186 • Tests 7, 8a & 8b were deemed outside the scope of this paper because they require 1D  
187 channel, rainfall and sewer models, respectively.

188 Before discussing the results from each test, simulation times are presented in Table 3 based on  
189 single core implementations of the model. Some of the simulation time differences between models  
190 were due to variations in the inundation dynamics between codes, particularly for test cases 3&6b  
191 where there was a larger variation in the number of wet cells. Simulated dynamics were more alike  
192 in tests 4&5 (see subsequent results sections) meaning the difference in simulation time between  
193 the models provides indicative data on their relative speeds. For test 4, which has the longest  
194 simulation time, LISFLOOD-Roe was 3.3 times slower than LISFLOOD-ACC and 116.2 times faster than  
195 LISFLOOD-ATS. These differences were not unexpected. For LISFLOOD-ATS, the time-step is

196 proportional to  $\frac{1}{\Delta x^2}$  rather than  $\frac{1}{\Delta x}$  as is the case with the other models, whilst the inclusion of  
197 potentially very small water surface slopes in the time stepping equation (see Table 2) will further  
198 reduce the time-step relative to the other codes, especially if the computational grid is aligned with  
199 the water surface contours. Unlike LISFLOOD-ACC, LISFLOOD-Roe includes absolute velocity in the  
200 time stepping equation because advection is included in the model physics. Another indicator of  
201 potential computational efficiency is the relative number of non-integer power functions needed to

202 calculate flow in each model, these are nine (LISFLOOD-Roe), one (LISFLOOD-ACC) and two  
 203 (LISFLOOD-ATS). Thus, LISFLOOD-ACC requires a similar amount of computation per time step to  
 204 LISFLOOD-ATS and significantly less computation per time-step than LISFLOOD-Roe. The relative  
 205 simulation times vary between the test cases due to resolution, velocity, depth and slope factors,  
 206 although in all but one case the rank order of the models in terms of simulation time was consistent.  
 207 Note that all three LISFLOOD-FP models are explicit in time and that implicit schemes would allow  
 208 longer time-steps to be used.

209 Mass balance errors for each model simulation are summarised in Table 4 using the cumulative  
 210 volume error ( $m^3$ ) at the end of each simulation, which is the sum of volume errors made by the  
 211 model calculated at regular intervals through the simulation. Such that:

$$212 \quad V_e = V^N - V^0 + \sum_{i=1}^N \Delta t^i Q_{in}^i - \sum_{i=1}^N \Delta t^i Q_{out}^i$$

213 where the volume error  $V_e$  over a period of  $N$  time-steps is the initial domain volume  $V^0$  plus the  
 214 sum of inflow volumes  $\Delta t Q_{in}$  over each of the  $N$  time-steps minus the sum of outflow volumes from  
 215 the domain  $\Delta t Q_{out}$  minus the domain volume at the end of the  $N$  time-steps  $V^N$ . As all the test cases  
 216 used here have closed boundaries at the edge of the domain the outflow volumes were zero.  
 217 LISFLOOD-ATS tended to have the smallest volume errors and these were always several orders of  
 218 magnitude below 1% of the domain volume. LISFLOOD-ACC either had the smallest or greatest mass  
 219 error depending on the test case. Reasons for larger mass errors in LISFLOOD-ACC simulations of test  
 220 case 3&6b and LISFLOOD-Roe simulations of tests 4&5 are discussed in subsequent test case specific  
 221 sections.

### 222 **3.1 Momentum conservation over a bump**

223 This case is designed to test a code's ability to simulate flow down a slope and over a bump. It is test  
 224 case 3 in the EA benchmarking study (Table 1) and includes a two metre resolution DEM re-sampled  
 225 to five metres (Fig. 1) with a Manning's coefficient of 0.01. Water enters the domain along the entire  
 226 western edge of the DEM for 30 seconds at up to  $65 m^3 s^{-1}$  and then flows downhill into a depression,  
 227 which is just large enough to hold the total inflow volume. As the water accelerates down the slope  
 228 a portion of the volume should overtop a bump 200 m in from the western edge and flow into a  
 229 second depression. Diffusive models, like LISFLOOD-ATS are not expected to overtop the bump due  
 230 to lack of an inertial term, meaning all the water should pond in the first depression. LISFLOOD-Roe  
 231 is expected to simulate flow over the bump, while LISFLOOD-ACC will be unstable at the low friction  
 232 value required by the test.

233 The depth after 300 seconds of simulation by LISFLOOD-Roe is plotted in Fig.1 and indicates the  
 234 presence of water in both depressions, as expected from a full shallow water model. Time series  
 235 results from the LISFLOOD-FP models and industry shallow water models at control points 1&2 on  
 236 Fig. 1 are plotted in Fig. 2. In the case of LISFLOOD-ACC a higher Manning's  $n$  of 0.03 was used as the  
 237 model is unstable at the 0.01 required by this test case because the friction is used to stabilise the  
 238 scheme (see Bates et al., 2010 for a complete explanation).

239 The diffusive type model LISFLOOD-ATS behaved as expected, with no water overtopping the bump  
 240 due to the absence of inertia in the model, meaning the depression simply fills from the bottom as

241 water flows down the slope from the western edge. Mass errors were small ( $1.02 \times 10^{-12} \text{ m}^3$  from an  
242 inflow of  $1310 \text{ m}^3$ ) and the arrival time of the flood edge at CP1 was within two seconds of  
243 LISFLOOD-Roe. LISFLOOD-ATS velocity at CP1 initially rises at the same time as the shallow water  
244 models but then peaks early around 25% below the magnitude of the shallow water model before  
245 decreasing rapidly as the depression fills and the water surface at CP1 levels out. LISFLOOD-Roe and  
246 LISFLOOD-ACC both overtopped the bump, although a positive mass error in LISFLOOD-ACC of  $25.5$   
247  $\text{m}^3$ , compared to  $0.025 \text{ m}^3$  in LISFLOOD-Roe meant it predicted higher water levels than LISFLOOD-  
248 Roe at both control points. This indicates that although LISFLOOD-ACC simulated water moving over  
249 the bump, the model was not stable throughout this test case, leading to a positive mass balance  
250 error during the early part of the simulation as water accelerated down slope from the inflow. The  
251 arrival time of the wetting front was later in LISFLOOD-ACC due to the higher Manning's coefficient  
252 used, although the peak velocities and final depths were within the range simulated by the industry  
253 codes. LISFLOOD-Roe provided a smoother simulation of depth and velocity transitions than  
254 LISFLOOD-ACC and some of the industry shallow water models. The EA study (Néelz and Pender,  
255 2010) suggests that models with shock capturing capabilities provide less oscillatory solutions and  
256 the LISFLOOD-Roe results support this conclusion. Different approaches to re-sampling the two  
257 metre resolution DEM to five metres account for the 25% difference in the final depths at CP2  
258 between the shallow water models (ISIS2D is the model that simulates the same final depth as  
259 LISFLOOD-Roe), assuming that the mass errors in these models are not significantly greater than  
260 LISFLOOD-Roe. Nevertheless, LISFLOOD-Roe filled CP2 at a slower rate than the industry codes. The  
261 reason for this can be explained based on the difference between LISFLOOD-Roe and its closest  
262 industry equivalent InfoWorks2D. InfoWorks2D uses a semi-implicit or a dual time-stepping Runge-  
263 Kutta scheme whereas LISFLOOD-Roe is first-order in time and space, which adds numerical diffusion  
264 and means smoother peaks and slower propagation times as seen in this test. All the other industrial  
265 schemes for the complete shallow water equations are second order in either space or time or both.

### 266 **3.2 Rate of flood propagation over extended floodplains**

267 This test case comprises a flat, initially dry floodplain and a point source on the centre west edge of  
268 the domain. It is designed to test the ability of the model to simulate symmetrical flooding over an  
269 extended floodplain and was test 4 in the EA benchmarking study. All three LISFLOOD-FP codes are  
270 expected to simulate the level dynamics for this test. However, the EA study results found that  
271 simpler models were unable to simulate velocity. The simplicity of the topography means the  
272 industrial codes can be compared to the LISFLOOD-FP simulations with relative confidence, although  
273 the implementation of the inflow in the industry codes might differ from the cell centred varying  
274 head method used by LISFLOOD-FP. The test case comprises of a 1000 by 2000 m floodplain at 5 m  
275 resolution, a Manning's roughness coefficient of 0.05 and uses the 5 hour inflow hydrograph shown  
276 in Fig. 3 at a 20 m wide source on the centre west edge of the domain.

277 The top row of Fig. 4 plots a snapshot of simulated depths from the LISFLOOD-FP models and the six  
278 commercial codes three hours into the simulation, shortly before the shoreline reached the eastern  
279 closed boundary of the domain. This allowed the greatest time period for differences between the  
280 models to emerge. All the shallow water codes, including LISFLOOD-Roe, have semi-circular  
281 shorelines with no discernable preferential flow in any direction. The flood extents simulated by  
282 LISFLOOD-ACC were greater in the diagonal indicating a preferential flow in these directions.  
283 LISFLOOD-ATS simulated similar but less pronounced preferential flow in the diagonal. Of the  
284 commercial diffusive type codes, JFLOW-GPU which is coupled in x and y, simulated a semi-circular

285 shoreline with a slight preference for flow perpendicular to the grid, whilst FlowRoute (which is  
286 decoupled) simulates a remarkably similar preferential flow to LISFLOOD-ACC, despite using the  
287 uniform flow formula implemented by LISFLOOD-ATS. A number of tests were conducted to attempt  
288 to recreate the greater diagonal flow simulated by FlowRoute and LISFLOOD-ACC with LISFLOOD-  
289 ATS, including changing the wetting and drying parameters and friction. However, using the fixed  
290 time-step of FlowRoute and removing the linearization of the LISFLOOD-ATS scheme for shallow  
291 water surface gradients (necessary to prevent the solution stalling in the adaptive time-stepping  
292 version) lead to the increased preferential flow in the diagonal seen in FlowRoute and LISFLOOD-  
293 ACC. These changes to the numerical scheme effectively return LISFLOOD-ATS to the version  
294 developed by Bates and De Roo (2000), which was subsequently upgraded by Hunter et al. (2005) to  
295 the version used throughout this paper.

296 Below the snapshots of simulated depth in Fig. 4 is a matrix plotting the differences between each of  
297 the simulations at this time. All the shallow water models differ from each other at the flood edge,  
298 presumably due to the wetting algorithm adopted. Away from the flood edge they are more alike  
299 with differences  $<0.005$  m rising up to 0.05 m within a few cells of the inflow. LISFLOOD-Roe was  
300 most like InfoWorks2D, which was not unexpected given that they both use Roe's approximate  
301 Riemann solver. Unlike all the other models, InfoWorks2D used an unstructured grid, indicating the  
302 choice of spatial discretisation had less effect on the outcome of this test case than the choice of  
303 numerical scheme, as would be expected over flat topography. JFLOW-GPU behaved in an almost  
304 opposite manner to LISFLOOD-ATS, with flow underestimated in the diagonal relative to the shallow  
305 water models. This led to greater depths (up to 0.025 m) 40 m diagonally from the source and lesser  
306 depths (up to 0.01 m) towards the flood edge in the diagonal. Although the LISFLOOD-ATS extents  
307 are similar to the shallow water models depths were also up to 0.01 m greater perpendicular to the  
308 inflow point. Perhaps the key point here is that all these differences are small relative to typical  
309 vertical errors in survey data and the accuracy required for strategic flood risk assessment.

310 Fig. 5 plots time series of depth and velocity at the four control points marked on the TUFLOW depth  
311 results in Fig. 4. To minimise confusion on the plots, simulations by the industry shallow water codes  
312 have been lumped into a single category, the interested reader is referred to the EA benchmarking  
313 study for a more detailed breakdown of these model results (Néelz and Pender, 2010). The industry  
314 shallow water codes and LISFLOOD-FP models simulated floodplain wetting to within 6 minutes of  
315 each other at the five points on the horizontal (CP1-4 or 1,3,5,6 in the EA study). On the diagonal  
316 (CP3 or 5 in the EA study) all LISFLOOD-FP and a number of the industrial models wetted at 60  
317 minutes ( $\pm 3$  minutes), although depth increased more rapidly over the next 20 minutes in the  
318 decoupled models. This more rapid increase in depth was reflected in the velocity simulations at this  
319 point, where velocity was 7.5% and 8.8% greater than LISFLOOD-Roe when simulated by LISFLOOD-  
320 ATS and LISFLOOD-ACC, respectively. The velocities on the horizontal were lower than LISFLOOD-Roe  
321 and the majority of the shallow water codes by a similar margin. Interestingly, LISFLOOD-ACC  
322 continued to simulate greater velocity on the diagonal for the remainder of the simulation, whilst  
323 the LISFLOOD-ATS velocities tended towards the shallow water codes then dropped below them  
324 after 175 minutes. This decrease in velocity was most noticeable in the depth simulations once the  
325 inflow hydrograph began to decrease at 250 minutes, demonstrating the effect of the LISFLOOD-ATS  
326 linearization at low slope (see Table 2). This is intuitively sensible since the inflow is driving the head  
327 change at the source and the water surface slope across the domain, which when shallow will  
328 initiate the linearization. A fixed time-step formulation or formulation without the linearization

329 would appear appealing on this basis, however as the water surface slope decreases towards zero  
330 the necessary time-step to avoid instability (checker boarding) will become infinitesimally small and  
331 computationally impractical.

332 Overall depths simulated by all the models were within 10% of each other, while inundation arrival  
333 times at CP 4 were spread over a <3 minute window after 60 minutes of simulation. In this test case  
334 it is not possible to pick out depth differences between the codes that can be attributed to the  
335 physical representation of the flow given the sensitivity to decoupling, linearization at low gradient,  
336 wetting method and the dominance of diffusion. At CP's 1-3 differences in peak velocity between  
337 the shallow water and simpler codes decreased with distance from the source where slopes and  
338 depths were lower, although the peak differences did not exceed 10%. It is worth noting that the  
339 differences between the industry models at CP1 were greater than the differences between the  
340 LISFLOOD-FP models, but also that the maximum velocities recorded in these plots were below a  
341 gentle  $0.5 \text{ ms}^{-1}$ . The next test case of a valley flooding following a dam failure represents a higher  
342 energy and less symmetrical test case.

### 343 **3.3 Valley flooding following dam failure**

344 This test case requires the models to simulate a valley flooding flood following a dam failure. For  
345 events of this type, flow depth, velocity and arrival time are all regarded as important factors for risk  
346 and hazard assessment because potentially dangerous velocities are expected. Given the EA results,  
347 LISFLOOD-Roe and LISFLOOD-ATS were expected to simulate maximum depths that were consistent  
348 with the industry codes, but the diffusive model was not expected to simulate velocity well due to a  
349 lack of inertia. LISFLOOD-ACC has not been tested on a case like this previously. The test case uses  
350 the hydrograph in Fig. 6 evenly spread over a 210 m inflow boundary (or 4 cells at 50 m resolution), a  
351 Manning roughness coefficient of 0.04 and a closed downstream boundary. A 50 m DEM was used in  
352 test case 5 in the EA benchmarking study. Here the LISFLOOD-FP simulations were also run at the 10  
353 m resolution of the best available DEM (Fig. 7) because significant simulation differences were  
354 observed at the higher resolution. Simulations from the industry codes are available at 50 m  
355 resolution from the EA study. However, the modellers in the EA study were asked to convert from  
356 the supplied 10 m resolution DEM to a 50 m resolution DEM and were given freedom to choose the  
357 lower left corner of the domain. This makes it difficult to perform a cell to cell overlay of the results  
358 in some cases and introduces topographic differences to the models (e.g. was the DEM re-sampled  
359 or smoothed to the 50 m resolution?). Therefore, flood extents from the industrial codes are not  
360 assessed for this test case, whilst the analysis of point time series should be interpreted with  
361 caution. This illustrates the need to implement different models in the same code in order to obtain  
362 sufficient experimental control in many benchmarking studies, especially as test cases become more  
363 complex.

