Reducing inconsistencies in point observations of maximum flood inundation level

Article

Published Version

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Reducing Inconsistencies in Point Observations of Maximum Flood Inundation Level

Brandon L. Parkes* and Hannah L. Cloke

Department of Geography, King's College London, London, United Kingdom

Florian Pappenberger

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

Jeff Neal

School of Geographical Sciences, University of Bristol, Bristol, United Kingdom

David Demeritt

Department of Geography, King's College London, London, United Kingdom

ABSTRACT: Flood simulation models and hazard maps are only as good as the underlying data against which they are calibrated and tested. However, extreme flood events are by definition rare, so the observational data of flood inundation extent are limited in both quality and quantity. The relative importance of these observational uncertainties has increased now that computing power and accurate lidar scans make it possible to run high-resolution 2D models to simulate floods in urban areas. However, the value of these simulations is limited by the uncertainty in the true extent of the flood. This paper

---

*Corresponding author address: B. L. Parkes, Department of Geography, King's College London, Strand, London WC2R 2LS, United Kingdom.

E-mail address: brandon.parkes@kcl.ac.uk

addresses that challenge by analyzing a point dataset of maximum water extent from a flood event on the River Eden at Carlisle, United Kingdom, in January 2005. The observation dataset is based on a collection of wrack and water marks from two postevent surveys. A smoothing algorithm for identifying, quantifying, and reducing localized inconsistencies in the dataset is proposed and evaluated showing positive results. The proposed smoothing algorithm can be applied in order to improve flood inundation modeling assessment and the determination of risk zones on the floodplain.

KEYWORDS: Hydraulic modeling; Model calibration; Model validation; Uncertainty; Error correction

1. Introduction

With the shift to more risk-based approaches to managing flooding, flood hazard maps and simulation models have assumed new prominence as instruments for informing policy decisions about the regulation of land use and spatial planning, pricing and availability of flood insurance, and the allocation of resources for flood defense schemes. With so much at stake in those decisions, it is important to reduce the uncertainties associated with scientific assessments of flood risk and inundation extent, not least because they provide grounds for political controversy over risk management decisions (Porter and Demeritt 2012).

To that end, flood inundation modeling provides vital information to help assess the risk of future flood events, the effectiveness of proposed defense schemes such as dikes and levees, and the potential consequences of their failing (Dawson et al. 2005; Porter 2010).

While improvements in computing power and in the accuracy of digital elevation models (DEMs) have made it possible to simulate urban flooding using high-resolution (<10 m) two-dimensional (2D) flood models (Dottori and Todini 2013; Smith et al. 2006), they have also increased the relative significance of observational uncertainties of flooding extent. 2D flood models have now advanced to the stage that it is possible to simulate large-scale urban floods at a resolution high enough to realistically represent the physical processes affecting both the extent and dynamics of flooding in complex, built-up areas (Bates 2012), which are particularly important to understand given the great risks to life and property posed by such urban flooding.

2D flood models apply a variety of approaches to estimating solutions of the shallow-water equations defined by Saint-Venant (Saint-Venant 1871). Reviews of commonly used models can be found in Hunter et al. (Hunter et al. 2008) and Pender and Néelz (Pender and Néelz 2010). The model used for this research, LISFLOOD-FP, divides the floodplain into a regular 2D grid of storage components (Bates and De Roo 2000) and uses a simple wave representation that includes diffusion and inertia to simulate floodplain flow between cells (Bates et al. 2010). One of the motivating factors driving the development of LISFLOOD-FP is the desire for a physically based model that can take advantage of high-resolution floodplain DEM data, using the simplest possible process capable of describing two-dimensional dynamic flood inundation (Bates and De Roo 2000). This is justified on the basis that simplification of the physical process can be compensated for through the use of effective parameters to represent channel and floodplain friction (Horritt and Bates 2001). The benefits of this approach are that the model is

simple to set up, computationally efficient, and easy to integrate with third party GIS systems (Bates and De Roo 2000), although, as is generally the case with 2D flood models, calibration against observational data is necessary before the model is able to make accurate predictions (Horritt and Bates 2002).

Despite efforts to make them physically realistic, 2D models necessarily involve some parameterization; even if the parameters are physically based, they will not be accurate representations of reality (Beven 2008). Observational data of previous flood events in the model domain are required to calibrate the model in the hope of identifying the parameters sets that give rise to a useful representation of reality (Hunter et al. 2005). However, the nature of floods is such that a parameter set that accurately simulates one flood may not be representative for subsequent flood events of a different magnitude. Since extreme flood events are, by definition, rare and difficult to access safely, the observational data available are always limited.