364 Table 5 is a contingency table comparing binary inundation extents from the 50 m resolution  
365 LISFLOOD-FP models. LISFLOOD-Roe simulated a greater inundation extent than both of the simpler  
366 models, with an additional 153 (4.2% of wet/wet) and 156 (4.3% of wet/wet) cells inundated  
367 compared to LISFLOOD-ACC and LISFLOOD-ATS, respectively. The simpler models were more alike,  
368 with LISFLOOD-ACC simulating 3 (0.1% of wet/wet) additionally wet cells compared to LISFLOOD-  
369 ATS. The maximum flood depths simulated by the LISFLOOD-FP models are plotted on the top row of  
370 Fig. 8(i, ii, iii), with the differences between models below (iv, v, vi). At this resolution LISFLOOD-Roe  
371 was clearly affected by instability at the flood edge where it simulated depths up to 0.6 m greater

372 than the simpler models, which also accounts for most of the additional cells inundated by this  
373 model. Note however, that the increased depths at the flood edge were short lived and that the  
374 mass balance errors in the model are below 0.1% of the water volume. Nevertheless, the mass  
375 errors for LISFLOOD-Roe ( $-8 \times 10^3$  Table 4) were significantly higher than those of LISFLOOD-ACC  
376 ( $8 \times 10^{-9}$ ) and LISFLOOD-ATS ( $4 \times 10^{-7}$ ). The mass errors from the simpler models are essentially a  
377 reflection of numerical precision of the continuity equation, with the mass error from LISFLOOD-ATS  
378 being greater because it need 194 times as many time-steps as LISFLOOD-ACC.

379 Water surface elevation dynamics were recorded at the six control points in Fig. 7. These points  
380 were also used by the industrial codes in the EA benchmarking study. Fig. 9 plots simulated water  
381 surface elevation over time, whilst Fig. 10 plots the corresponding velocity. LISFLOOD-Roe simulated  
382 a later arrival time of the wetting front and a slower increase in water depth than the industrial  
383 codes. Although the slower increase in depth was also seen at CP2 in the flow over a bump test case,  
384 the differences to the shallow water models were larger as the travel distance is larger too.  
385 LISFLOOD-Roe had difficulty simulating the wet/dry edges at this resolution despite reducing the  $\alpha$   
386 coefficient in the time-step equation (Table 1) to 0.3 for this case. Furthermore, the way inflow to  
387 the domain is handled (simply changing head in the inflow cell) may not being adequate in this case.

388 Despite the timing issues, peak velocities for all LISFLOOD-FP models were always within the range  
389 simulated by the industry codes, while peak levels were within the range at CP's 1,3,4&6 and <10%  
390 lower at CP's 2&5. LISFLOOD-ATS and ACC simulated lower water surface elevations than the  
391 industrial codes at 50 m resolution, except at the bottom of the reach where water ponds due to the  
392 closed downstream boundary. Although arrival times were within 15 minutes of the shallow water  
393 models, the rate of rise in water surface elevation was consistently quicker (as seen in the previous  
394 test case) and the discrepancy between the models increased with distance downstream, this could  
395 potentially indicate greater numerical diffusion in the model which simulated smoother depth  
396 increases. As mass balance errors were insignificant, the higher rate of water level rise tended to  
397 result in greater peak velocities. Peak velocities for LISFLOOD-ACC and -ATS were within the shallow  
398 water model estimates at CP 5, <10% greater at CP's 1&4 and <20% greater at CP's 1, 2&6 at 50 m  
399 resolution. For this test case, the EA benchmarking study found that the diffusive type models  
400 produced oscillatory estimates of velocity (not shown here) that were sometimes over 100%  
401 different from the shallow water model simulations at points 4 and 5. However, this was not the  
402 case with the LISFLOOD-ATS because velocity simulations were within 20% of the shallow water  
403 models. Therefore, the industry diffusive models failed this test due to some unreported aspect of  
404 their implementation rather than the lack of flow process representation in the diffusive type model.

405 Velocity and water surface elevation data were recorded at 1 minute intervals by the industry  
406 models, so the same convention was adopted here. To evaluate the sensitivity to how frequently  
407 results were recorded the sampling rate was increased to 5 seconds. This increased water surface  
408 elevation by at most 0.003 m at CP3 but had a greater effect on velocity with peak values increasing  
409 by up to  $0.189 \text{ ms}^{-1}$  (8.9%) at CP1 and CP3. This temporal resolution effect is significant when  
410 comparing peak velocities from the models at these two control points because it is of similar  
411 magnitude to the differences between models.

412 The differences in water surface elevations at the beginning of the simulation reflect the differences  
413 in DEM elevation between the models (e.g. dry bed) that result from allowing the modeller to decide

414 how to convert from a 10 m to 50 m resolution DEM. These bed elevation differences were  
415 sometimes over 50% of the differences between the model simulations (See CP 4 in particular).  
416 Therefore, before examining the 10 m results in detail, a quick test was implemented to estimate the  
417 magnitude of the resolution effect on model simulations relative to the differences between model  
418 formulations at 50 m resolution. For this experiment the 10 m resolution DEM was re-sampled to 20,  
419 40, 50, 60, 80 & 100 m resolutions using a nearest neighbour approach. These DEM's were then used  
420 for simulation by the LISFLOOD-ACC model, as this was the most scalable model formulation in  
421 terms of computation time and model stability. Fig. 11 plots the effect of model and DEM resolution  
422 change on peak water surface elevations and velocities as well as the timings of velocity peaks. Each  
423 block of bars is one of six control points from Fig. 7, with the individual bars in each block  
424 representing the different DEM resolutions from 10 m (left) to 100 m (right). The affect of resolution  
425 on maximum water surface elevation (Fig. 11a) was up to 20 cm or 5% of the depth, with the  
426 greatest difference at control point 5. For velocities the affect of resolution was up to  $0.612 \text{ m}^3\text{s}^{-1}$  or  
427 20% of the velocity at control point 3 (Fig. 11c). The changes in velocity with resolution have both a  
428 random component due to alterations in flow pathways with resolution and a systematic decrease in  
429 wave speed with courser resolution, which is better represented by the up to 25 minute changes in  
430 peak velocity arrival times (Fig. 11d). This is a rather simple exploration of the model sensitivity to  
431 DEM resolution and does not separate any scalability issues with the model formulation from affects  
432 of changing topography, while assuming the 10 m DEM is error free. However, the various  
433 treatments of the DEM in this paper and by the industry models demonstrate that for this test case  
434 the resolution and sampling of the topography had a similar or greater magnitude effect on model  
435 simulations than model formulation, highlighting the importance of floodplain topography, as  
436 demonstrated by numerous studies (Fewtrell, 2008; Yu and Lane, 2006; Sanders 2007; Wilson and  
437 Atkinson 2007). Therefore, although the diffusive type models were less able to simulate the  
438 hydraulics over the transitions in slope seen on this reach, the simulations of hazard were similar  
439 given the sensitivity to factors such as sampling intervals and DEM treatment.

440 For the 10 m resolution test the relative behaviour of the three models changed. LISFLOOD-Roe and  
441 LISFLOOD-ACC simulations of depth and velocity were more alike than LISFLOOD-ACC and ATS (see  
442 plots of maximum depths in Fig. 8 and time series data in Fig. 9&10), with both converging towards  
443 the results obtained from the 50 m resolution industry shallow water models. At 50 m resolution  
444 differences between the LISFLOOD-ACC and LISFLOOD-ATS simulations of maximum depth were an  
445 order of magnitude smaller than the differences between these simpler models and LISFLOOD-Roe,  
446 with differences  $<0.01$  m in the area where water ponds at the bottom of the reach (northeast  
447 corner). However, at 10 m resolution LISFLOOD-ATS under predicted the depths from the other two  
448 models, with a difference in maximum water surface elevation of  $<0.03$  m, while also simulating  
449 depths and velocities within a few percent of the 50 m resolution simulation from this model.  
450 Therefore, the increase in resolution to 10 m has caused the LISFLOOD-ACC model to behave more  
451 like a full shallow water model, whereas at 50 m resolution it behaved in a similar manner to the  
452 diffusive model. The LISFLOOD-Roe 10m simulations fall within the range of levels and velocities  
453 simulated by the 50 m resolution industry models, while the mass errors (Table 4) have decreased.  
454 Furthermore, the maximum depth plots in Fig. 12 show no evidence of the instability at wet/dry  
455 edges seen at 50 m resolution.

456 At 10 m resolution the greatest differences between the LISFLOOD-Roe and LISFLOOD-ACC models  
457 occurred in areas of deep water at the base of steep slopes. Typically, the difference between the

458 models are  $<0.3$  m, however LISFLOOD-ACC over predicted LISFLOOD-Roe by up to 1.6 m for a  
459 roughly 500 m by 500 m region of deep water at the bottom of a slope close to the dam breach. This  
460 is a region where we find transitions from supercritical to subcritical flow so it is very unlikely to be  
461 simulated well by LISFLOOD-ACC. However, it is interesting that these locally large errors have not  
462 propagated down the valley, where the levels, velocities and timings are within the range simulated  
463 by the industry shallow water models. This is consistent with the finding of Hunter *et al.* (2005) that  
464 local hydraulic shocks do not necessarily impact on wave propagation and that where these do not  
465 dominate a test case or results in large mass balance errors it may still be possible to use a simplified  
466 model.

467 The differences between the models are further illustrated by the long section plots of bed elevation  
468 and maximum depth for the top 10,000 m of the reach in Fig. 12. Plot (a) on this figure shows the  
469 maximum water surface elevations for the three LISFLOOD-FP model simulations at 10 m resolution.  
470 As stated previously, the models are most alike on the steeper sections of the domain except at  
471 1,500 to 2,000 m where LISFLOOD-ACC over-predicted the other models. In areas of shallow  
472 gradient LISFLOOD-ATS under-predicted the other two models as noted in Fig. 8(xi & xii). Also  
473 plotted on this long section are the maximum velocity (c) and depth at the time of maximum velocity  
474 (b) from LISFLOOD-Roe. This was done to demonstrate that maximum depth was not coincident with  
475 maximum velocity on shallower sections of the model domain and also that LISFLOOD-Roe  
476 maximum velocities could sometimes occur during cell wetting due to the use of a momentum  
477 threshold that required a depth of flow between cells of 0.01 m before the flow equation was  
478 implemented (See Table 2). The implication for hazard estimation where the hazard is a product of  
479 both depth and velocity is that maximum simulated depth and velocity may only be appropriate for  
480 hazard estimation on the steeper sections of the domain, but are likely to overestimate hazard on  
481 sections with lower gradients. Calculating hazard at each time-step through a simulation and taking  
482 the maximum will be necessary in these locations and the method adopted for this is more  
483 significant in terms of resulting hazard than the choice of model for this test.

### 484 3.4 Dam break

485 This test case (EA test 6B) is designed to evaluate the code's ability to simulate hydraulic jumps and  
486 wake zones behind buildings and is a 20x scaled up version of the flume study by Soares-Frazao and  
487 Zech (2002). Only LISFLOOD-Roe is expected to give satisfactory simulations for this test due to the  
488 dominance of supercritical flow in this test case and lack of shock capturing capability in the other  
489 LISFLOOD-FP models. The domain comprises an 8 m deep, 135 m by 72 m reservoir that flows  
490 through a 20 m wide gate into a 72 m wide flume with a single building 68 m from the gate. The  
491 flume has an initial water depth of 0.4 m and a total length of 2020 m. Simulations were run for 300  
492 seconds from initially still water conditions with a Manning's coefficient of 0.05.

493 To illustrate the test case results, inundation depths simulated by LISFLOOD-Roe are plotted at 5  
494 second intervals for the first 30 seconds of the simulation in Fig. 13. The model performed as  
495 expected, with a hydraulic jump developing in front of the building from 15 seconds onwards and a  
496 wake zone behind the building from 20 seconds onwards. To compare the LISFLOOD-FP models and  
497 industry shallow water codes water level time-series were recorded at the six control points in Fig.  
498 14, although neither LISFLOOD-ATS or LISFLOOD-ACC were expected to simulate this test case  
499 adequately as they lack the necessary physics. LISFLOOD-ATS provided a smooth but inaccurate  
500 solution to the test without simulating the hydraulic jump, although the mass conservation was the

501 best of the three models. LISFLOOD-ACC was the least accurate of the LISFLOOD-FP models and had  
502 a 30% volume error because some flows between cells were sufficiently high to cause negative cell  
503 water depths when the continuity equation was implemented. This confirms, as expected, the  
504 unsuitability for this scheme for this test and situations where a significant proportion of the flow  
505 will be supercritical at times and in areas of interest. It is not clear from these results if this model  
506 failed primarily due to the lack of advection terms and/or because of the numerical solver used.  
507 However, advection will be necessary when velocities vary rapidly in time (e.g. transitional flows),  
508 while the results from the industry schemes and other LISFLOOD-FP models demonstrate, as  
509 expected, that a shock capturing shallow water model is necessary for these conditions.

510 Overall LISFLOOD-Roe provided a similar solution to the industry shallow water models. The  
511 simulated depth and velocity dynamics were smoother than the codes without shock capturing  
512 capabilities and most like those of InfoWorks2D, indicating the importance of the choice of shallow  
513 water solver in this test as discussed by Néelz and Pender (2010). This test has demonstrated that  
514 both simpler LISFLOOD-FP models should be avoided in situations where hydraulic jumps are  
515 expected to affect flood wave propagation.

## 516 **4 Discussion**

517 This paper has applied three versions of the LISFLOOD-FP model with different process  
518 representations to four test cases that were used for benchmarking industry standard two-  
519 dimensional model codes. Differences between the LISFLOOD-FP models were evaluated using a set  
520 of controlled tests that would have been difficult to implement without a universal code  
521 environment to manage the model state variables and parameters. This discussion will be structured  
522 in two parts, with the first part dealing with the results from the three models from each test and  
523 the degree of physical complexity needed to simulate inundation under versions scenarios, followed  
524 by a second section on the implication of these results on benchmarking best practice.

### 525 **4.1 How much physical complexity is required?**

526 For the test cases where flows were subcritical and varied gradually in time simulations of velocity,  
527 depth and inundation extent from the three models and the industry codes were broadly consistent  
528 with differences between models due to physical complexity often obscured by more subtle issues.  
529 The simulations of flood propagation over an extended floodplain provide an example of this  
530 problem for depth simulation because of the sensitivity to decoupling, linearization of the diffusive  
531 model at low slopes and to a lesser extent wetting and drying parameters. Despite these factors,  
532 depths simulated by industry shallow water and the three LISFLOOD-FP models were within 10% of  
533 each other for this test, while inundation arrival times were spread over a <6 minute, but often <3  
534 minute, window after up to 60 minutes of simulation. The velocity dynamics showed more variation  
535 between the codes, with the two simpler LISFLOOD-FP models, where the flow equations are  
536 decoupled in x and y, tending to under-predict the LISFLOOD-Roe velocity when aligned with the grid  
537 and over-predict on the diagonal, except when the time-step linearization takes effect in LISFLOOD-  
538 ATS. Although this is a limitation, being unable to simulate symmetry was not an obvious problem in  
539 the real world test cases and given the results from JFLOW GPU not a problem that relates to  
540 physical complexity. Nevertheless, if symmetry is essential then decoupled schemes should be  
541 avoided. An ability to simulate symmetry may thus be a theoretically interesting property for a  
542 hydraulic model, but one which may not have great practical relevance.

543 For the valley flooding following dam failure at 50 m resolution, maximum simulated depths were  
544 lower in LISFLOOD-ATS and -ACC, although the sensitivity to subtle choices over how to sample the  
545 topography from the 10 m DEM and the grid resolution of the model were as important in  
546 determining local variations in depth and velocity. LISFLOOD-Roe simulated later arrival times and  
547 slower increases in water levels than the other industry shallow water models at 50 m resolution,  
548 indicating the model had too much numerical diffusion at this scale, while being unable to simulate  
549 the wet/dry edges in a satisfactory manner. At 10 m resolution simulations from LISFLOOD-ACC and  
550 LISFLOOD-Roe were within the range of industry codes. Further work to improve the scalability of  
551 the codes, particularly LISFLOOD-Roe, when applied to this test is needed. For LISFLOOD-ATS the  
552 absence of inertia was evident around the regular transitions in slope along the reach. In percentage  
553 terms, the consistency in velocity simulation was similar to the consistency in simulation of depth,  
554 although velocity was more sensitive to local DEM changes than depth which made this variable  
555 difficult to compare with the industry codes due to uncertainties in topographic sampling.

556 Simulation times varied dramatically between test cases, although the LISFLOOD-FP model with the  
557 simplest physical representation (LISFLOOD-ATS) required the longest simulation time by between  
558 one and two orders of magnitude for all tests. The simulation times of LISFLOOD-Roe were  
559 consistent with the quicker explicit industry shallow water models in Néelz and Pender (2010), but  
560 the inertial model LISFLOOD-ACC was 3.2 times faster than LISFLOOD-Roe for the flooding over an  
561 extended floodplain. The flow over an extended floodplain test provides the most rigorous  
562 comparison of simulation times here because simulated depths and inundation extents were more  
563 consistent between the models in this test than the others. The simulation times for LISFLOOD-ATS  
564 are likely to seriously limit its suitability for large area, fine resolution or Monte Carlo type studies,  
565 even when the simulations are considered to be accurate enough for the task. Although not  
566 reported here all the models were tested with inappropriately long time-steps and found to be  
567 inaccurate, particularly in terms of timings and velocities meaning this should be avoided. This is  
568 especially relevant in the case of LISFLOOD-ATS where a similar time-step to that used in the models  
569 with inertia will lead to inaccurate simulation. The high computational cost of LISFLOOD-ATS means  
570 it is tempting to use an adaptive time-step similar to the shallow water models, whilst implementing  
571 a flow limiter to prevent the solution from oscillating. Hunter et al. (2005) provide a description of  
572 this approach, however the flow limiter should be avoided because it leads to a significant  
573 deterioration in the quality of simulated wave propagation.

574 At this point it is useful to discuss the explicit diffusive model results within the historical context of  
575 their use. These models were critical in highlighting the advantages of 2D modelling of floodplain  
576 flows over 1D approaches (Horritt and Bates, 2002), while bringing flood simulation to a wider  
577 audience by being relatively easy to code, understand and visualise. Furthermore, the results here  
578 and elsewhere (e.g. Yu and Lane, 2006; Tayefi et al., 2007; Hunter et al., 2008; Neal et al., 2009a)  
579 indicate that the inundation extents and depths typical of previous mapping work with these models  
580 would not change markedly if they were re-calculated using a more complex methodology, at least  
581 for sites where flows vary gradually and model time-steps were appropriate. Explicit diffusive model  
582 also benefit from being simple, however any perception from previous work that this simplicity leads  
583 to relative computational efficiency should be rejected in almost all cases. Thus, in an operational  
584 context the approaches available for inundation simulation have moved on from the LISFLOOD-ATS  
585 type formulation.

586 The flow over a bump and dam break test cases require the simulation of conditions that were  
587 expected to challenge the two simpler LISFLOOD-FP formulations. Only LISFLOOD-Roe was able to  
588 simulate the flow over the bump test case correctly. LISFLOOD-ACC could also simulate water  
589 overtopping the bump but only by increasing the roughness, while LISFLOOD-ATS did not overtop  
590 the bump as would be expected for a model which lacks inertia. Thus the diffusive and shallow water  
591 model results were consistent with the EA study, while the LISFLOOD-ACC results indicate that this  
592 model may be suitable for similar test cases where flows are subcritical and friction is greater than  
593  $n=0.03$ . Developments to this scheme for urban applications should focus on methods to maintain  
594 stability at low friction without compromising on speed, or the development of hybrid models where  
595 the numerical scheme adapts to the flow conditions.

596 LISFLOOD-Roe simulated similar dynamics to the other shock capturing shallow water codes for the  
597 dam break test case. LISFLOOD-ATS and LISFLOOD-ACC were unable to simulate the hydraulic jump  
598 as expected and should not be used if such features are essential to the simulation, i.e. where the  
599 influence of the shocks extends away from their local vicinity and affects wave propagation globally  
600 in the model. Where this occurs appears easy to identify for LISFLOOD-ACC as in every such case  
601 examined here the mass balance errors from the model become unacceptably large (see Table 3).  
602 Hence when applying the LISFLOOD-ACC model to test cases where it was unable to emulate the full  
603 shallow water models depths and velocities to within  $\sim 10\%$ , its mass balance error increased by  
604 many orders of magnitude. If we conclude that the model is applicable to a smaller range of  
605 scenarios than the full shallow water models, then mass balance would appear to be a good proxy  
606 for determining appropriateness and should thus always be reported. Furthermore, although  
607 LISFLOOD-ATS was unable to simulate key aspects of the flow over a bump and the dam break the  
608 model remained stable and conserved mass for all the tests undertaken here unlike the other two  
609 models.

#### 610 **4.2 Implications for inundation model benchmarking**

611 The previous discussion on model complexity highlights the difficulty of benchmarking complex  
612 models, where aspects of model setup that might usually be considered as minor can obscure the  
613 headline differences between models such as the type of solver used or physical complexity.  
614 Benchmarking is undoubtedly made easier by models that share common sub-routines and input  
615 data, such as the three used here, but as this is not a practical solution for industry models. The tests  
616 conducted in the EA study established the magnitude of differences between models given a  
617 number of test cases, which allowed model responses to be classified and approaches that simulated  
618 non-behavioural dynamics to be identified. In terms of model suitability for various applications, the  
619 finding here support those of the EA model benchmarking study (Néelz and Pender, 2010), except  
620 that the performance of LISFLOOD-ATS for velocity simulation was significantly better than the  
621 industry equivalents of this code and there was no industry implementation of LISFLOOD-ACC. An  
622 important question is how significant the choice of model is in relation to other factors, including  
623 both controllable model setup decisions (e.g. resolution, mesh type and the frequency with which  
624 results are recorded) and model uncertainties (e.g. possible input flow data and DEM errors).

625 The valley filling test provides a convenient example of how simple model setup decisions can have  
626 as much impact on hazard mapping as choosing between the three LISFLOOD-FP models. Peak  
627 velocities at each control point were short-lived to the extent that increasing the rate at which  
628 velocity was recorded from 1 minute to 5 seconds increased peak velocity by up to 8.9 % at selected

629 control points. This has significant implications for risk assessment because methods that take  
630 infrequent snapshots of model state variable may not capture maximum velocities. Furthermore,  
631 maximum velocity and depth were broadly coincident in time on the steeper sections of the domain,  
632 with maximum depth occurring some time after maximum velocity on the shallower sections. The  
633 implication of this for hazard estimation is that the product of maximum simulated depth and  
634 maximum velocity may overestimate hazard in particular locations if these are determined  
635 separately. Also, since peak velocity is short in duration, hazard will also change rapidly.

636 In addition to model setup issues that can be controlled, the significance of model choice in relation  
637 to the principal sources of uncertainty would be a useful addition to future benchmarking work.  
638 Here it was relative simple to demonstrate that simulations of the valley flooding event were as  
639 sensitive to the sampling of the 10 m topography to coarser resolutions as they were to model  
640 choice, and that the model had not converged on a grid-independent estimate of velocity by 10 m.  
641 However, this should go further in future benchmarking work by evaluating the choice of model  
642 given uncertainty in the elevation and inflow data typically used for the applications being tested.

643

## 644 5 Conclusions

645 Three two-dimensional hydraulic models with different physical representations have been  
646 benchmarked using four test cases. Well known factors such as topography were found to influence  
647 simulations, but a number of less obvious factors also cause differences in simulations as great or  
648 greater than physical complexity. A number of specific conclusions can be made:

- 649 1) Explicit diffusive type model required much longer simulation times than the models with  
650 inertia for the 2-50 m resolution applications considered here. This problem cannot be  
651 solved by using fixed longer time steps and a flow limiter because of the poor simulation of  
652 wave propagation with such methods.
- 653 2) Decoupled schemes were unable to simulate symmetry over flat topography, although  
654 similar effects on the irregular topographies tested here were not identifiable given other  
655 factors. The simulation of symmetry is therefore an interesting technical test but may have  
656 limited relevance to real world flows.
- 657 3) For test cases with gradually varying flows, simulations of velocity were surprisingly similar  
658 between the codes and usually as alike as depth in terms of % difference. This means that  
659 the simplified models may be appropriate for velocity simulation for a wider range of  
660 conditions than suggested by the EA study where gradually varied subcritical flows are  
661 expected.
- 662 4) For test cases where flows change gradually with time the difference between models are  
663 only as large as differences caused by other modelling choices (e.g. topography sampling,  
664 recording of results etc).
- 665 5) The diffusive model LISFLOOD-ATS was the least like the other codes in the vicinity of slope  
666 transitions for the valley flooding following a dam failure test case. The momentum  
667 conservation over a bump test case demonstrates how the diffusive model loses momentum  
668 too quickly when the DEM slope decreases.
- 669 6) LISFLOOD-Roe as a pure first-order in time and space scheme simulated later arrival times  
670 and slower increases in water levels than the other shallow water models at the 50 m

671 resolution version of the valley flooding test, and the CFL number had to be reduced in order  
672 to simulate the wet/dry edges in a stable manner at 10 m resolution.

- 673 7) Simple decisions over arbitrary modelling choices, such as how frequently to record velocity  
674 and assumptions about depth velocity correlations, can have greater impacts on hazard  
675 assessment than decisions over model physical complexity.
- 676 8) The simpler models were unable to simulate hydraulic jumps and wake zones, as expected.
- 677 9) Rigorous control in benchmarking studies is difficult to achieve, especially when undertaken  
678 with multiple codes and modellers.
- 679 10) LISFLOOD-Roe was required when there were subcritical to supercritical transitions in the  
680 flow that affect the wave propagation, but unless contra-indicated by large mass errors  
681 LISFLOOD-ACC was a faster alternative to a full shallow water model for gradually varied  
682 subcritical flows where domain-average friction typically exceeds  $n=0.03$ .

683 LISFLOOD-Roe was applicable to the widest range of flow conditions, although it was inaccurate at  
684 the flood edge when adjacent velocity was high and grid resolution coarse. LISFLOOD-ACC was  
685 usually the quickest model, which would be particularly advantageous for real-time inundation  
686 forecasting, Monte Carlo type analysis or large model applications. However, for conditions where  
687 this code was not physically suitable (e.g. low friction, Froude number  $> 1$ ) the model became  
688 unstable and mass balance errors became large. For LISFLOOD-ATS, the ease of use, simplicity,  
689 stability and small mass errors may be desirable where the model is applied to cases where it is  
690 difficult to check model results, where only diffusive process representation is required and where  
691 coarse resolution models are needed. However, the computational cost will be relatively high for  
692 typical strategic flood risk management applications.

693 Through a rigorous series of benchmarks tests of 2D hydraulic models this paper is able to draw  
694 conclusions on the degree of physical complexity required to model flood inundation. We show that  
695 for gradually varied flow full shallow water models may be unnecessarily complex, and simpler,  
696 cheaper schemes, such as the inertial wave formulation in LISFLOOD-ACC, can perform just as well,  
697 both in terms of velocity and depths. Moreover we show that subtle modelling decisions can often  
698 have more effect on results than selecting a more physically complex model. The results of this  
699 study therefore provide additional guidance to help 2D model users select the appropriate scheme  
700 for any given situation.

## 701 **Acknowledgement**

702 Jeffrey Neal was funded by the Flood Risk Management Research Consortium, which is supported by  
703 grant number EP/F20511/1 from the EPSRC and the DEFRA/EA joint Research Program on Flood and  
704 Coastal Defence. We would like to thank the three anonymous reviewers for their very helpful  
705 comments and suggestion, along with Jim Walker and Sylvain Néelz for arranging access to the EA  
706 2D model benchmarking data sets.

## 707 **References**

708 Apel, H., Aronica, G. T., Kreibich, H. & Thielen, A. H. 2009. Flood risk analyses-how detailed do we  
709 need to be? *Natural Hazards*, 49, 79-98.

710 Aronica, G.T., Nasello, C., Tucciarelli, T. 1998. A 2D Multilevel Model for Flood Propagation in Flood  
711 Affected Areas. *ASCE - Journal of Water Resources Planning and Management*, Vol.124, 4.

712 Bates, P. D. & De Roo, A. P. J. 2000. A simple raster-based model for flood inundation simulation.  
713 *Journal of Hydrology*, 236, 54-77.

714 Bates, P. D., Horritt, M. S. & Fewtrell, T. J. 2010. An simple inertial formulation of the shallow water  
715 equations for efficient two dimensional flood inundation modelling. *Journal of Hydrology*, 387, 33-  
716 45.

717 Brufau, P., García-Navarro, P., and Vázquez-Cendón, M. E. 2002. "A numerical model for the flooding  
718 and drying of irregular domains." *Int. J. Numer. Methods Fluids*, 39, 247–275.

719 Bradbrook, K. F., Lane, S. N., Waller, S. G. & Bates, P. D. 2004. Two dimensional diffusion wave  
720 modelling of flood inundation using a simplified channel representation. *International Journal of*  
721 *River Basin Management*, 2, 1-13.

722 Crowder, RA., Pepper, AT., Whitlow, C., Sleigh, A Wright, N. and Tomlin C. 2004. Benchmarking of  
723 hydraulic river modelling software packages. SCHO0305BIXN-E-P. Environment Agency, Bristol.

724 Cobby, D.M., Mason, D.C., Davenport, I.J., 2001. Image processing of airborne scanning laser  
725 altimetry data for improved river flood modelling. *ISPRS Journal of Photogrammetry and Remote*  
726 *Sensing*, 56(2): 121-138.

727 Fewtrell, T.J., Bates, P.D., Horritt, M., Hunter, N.M., 2008. Evaluating the effect of scale in flood  
728 inundation modelling in urban environments. *Hydrological Processes*, 22(26): 5107-5118.

729 Gouldby, B., Sayers, P., Mulet-Marti, J., Hassan, M. A. A. M. & Benwell, D., 2008. A methodology for  
730 regional-scale flood risk assessment. *Proceedings of the Institution of Civil Engineers - Water*  
731 *Management*, 161, 169-182.

732 Guinot, V. & Soares-Frazae, O. S. 2006. Flux and source term discretization in two-dimensional  
733 shallow water models with porosity on unstructured grids. *International Journal for Numerical*  
734 *Methods in Fluids*, 50, 309-345.

735 Hall, J.W., Dawson, R.J., Sayers, P.B., Rosu, C., Chatterton, J.B. & Deakin, R. 2003. A methodology for  
736 national-scale flood risk assessment. *Proceedings of the Institution of Civil Engineers-Water and*  
737 *Maritime Engineering*, 156, 235-247.

738 Hall, J.W., Sayers, P.B. & Dawson, R.J., 2005. National-scale assessment of current and future flood  
739 risk in England and Wales. *Natural Hazards*, 36, 147-164.

740 Hall, J.W., Manning, L.J. & Hankin R.K.S., 2011. Bayesian calibration of a flood inundation model  
741 using spatial data. *Water Resources Research*, 47, W05529.

742 Horritt M. S., 2000. Calibration of a two-dimensional finite element flood flow model using satellite  
743 radar imagery. *Water Resources Research*, 36, No. 11, 3279–3291.

744 Horritt, M.S., Bates, P.D., 2001. Predicting floodplain inundation: raster-based modelling versus the  
745 finite-element approach. *Hydrological Processes*, 15(5): 825-842.

746 Horritt, M.S., Bates, P.D., 2002. Evaluation of 1D and 2D numerical models for predicting river flood  
747 inundation. *Journal of Hydrology*, 268(1-4): 87-99.

748 Horritt, M.S., Di Baldassarre, G., Bates, P.D., Brath, A., 2007. Comparing the performance of a 2-D  
749 finite element and a 2-D finite volume model of floodplain inundation using airborne SAR imagery.  
750 Hydrological Processes, 21: 2745-2759.

751 Hunter, N.M., Horritt, M.S., Bates, P.D., Wilson, M.D. and Werner, M.G.F., 2005a. An adaptive time  
752 step solution for raster-based storage cell modelling of floodplain inundation. Advances in Water  
753 Resources, 28, 975-991.