There is a variety of different sources of data on flood inundation extent, marked by their own limitations. Recordings of water levels from river gauges in or near the affected area can be a valuable source of point data, especially if they produce frequent and reliable readings of water level. However, the utility of river gauge readings in model calibration can be limited: they only provide water level readings within the channel and so cannot inform the model on water extent across the floodplain; they are often known to become unreliable at times of out of bank flow (Brakenridge et al. 1998); and they are often used to provide the data for the inflow (and outflow) boundary conditions of the hydraulic model itself (see, e.g., Neal et al. 2011).

Surveys to collect remotely sensed imagery of the extent of the flood are an extremely valuable source of observational data, especially for 2D models where the image cannot only be used as a global calibration dataset of the model domain but also as a source of information for understanding and improving the model structure (Schumann et al. 2009). Prior to the invention of digital imagery, aerial photography has been used for the purposes of flood assessment since the early twentieth century (Bhavsar 1984). Even before the widespread use of computerized image processing, a skilled analyst could use information on aircraft height, camera specification, image size, and angle to accurately measure the spatial extent of flooding (Marcus and Fonstad 2008). Although aerial photography remains potentially the most accurate method of measuring flood extent remotely, the costs and logistics of rapidly commissioning airborne surveys, their dependence on fine weather, the difficulty of orthorectifying and building a mosaic from the images, and the requirement to manually identify water extents all conspire to limit their use.

Satellite remote sensing is increasingly being used to fill that gap in observational data (Schumann et al. 2009). Images from synthetic aperture radar (SAR) have often been used in the United Kingdom and elsewhere for monitoring floods (see, e.g., Biggin 1996; Dung et al. 2011; Horritt et al. 2001; Hunter et al. 2006; Mason et al. 2009; Matgen et al. 2010), and sophisticated algorithms have been developed to extract the water outline automatically (Mason et al. 2009; Schumann et al. 2007a). In spite of complications such as double-bounce reflections from buildings, the high-resolution SAR images from recently launched satellites such as TerraSAR-X are approaching the accuracy of aerial photos even in inundated urban areas (Schumann et al. 2011).

Figure 1. A wrack mark of debris deposited by a flood near Tewkesbury, United Kingdom. Photo taken 3 May 2012.

These limitations with conventional scientific instrumentation have spurred efforts to find alternative sources of data. For example, Connell et al. (Connell et al. 1998) interviewed 40 residents directly affected by floods of the Waiho River, New Zealand, in 1986 and 1994. As well as the descriptions of maximum flood levels, residents also provided photos and videos. Interpretation of the images was hampered by the oblique angle of the photos and lack of information on the time the images were taken, but Connell et al. estimate the water depth accuracy to be within ± 0.2 m. The recent expansion of digital cameras, videos, and mobile phones with cameras means there are now likely to be extensive photographic records of floods with time stamps, and the popularity of online image sharing sites such as Flickr eases the process of data collection. After a flood at Morpeth, United Kingdom, in September 2008, Parkin (Parkin 2010) used almost 1500 pieces of information to build a series of hourly snapshots of the rising limb of the flood.

Instead of interviewing eyewitnesses or collecting their photographs, postevent surveys in the immediate aftermath of flooding can also seek to identify and map signs of the recent flood, in the form of either water marks (stains on vertical surfaces such as walls) or wrack marks (lines of debris deposited on the ground by the water as it begins to recede) (see Figure 1). While it is possible to identify high water marks from remotely sensed images of high quality and resolution (e.g., see Lane et al. 2003), local surveys are considered less uncertain (E. M. Stephens and P. D. Bates 2013, personal communication). Local surveys of wrack marks using handheld global positioning system (GPS) devices allow for straightforward recording of the horizontal location of high water marks that can subsequently be intersected with the DEM to give the location in the vertical (Schumann et al. 2007b); the accuracy of this method is dependent on the local topographic slope as well as the precision of the GPS measurement. More recently, the emergence of differential GPS (dGPS) devices allow direct, in situ vertical measurements of the elevation of water and wrack marks accurate to within 0.01 m (Horritt et al. 2010), although the accuracy of those marks as measures of the water height itself may not

be quite so precise. For example, the surveyor may miss the peak deposition line, or the surface may retain marks higher than the extent of the water caused by upward diffusion of the deposition surface. Often these high water mark data are used as a supplement to more conventional sources of flood extent data (Hunter et al. 2005; Pappenberger et al. 2006; Schumann et al. 2007a), but sometimes data from extensive surveys have provided the primary data source for model calibration: Mignot et al. (Mignot et al. 2006) used a set of 99 high water marks on buildings in the town of Nîmes, France, to calibrate a model of a severe flood in 1988; a small, but destructive flood in Boscastle, United Kingdom, in 2004 was modeled by Lhomme et al. (Lhomme et al. 2010) using a survey of 72 wrack marks for model validation; and the aftermath of the flood in Carlisle, United Kingdom, in January 2005 was extensively surveyed by two independent teams giving rise to a dataset consisting of 217 wrack marks and 46 water marks. In the absence of any remotely sensed images of the flood, the data from the surveys have been used as the primary source of calibration data for several projects modeling the event (Fewtrell et al. 2011; Horritt et al. 2010; Leedal et al. 2010; Neal et al. 2009).