754 Hunter, N.M., Bates, P.D., Neelz, S., Pender, G., Villanueva, I., Wright, N.G., Liang, D., Falconer, R.A.,  
755 Lin, B., Waller, S., Crossley, A.J. and Mason, D.C., 2008. Benchmarking 2D hydraulic models for urban  
756 flooding. Proceedings of the Institution of Civil Engineers-Water Management, 161(1), 13-30.

757 Lamb, R., Crossley, A. & Waller, S., 2009. A fast 2d floodplain inundation model. Proceedings of the  
758 Institution of Civil Engineers - Water Management, 162, 363-370.

759 Leopardi A., Oliveri E. and Greco M., 2002. Two-dimensional modeling of floods to map risk-prone  
760 areas. ASCE Journal of Water Resources Planning and Management, 128, No. 3, 168–178.

761 Lhomme, J., Gutierrez-Andres, J., Weisgerber, A., Davison, M., Mulet-Marti, J., Cooper, A. And  
762 Gouldby, B., 2010. Testing a new two-dimensional flood modelling system: analytical tests and  
763 application to a flood event. Journal of Flood Risk Management 3(1): 33-51.

764 Liang, D., Falconer, R. A. & Lin, B. L., 2006. Comparison between tvd-maccormack and adi-type  
765 solvers of the shallow water equations. Advances in Water Resources, 29, 1833-1845.

766 Liang D., Lin B. and Falconer R. A., 2007. Simulation of rapidly varying flow using an efficient TVD-  
767 MacCormack scheme. International Journal for Numerical Methods in Fluids, 53(5), No. 3, 811–826.

768 McMillan, H. K. & Brasington, J., 2007. Reduced complexity strategies for modelling urban floodplain  
769 inundation. Geomorphology, 90, 226-243.

770 Mignot, E., Paquier, A. & Haider, S., 2006. Modeling floods in a dense urban area using 2d shallow  
771 water equations. Journal of Hydrology, 327, 186-199.

772 Neal, J., Bates, P., Fewtrell, T., Hunter, N. M., Wilson, M. & Horritt, M., 2009a. Distributed whole city  
773 water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic  
774 model simulations, Journal of Hydrology, 368, 42-55.

775 Neal, J. C., Fewtrell, T. J. & Trigg, M. A., 2009b. Parallelisation of storage cell flood models using  
776 openmp. Environmental Modelling and Software, 24, 872-877.

777 Néelz, S. & Pender, G. 2010. Benchmarking of 2D Hydraulic Modelling Packages. SC080035/SR2.  
778 Environment Agency, Bristol.

779 Roe, P. (1981). "Approximate Riemann Solvers, Parameter Vectors, and Difference-Schemes."  
780 Journal Of Computational Physics 43(2): 357-372.

781 Sanders, B. F., 2007. Evaluation of on-line dems for flood inundation modeling. Advances in Water  
782 Resources, 30, 1831-1843.

783 Schubert, J.E., Sanders, B.F., Smith, M.J., Wright, N.G., 2008. Unstructured mesh generation and  
784 landcover-based resistance for hydrodynamic modeling of urban flooding. *Advances in Water*  
785 *Resources*, 31(12): 1603-1621.

786 Soares-Fraza, S. and Zech, Y., 2002 Dam-break flow experiment: The isolated building test case.  
787 Available online at: [http://www.impact-project.net/wp3\\_technical.htm](http://www.impact-project.net/wp3_technical.htm)

788 Stelling, G.S., Kernkamp, H.W.J. & Laguzzi, M.M., 1998. Delft Flooding System: a powerful tool for  
789 inundation assessment based upon a positive flow simulation. *Hydroinformatics '98*, eds. Babovic  
790 and Larsen, Balkema: Rotterdam, 449-456.

791 Syme, W.J., 1991. Dynamically linked two-dimensional/one-dimensional hydrodynamic modelling  
792 program for rivers estuaries and coastal waters. MEngSc thesis, University of Queensland, Australia.

793 Tayefi, V., Lane, S.N., Hardy, R.J., Yu, D., 2007. A comparison of one- and two-dimensional  
794 approaches to modelling flood inundation over complex upland floodplains. *Hydrological Processes*,  
795 21(23): 3190-3202.

796 Trigg, M.A. et al., 2009. Amazon flood wave hydraulics. *Journal of Hydrology*, 374(1-2): 92-105.

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

799 Werner, M.G.F., 2004. A comparison of flood extent modelling approaches through constraining  
800 uncertainties on gauge data. *Hydrology and Earth System Sciences*, 8(6): 1141-1152.

801 Werner, M.G.F., Hunter, N.M., Bates, P.D., 2005. Identifiability of distributed floodplain roughness  
802 values in flood extent estimation. *Journal of Hydrology*, 314(1-4): 139-157.

803 Wilson, M.D., Atkinson, P.M., 2007. The use of remotely sensed land cover to derive floodplain  
804 friction coefficients for flood inundation modelling. *Hydrological Processes*, 21(26): 3576-3586.

805 Yu, D. & Lane, S. N., 2006. Urban fluvial flood modelling using a two-dimensional diffusion-wave  
806 treatment, part 2: Development of a sub-grid-scale treatment. *Hydrological Processes*, 20, 1567-  
807 1583.

## 809 **List of Figures**

- 810 Fig. 1: DEM and simulated depth after 900 seconds from LISFLOOD-Roe.
- 811 Fig. 2: Simulated levels and velocities at control points from LISFLOOD-FP models and industry  
812 shallow water models.
- 813 Fig 3: Inflow hydrograph for flow over an extended floodplain test.
- 814 Fig 4: Simulated depth and difference in depth after 3 hours for the LISFLOOD-FP and Industry codes.
- 815 Fig. 5: Simulated depth (left) and velocity (right) at the four control points in Fig. 4.
- 816 Fig. 6: Upstream inflow boundary for valley flooding test case.
- 817 Fig. 7: DEM at 10 m resolution for valley flooding test case. The plot includes the model inflow  
818 boundary, the long section used by Fig. 12 and the control points used by Fig's 9 & 10.
- 819 Fig. 8: Maximum simulated depth in meters for 50 m (a) & 10 m (b) resolution simulations of valley  
820 flooding test case. Differences between model maxima are also plotted.
- 821 Fig. 9: Simulated depth at six control points for valley flooding test case.
- 822 Fig. 10: Simulated scalar velocity at six control points for valley flooding test case.
- 823 Fig. 11: Plots of the effect of model and DEM resolution change on peak water surface elevations (a)  
824 and velocities (c) as well as the timings of max water surface elevation (b) and maximum velocity (d).  
825 Each block of bars is one of six control points from Fig. 7, with the individual bars in each block  
826 representing the different DEM resolutions from 10 m (left) to 100 m (right).
- 827 Fig. 12: Plots of maximum water surface elevation from the three models (a) and elevation at the  
828 time of maximum v for LISFLOOD-Roe (b) and maximum velocity for LISFLOOD-Roe for the 10,000m  
829 long section marked on Fig. 7.
- 830 Fig. 13: LISFLOOD-Roe evolution of depth during the first 30 seconds of simulation for the dam break  
831 test.
- 832 Fig. 14 Simulated water depths at control points 1-6 on Fig. 13 from the dam break test.

833

834 Table 1: Summary of test cases

EA test	Description	Tested here
1	Flooding a disconnected water body.	No
2	Filling of floodplain depressions.	No
3	Momentum conservation over a small (0.25m) obstruction.	Yes
4	Speed of flood propagation over an extended floodplain.	Yes
5	Valley flooding following a dam failure.	Yes + finer resolution
6a&b	Dam break. a) Flume scale, b) Field scale.	Yes, b only
7	River to floodplain linking.	No
8a&b	Urban flood. a) Rainfall, b) Rainfall and sewer surcharge.	No

835

837 Table 2: Model attributes

Model	ATS	ACC	Roe
Key reference	(Hunter et al., 2005)	(Bates et al., 2010)	(Villanueva and Wright, 2006)
Wave properties	Diffusive	Inertial shallow water	Full shallow water
Scheme	Finite difference (forward differences) explicit		Finite volume explicit
Mesh	Cartesian Grid with staggered $h$ and $Q$		
Solver	None – Analytical		Approximate Roe Riemann solver
Time stepping	$\Delta t_{max} = \frac{\Delta x^2}{4} \frac{2\pi}{h_{flow}} \left  \frac{\Delta h}{\Delta x} \right $	$\Delta t_{max} = \alpha \frac{\Delta x}{\sqrt{gh}}$	$\Delta t_{max} = \alpha \frac{\Delta x}{ v  + \sqrt{gh}}$
Mass conservation	$h_{i,j}^{t+\Delta t} = h_{i,j}^t + \Delta t \frac{Q_{x,i,j}^{t-\Delta t} - Q_{x,i,j}^t + Q_{y,i,j-1}^{t-\Delta t} - Q_{y,i,j}^t}{\Delta x^2}$		
Courant number ( $\alpha$ )	Effectively 1.0	0.7 unless stated	
Shock capturing	No	Yes	
Coupling in x and y	No - 1D across each cell edge		Yes
Linearization of scheme at low slope	if $\Delta h < \Delta x c$ where $c = 0.0002$	No	
Roughness	Manning's: Global or distributed (linear interpolation between cells). However, as the physics represented in the three models is different, the physical meaning of friction in the models differs.		
Pre-processing	Interpretation of test case was ensured to be identical between models because all access the same state variables.		
Topography ( $z$ )	Raster grid, depth of flow ( $h_{flow}$ ) defined as the difference between the highest water free surface in the two cells and the highest bed elevation.		
Domain boundary/ Internal boundary (wall)	Zero flux if cell elevation exceeds adjacent water surface elevation	Brufau et al. (2002)	
Point/boundary inflows	Head change only for these test cases.		
Wet/dry threshold	Depth threshold (0.001 m)		Depth threshold for wetting (0.001 m), additional threshold for momentum (0.01 m)
Numerical precision	All state variables and parameters are double precision.		
Executable	LISFLOOD-FP version 4.4.13 compiled with the 64-bit Intel C++ compiler for Linux version 10.1.015. Static executable using O3 compiler optimisation with OpenMP parallelisation (see Neal et al., 2009).		
Hardware	Linux operating system running on two quad-core 2.8 GHz Intel Xeon processor (E5462) with 6 MB cash each and 16 GB of RAM.		

Where  $x$  is distance,  $n$  is Manning's roughness coefficient,  $Q$  is flow rate,  $h$  is water depth,  $h_{flow}$  is the depth of water through which water can flow,  $g$  is acceleration due to gravity,  $\alpha$  is the Courant number,  $v$  is flow velocity and  $c$  is a typically small water depth threshold.

838 Table 3: Summary of test case simulation times (minutes).

	Test 3		Test 4		Test 5 – 50 m		Test 5 – 10 m		Test 6b	
	Simulation time	Number of time-steps	Simulation time	Number of time-steps	Simulation time	Number of time-steps	Simulation time	Number of time-steps	Simulation time	Number of time-steps
Roe	0.07	905	6.48	15291	2.55**	56211	302.19 <sup>+</sup> *	207487	4.67	18176
ACC	0.03	900	1.97	10551	0.68	21147	344.11 <sup>+</sup> **	260676	0.67	18001
ATS	1.52	82835	228.95	654581	161.13	4102887	6415.21 <sup>++</sup>	1.02x10 <sup>8</sup>	182.15	4819760

839 \*cfl number reduced to 0.5 for stability

840 \*\*cfl number reduced to 0.3 for stability

841 <sup>+</sup>run on two CPU cores

842 <sup>++</sup>run on eight CPU cores

843

844 Table 4: Summary of mass balance errors and mass balance errors as a percentage of input volume  
845 at the end of each model simulation. Note: water does not leave the domain in any of these test  
846 cases.

	Test 3	Test 4	Test 5 – 50 m	Test 5 – 10 m	Test 6b
Roe	$2.50 \times 10^{-2}$	$-2.00 \times 10^3$	$-8.81 \times 10^3$	$-7.05 \times 10^3$	$-2.19 \times 10^0$
ACC	$2.55 \times 10^1$	$1.85 \times 10^{-12}$	$8.01 \times 10^{-9}$	$2.47 \times 10^4$	$-3.68 \times 10^2$
ATS	$1.02 \times 10^{-12}$	$6.05 \times 10^{-9}$	$4.22 \times 10^{-7}$	$4.72 \times 10^3$	$-1.42 \times 10^{-7}$
Domain Volume	$1.31 \times 10^3$	$2.85 \times 10^5$	$9.44 \times 10^6$	$9.44 \times 10^6$	$1.29 \times 10^3$
Mass error as a percentage of domain volume					
Roe	0.002%	0.702%	0.093%	0.075%	0.170%
ACC	1.947%	<0.001%	<0.001%	0.262%	28.527%
ATS	<0.001%	<0.001%	<0.001%	0.050%	<0.001%

847

848

849 Table 5: Contingency tables for wet dry comparisons between LISFLOOD-FP models for 50 m  
850 resolution valley flooding test case

		LISFLOOD-Roe		LISFLOOD-ATS	
		Wet	Dry	Wet	Dry
LISFLOOD-ACC	Wet	3597	1	3595	3
	Dry	153	53624	0	53777
LISFLOOD-ATS	Wet	3594	1		
	Dry	156	53624		

851

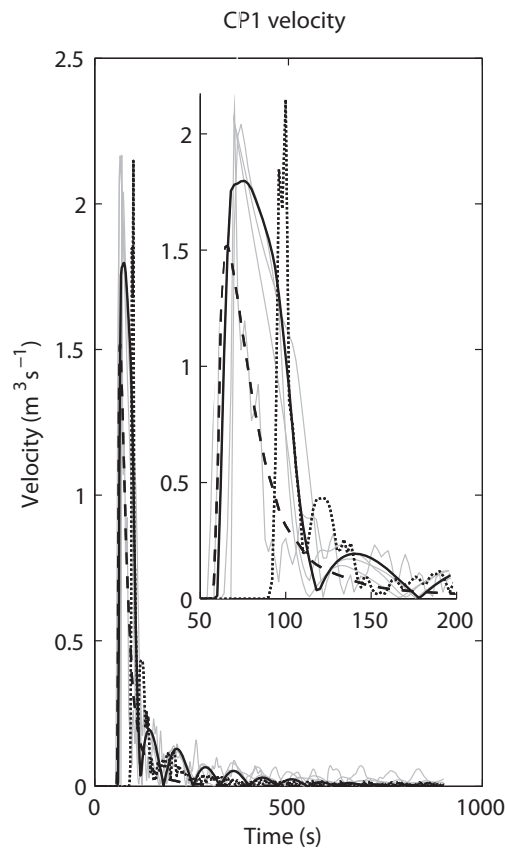
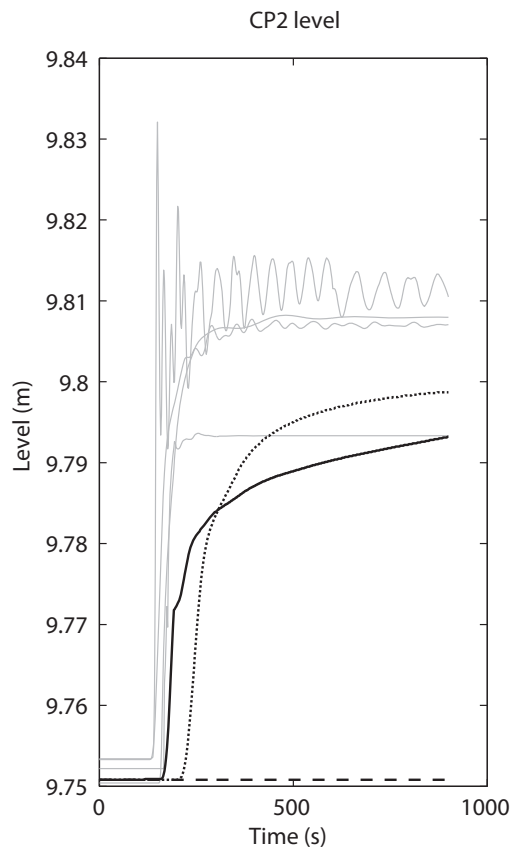
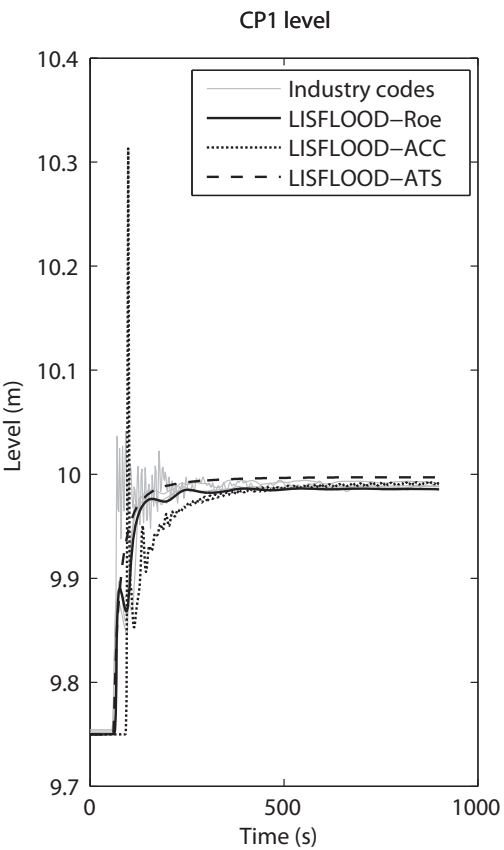
852

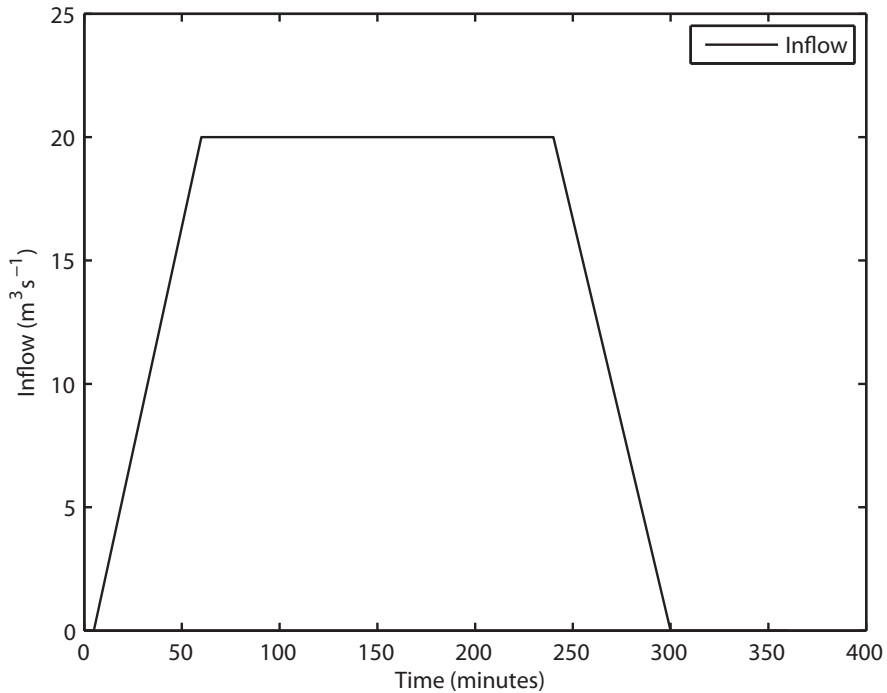
853

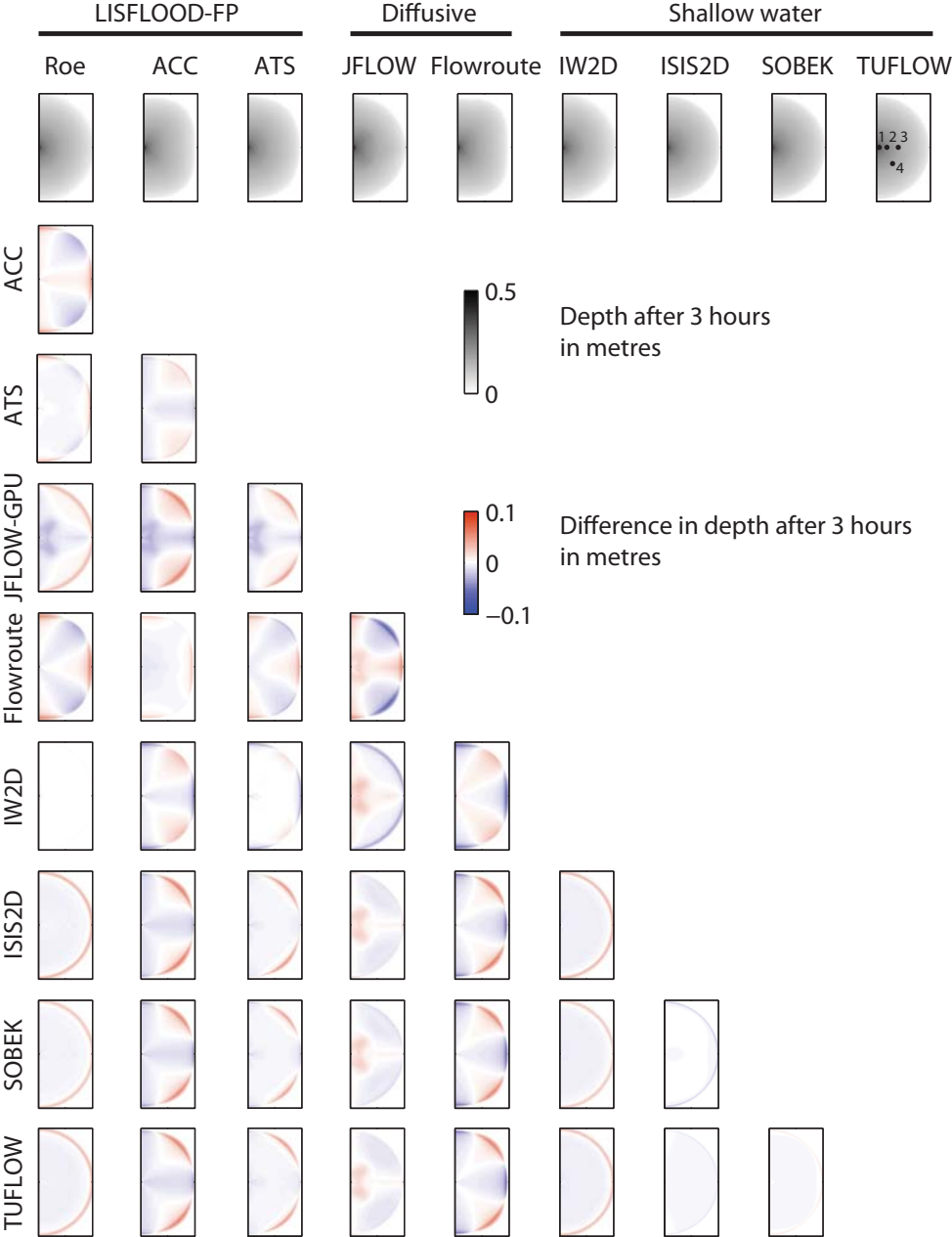
854

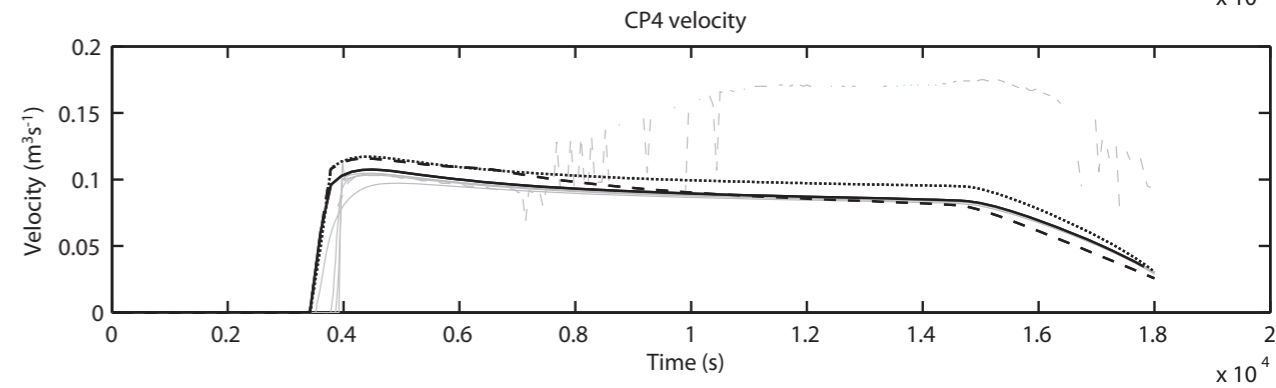
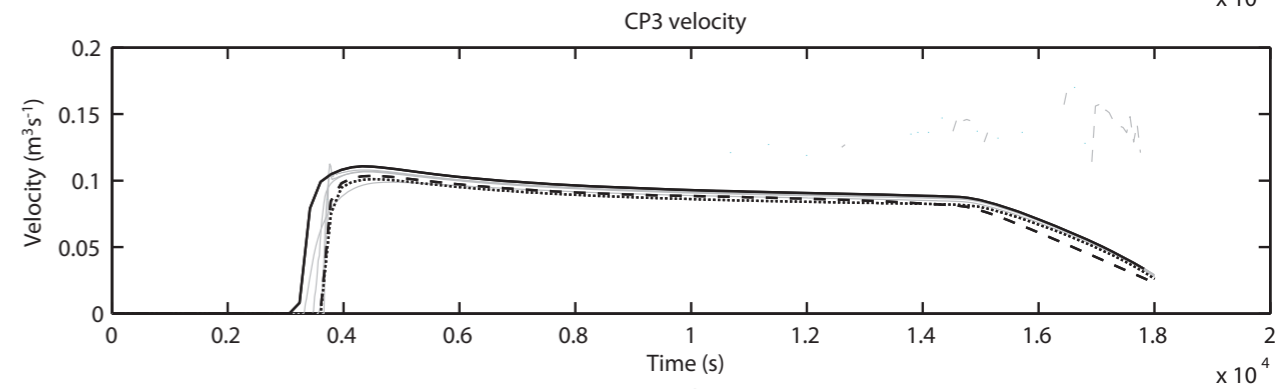
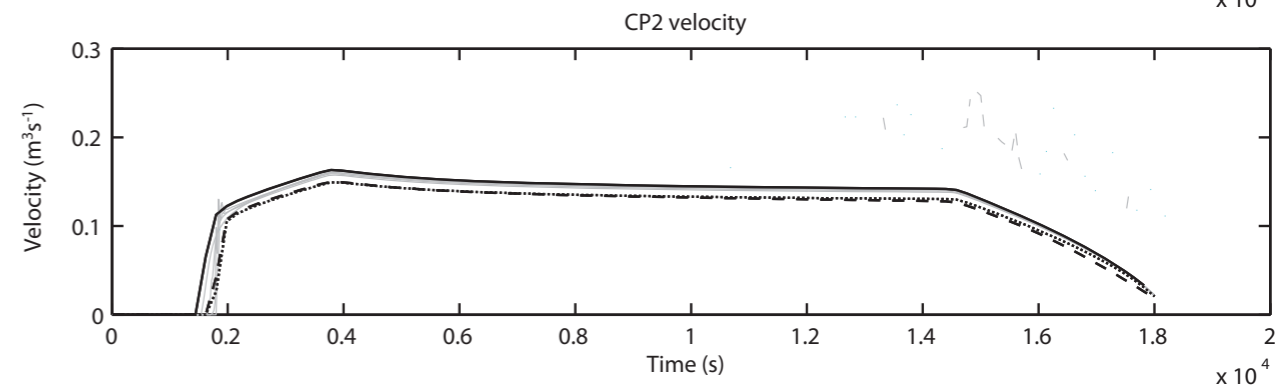
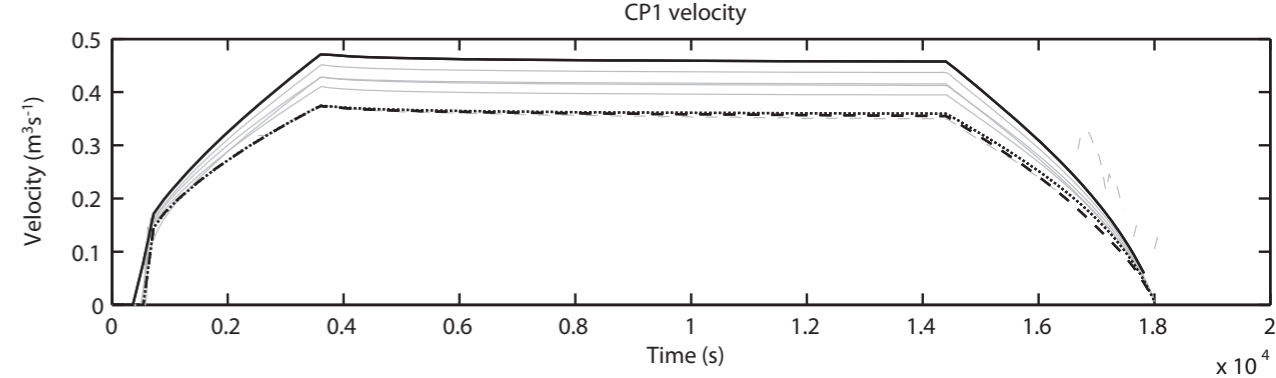
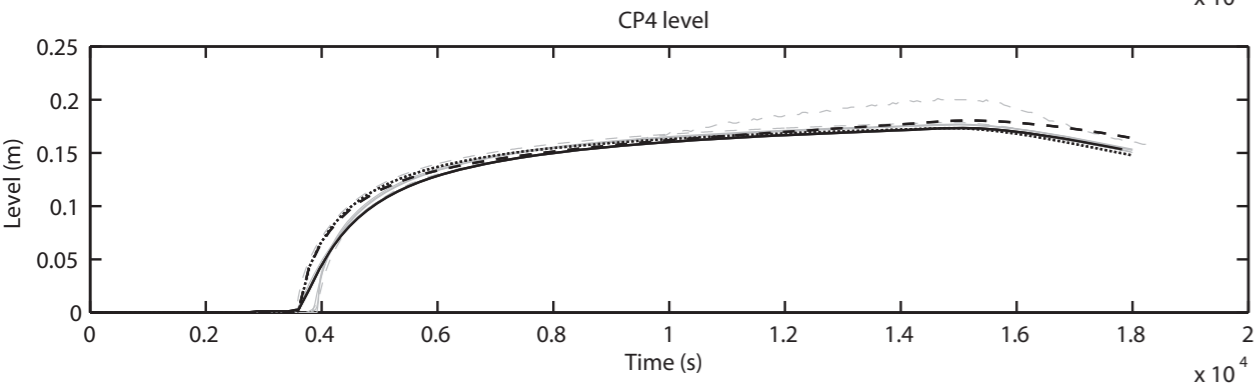
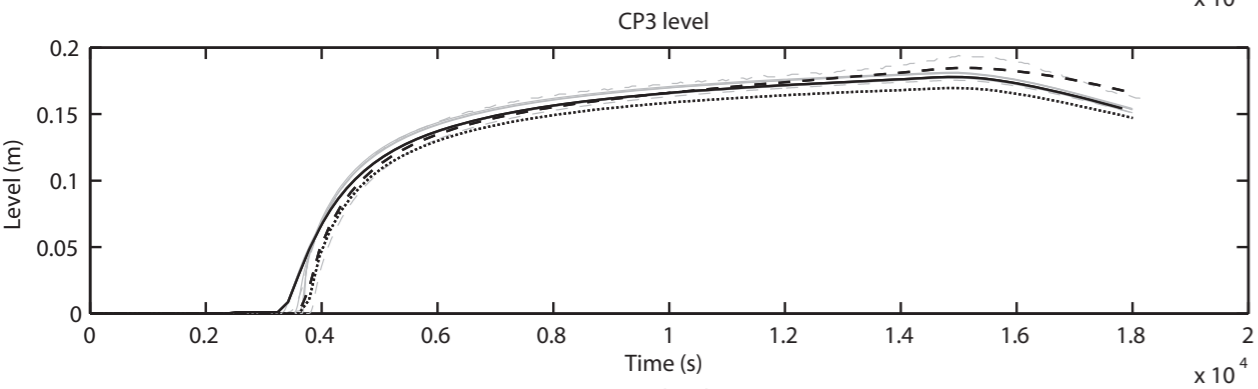
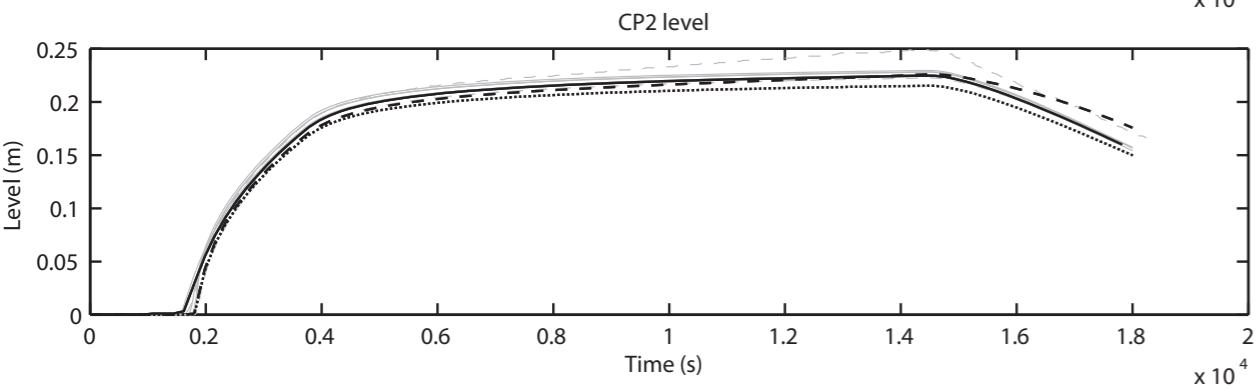
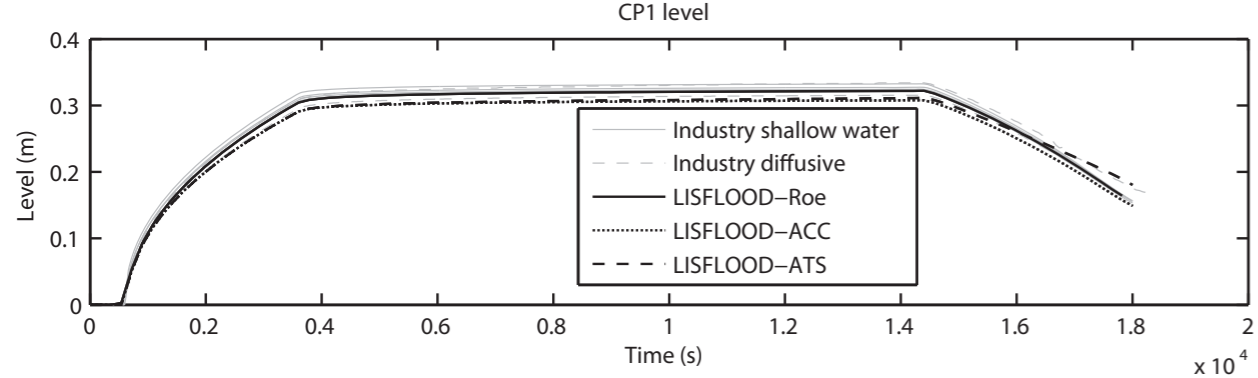
855

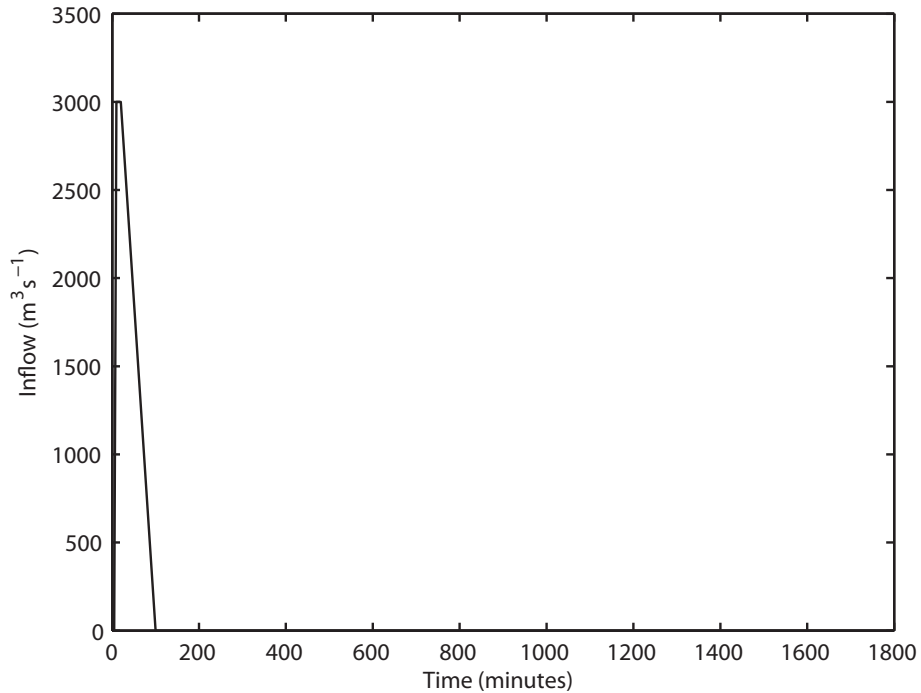










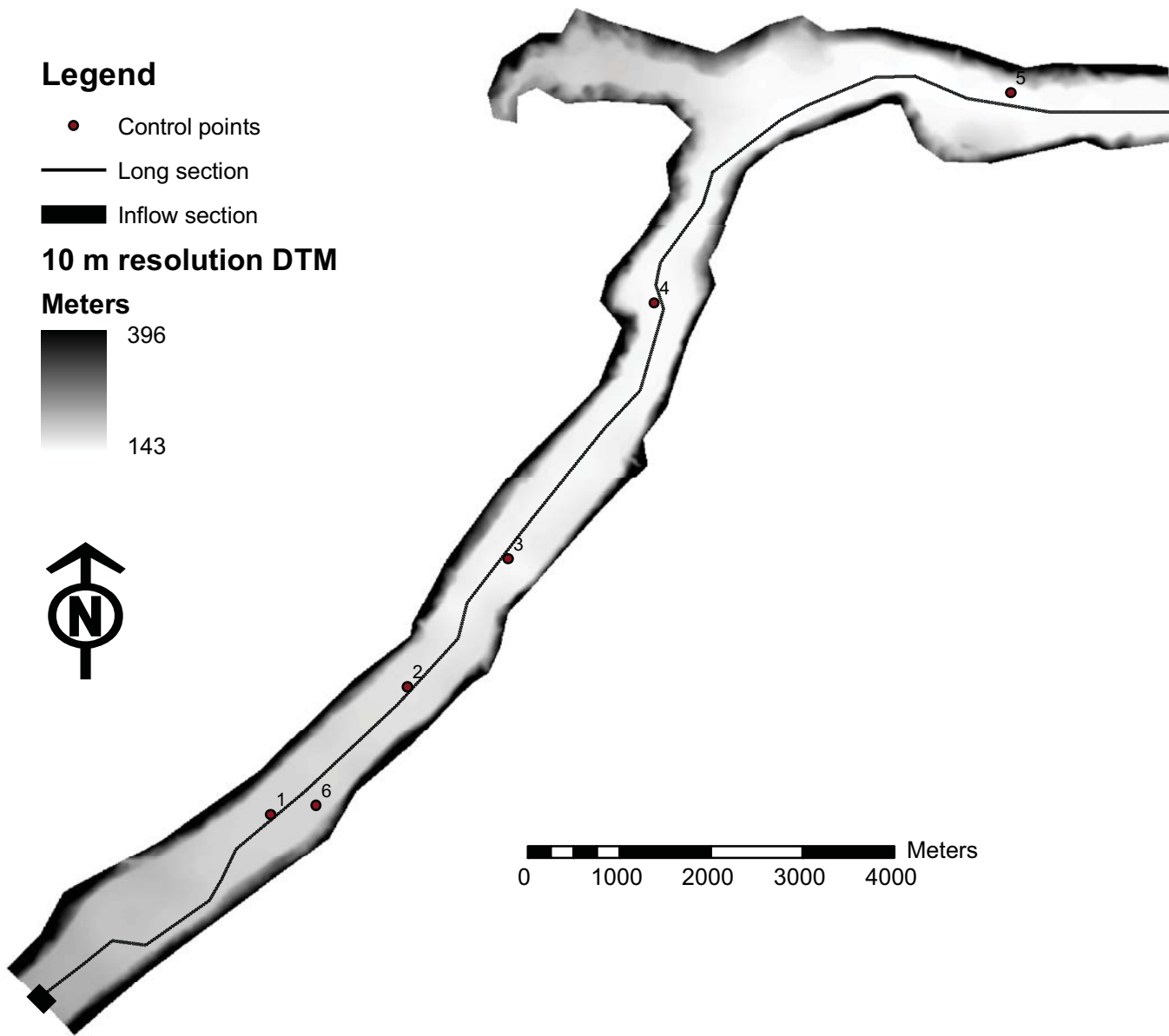
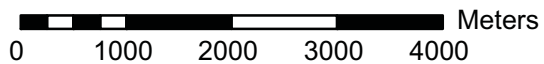
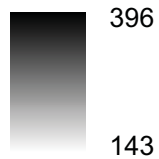


## Legend

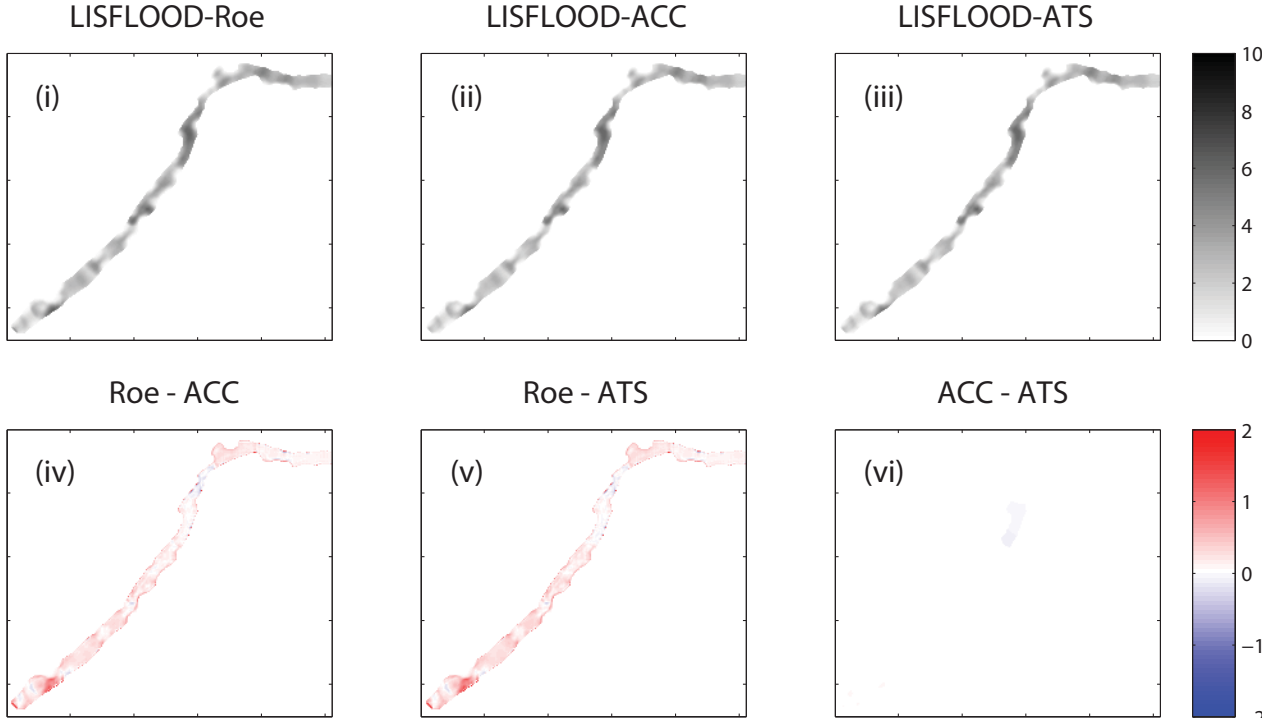
- Control points
- Long section
- Inflow section

10 m resolution DTM

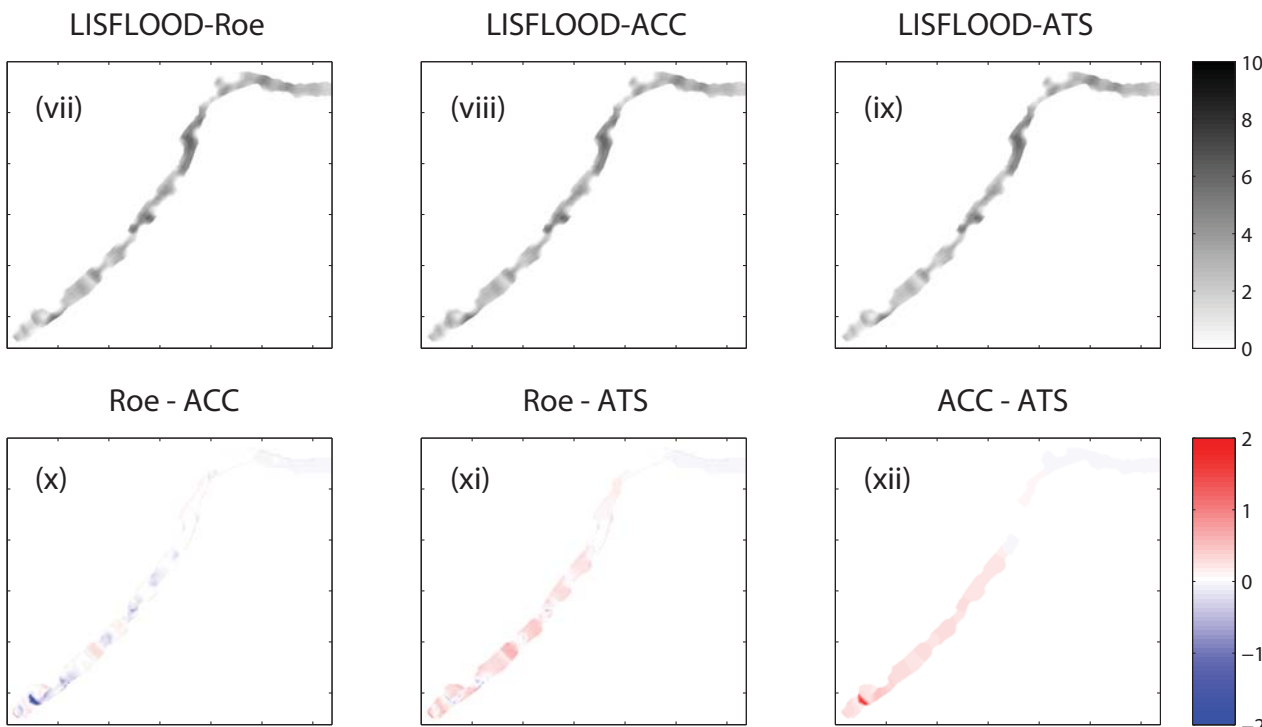
Meters

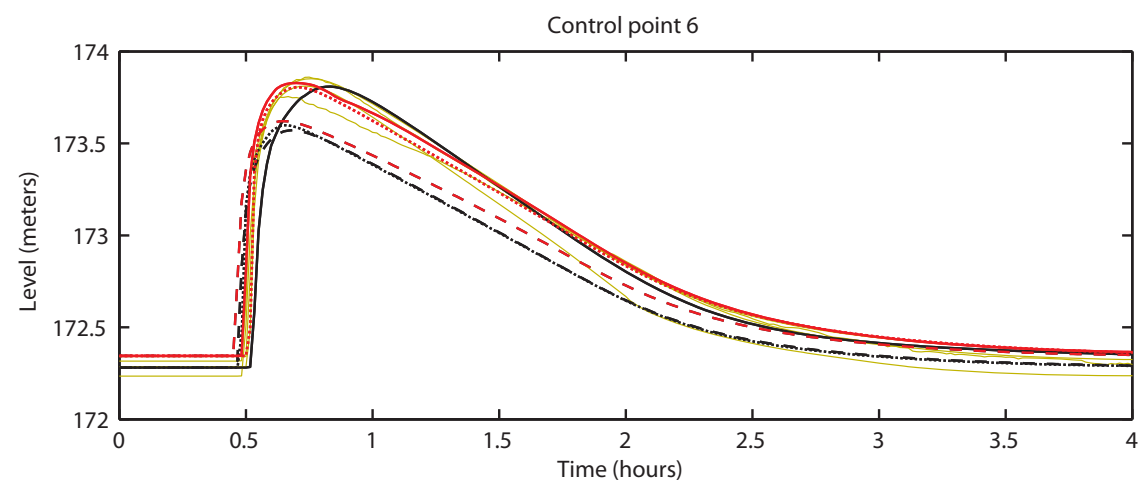
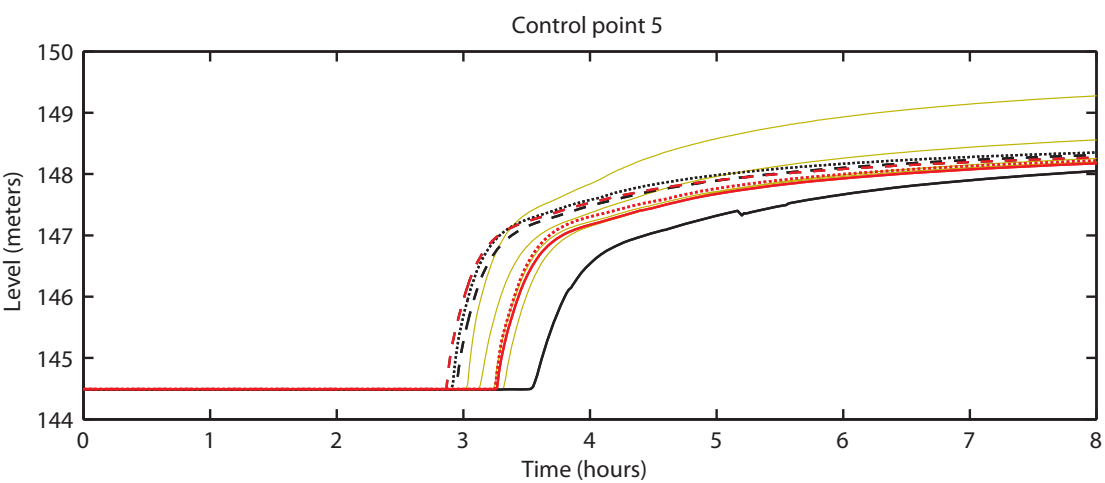
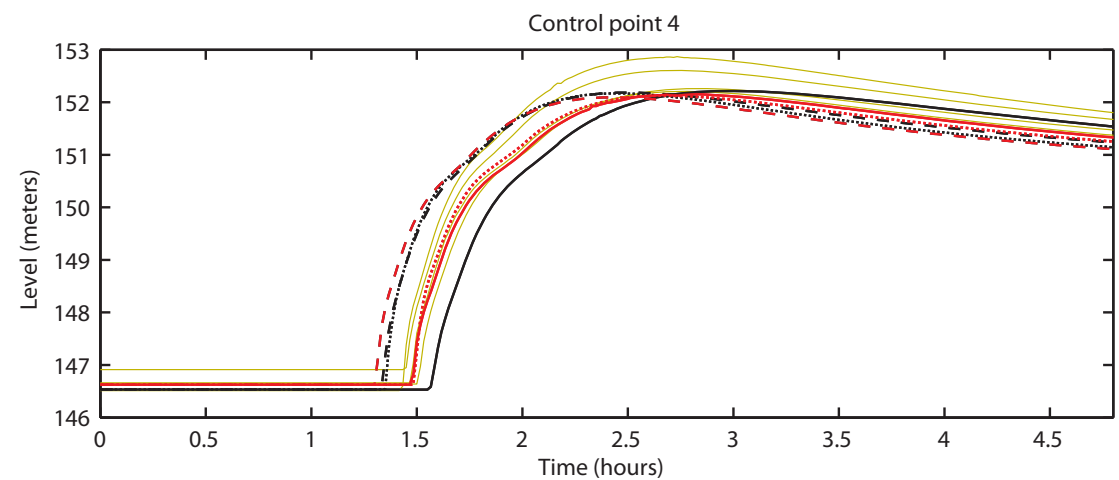
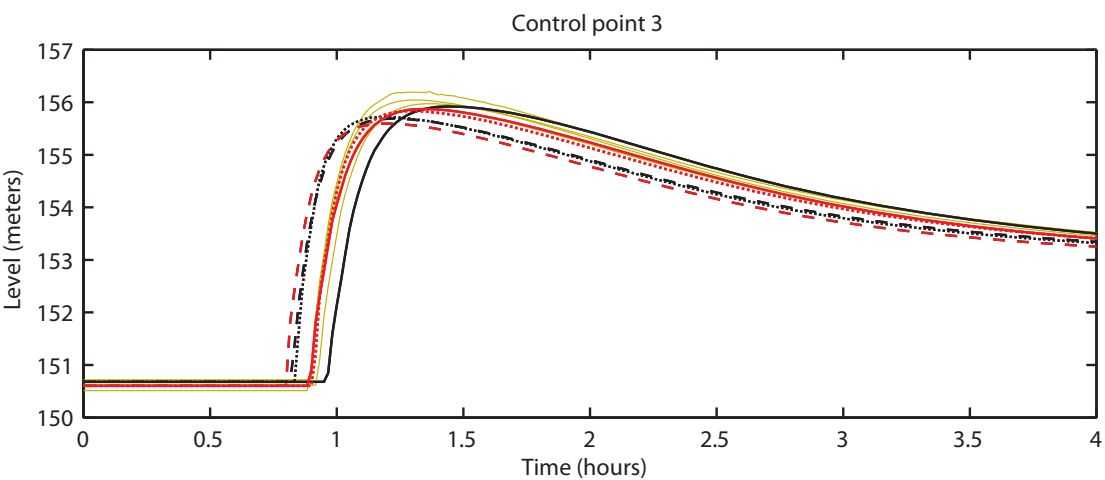
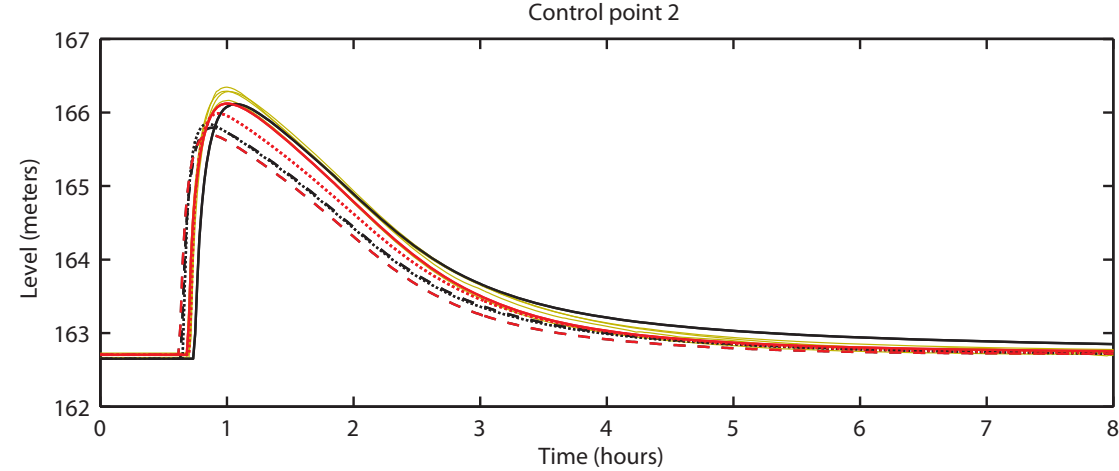
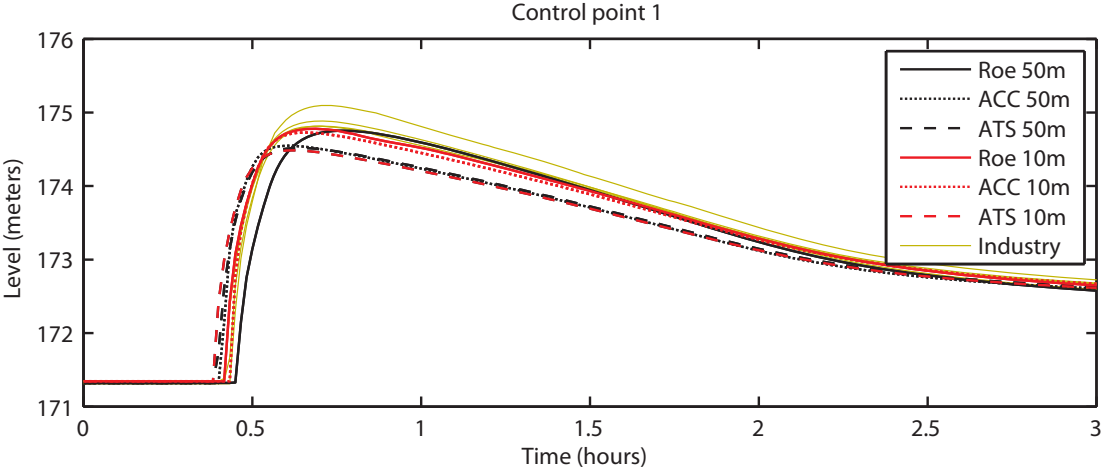


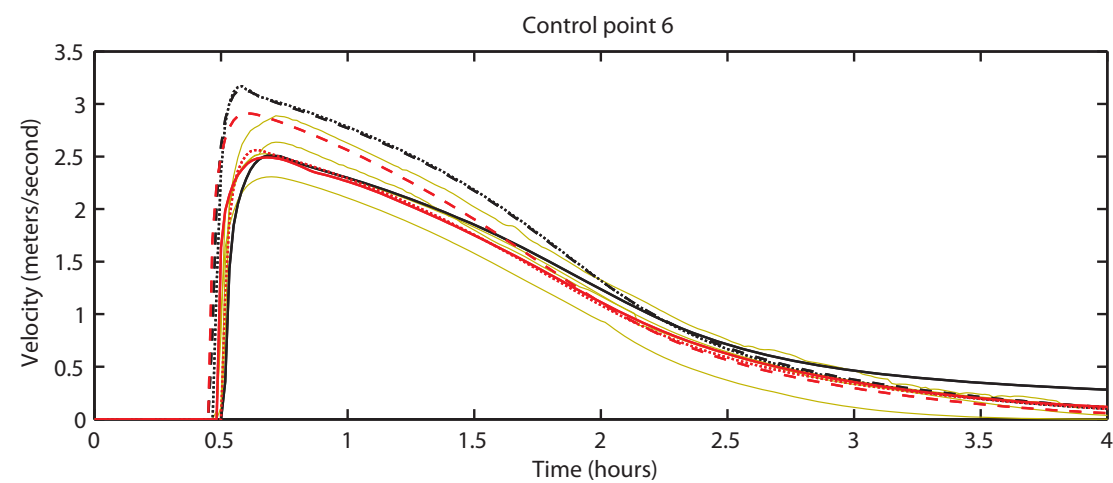
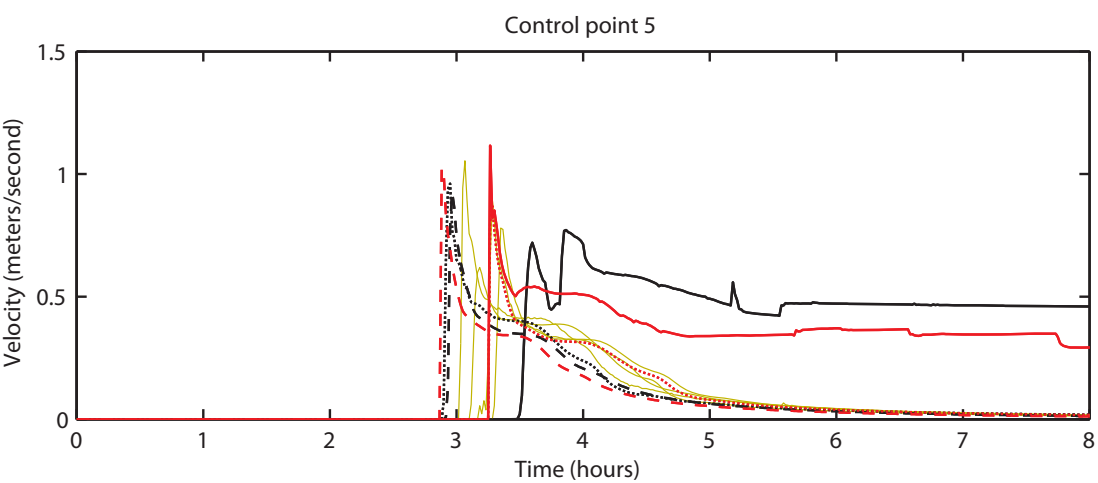
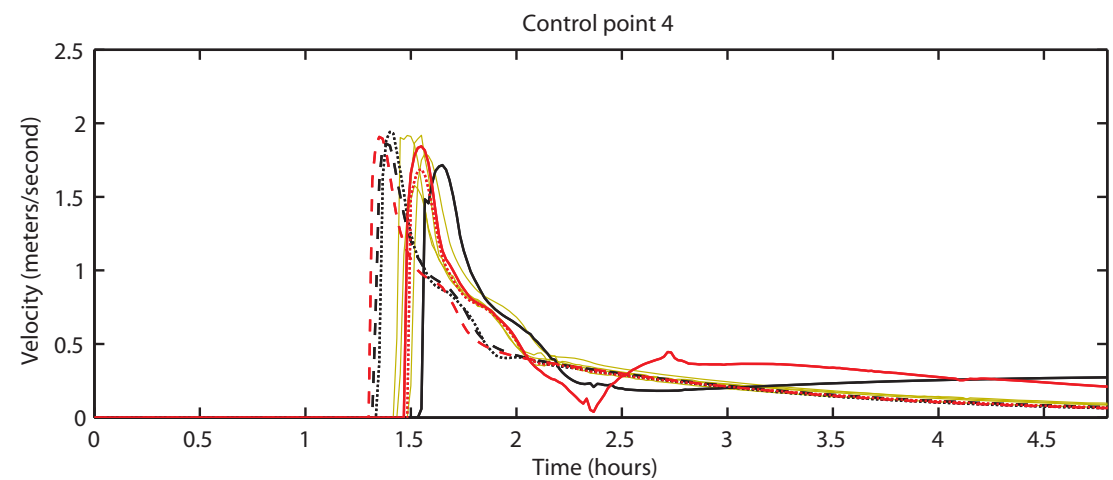
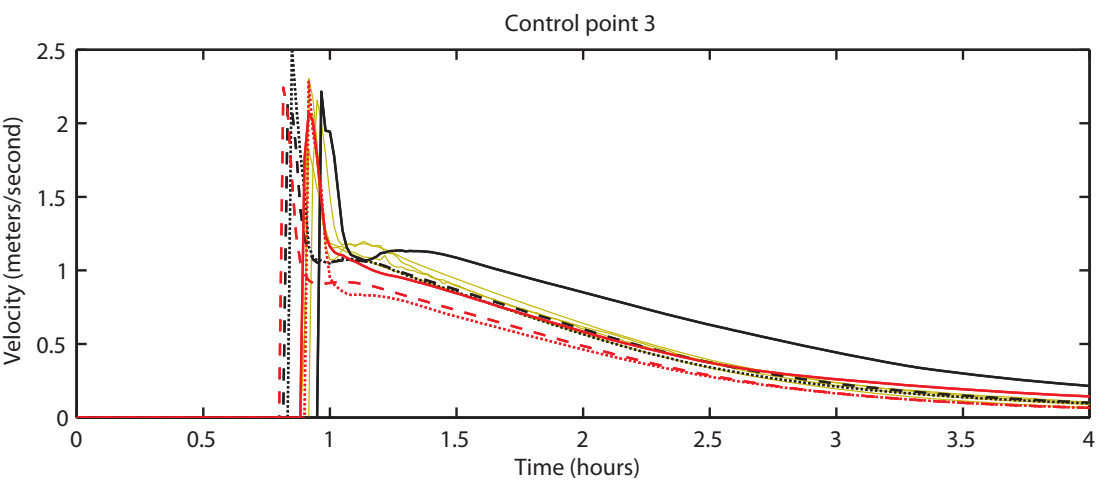
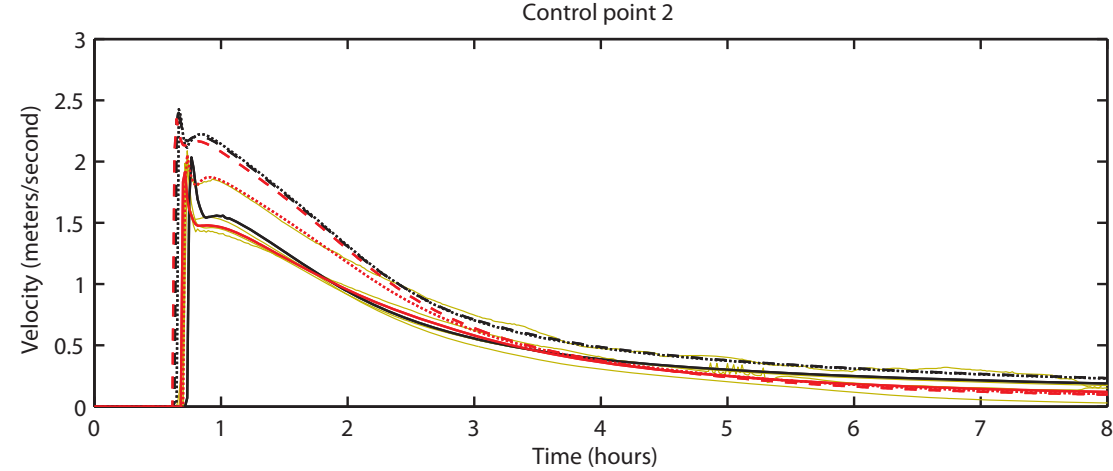
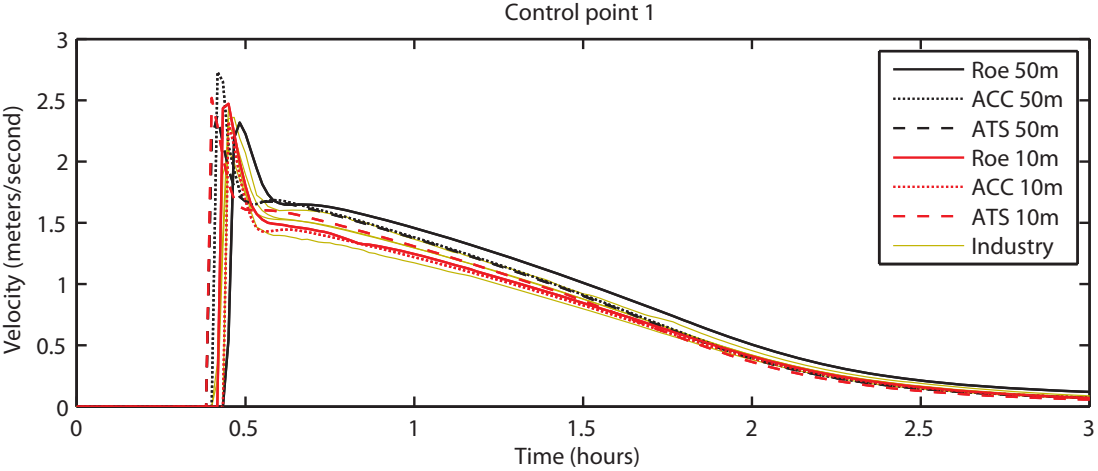
a) 50 m resolution models

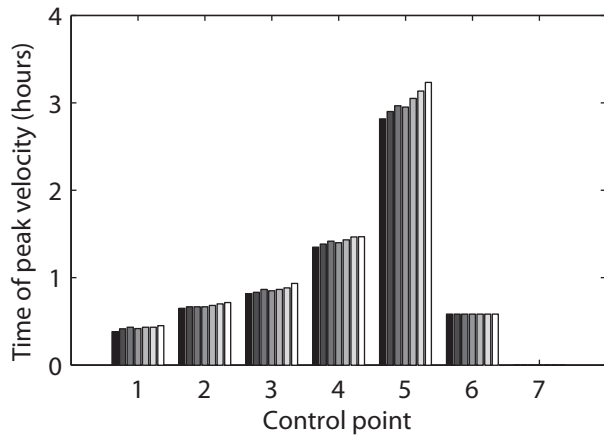
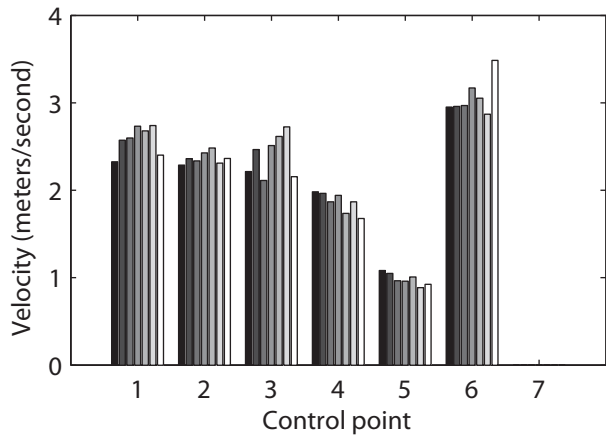
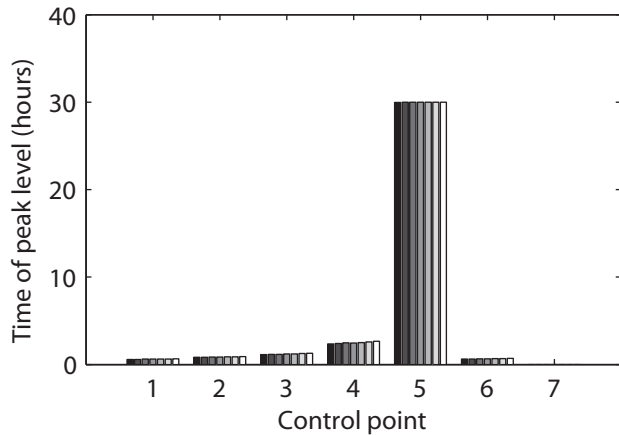
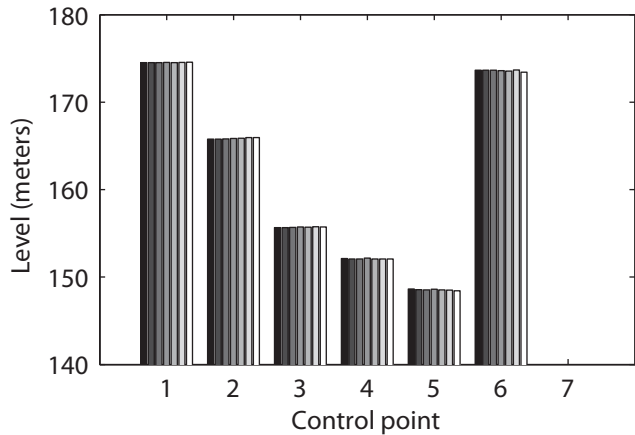


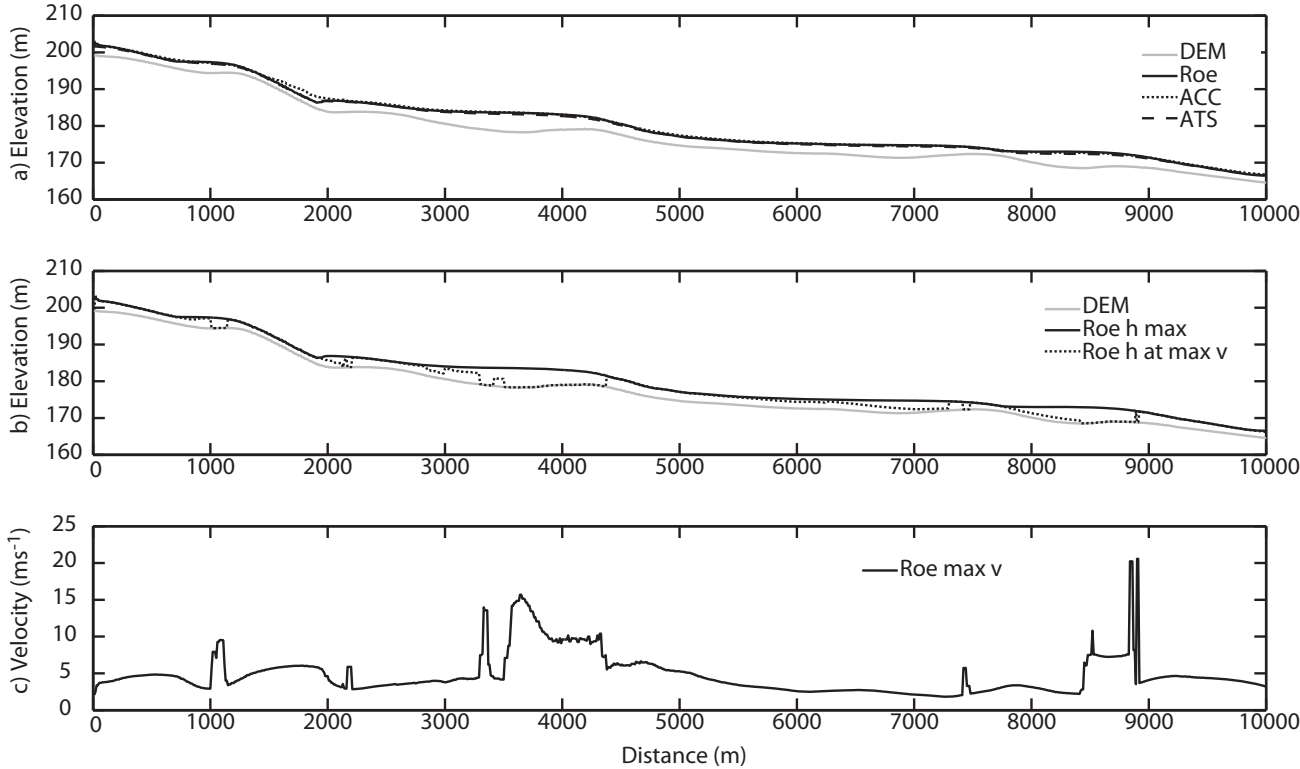
b) 10 m resolution models

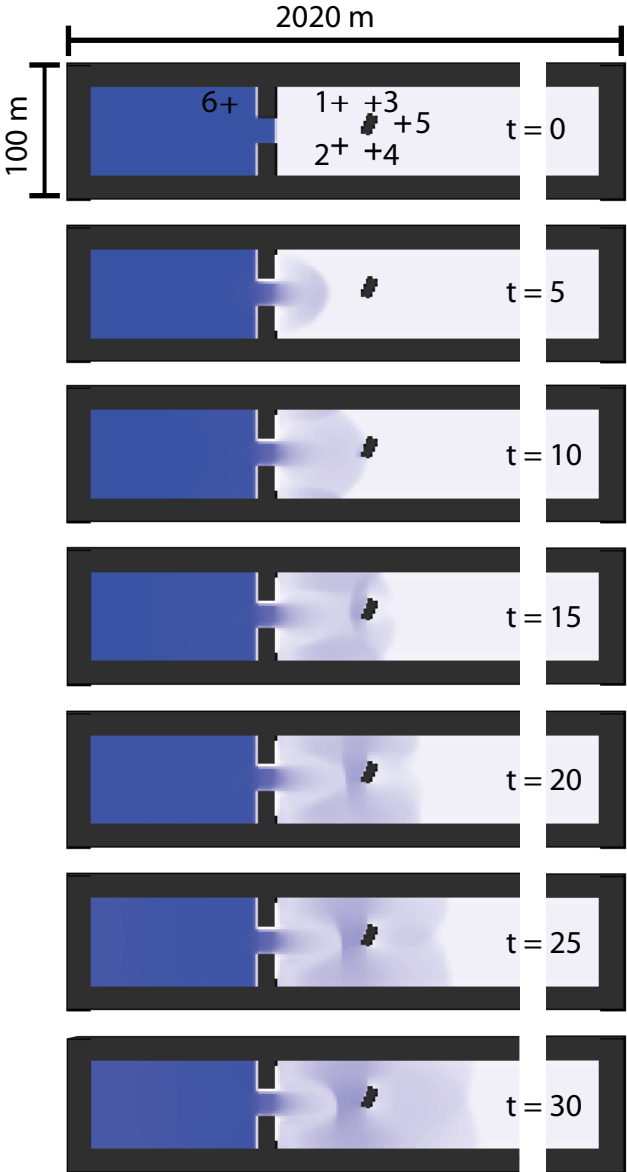




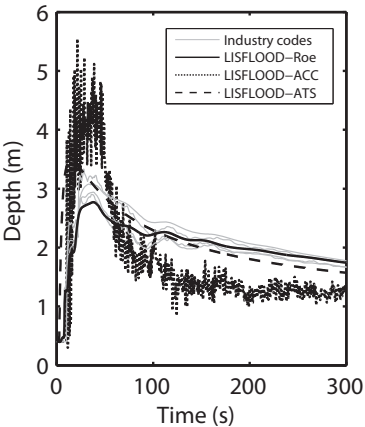




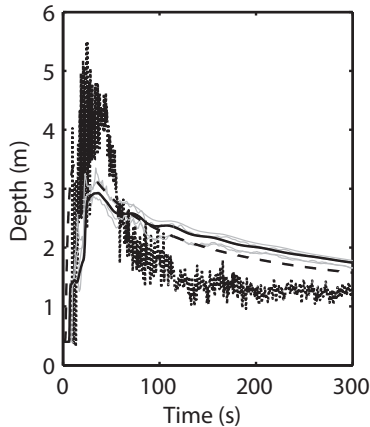




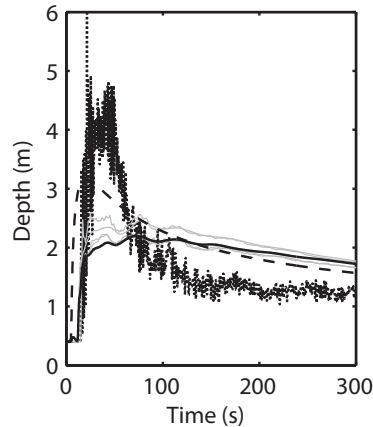
CP1



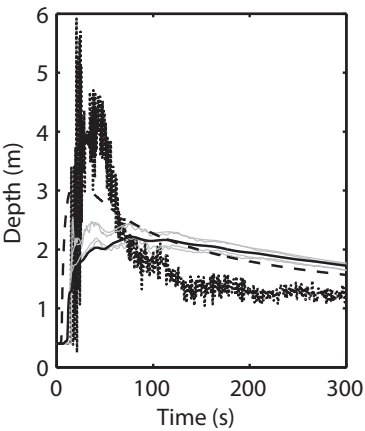
CP2



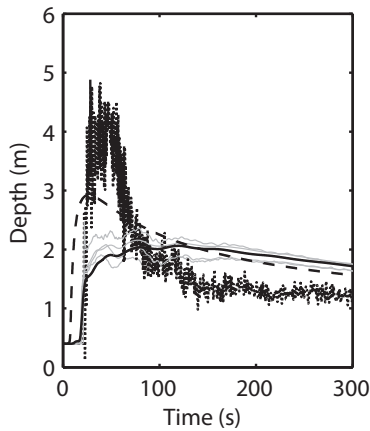
CP3



CP4



CP5



CP6