That dataset from the 2005 Carlisle flood is the focus for the analysis presented in this paper. Some analysis of the accuracy and characteristics of the wrack and water marks was performed by both Neal et al. (Neal et al. 2009) and Fewtrell et al. (Fewtrell et al. 2011). Neal et al. (Neal et al. 2009) found evidence that wrack marks underestimate the peak water level when compared to water marks by 0.51 m on average, a difference they attributed to some combination of the surveyor missing the highest wrack mark in locations where multiple depositions of debris occurred because of the receding waters remaining stationary for multiple periods and, in the case of water marks on buildings, the staining maybe being higher than the actual water level because of capillary action. Fewtrell et al. (Fewtrell et al. 2011) went further with their analysis, comparing wrack and water marks not only against each other but also with the peak water height recorded by a nearby river gauge. They suggested the mean difference between proximate observations to be 0.1 m, which is within the accuracy of the DEM, although they did find greater average differences when comparing wrack marks to water marks (Fewtrell et al. 2011). Other than the examples of Neal et al. (Neal et al. 2009) and Fewtrell et al. (Fewtrell et al. 2011), both from the 2005 Carlisle event, it seems there has been little or no effort to systematically measure the uncertainty in wrack and water marks as measures of maximum flood extent and depth. Apart from the lack of reference data against which to measure the wrack marks, this is also perhaps because other uncertainties in flood models, or indeed the accuracy of the DEM itself, were thought to dwarf inaccuracies in physical records of the water extent. However, as highly accurate lidar surveys become more widely available and the resolution of 2D models increases, the uncertainty of the measurements will become relatively more significant (Bates 2012).

The purpose of this paper is to assess the inaccuracies associated with observational datasets from locally surveyed wrack and water marks and to propose a "smoothing" algorithm that can be used both to facilitate the identification of specific errors and for reducing overall observational uncertainty. The paper is organized as follows: in the following section (section 2) the study event and datasets are introduced. Next, in the methods section (section 3) the smoothing algorithm is described and applied to the dataset, the results of which are shown in
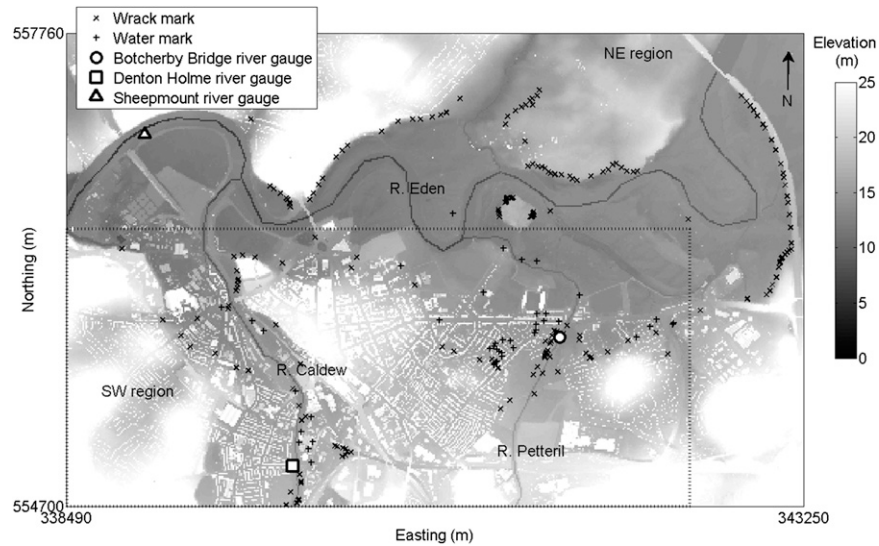
Figure 2. Digital elevation map of Carlisle, United Kingdom, showing the main water courses, locations of river gauges, and observations of maximum water extent from records of wrack marks (x's) and water marks (crosses). The dotted line delineates the southwest, largely urban, subset of wrack and water marks from the largely rural subset in the northeast.

section 4. Finally, conclusions are drawn on the applicability of the method and its potential benefit to flood inundation modeling.

2. Case study: January 2005 River Eden flood at Carlisle, United Kingdom

Carlisle is a town in Cumbria, northwestern England, with a population of approximately 100 000. The town was an important Roman settlement established to serve forts on Hadrian's Wall, remnants of which still stand in the town today. Carlisle is located on the River Eden with two notable tributaries, the Caldew and the Petteril, joining the Eden at Carlisle (see Figure 2). The catchment area of the River Eden and its tributaries is largely in the English Lake District, a national park with the highest annual rainfall in England (Barker et al. 2004).

In January 2005, up to 175 mm of precipitation fell in the Eden catchment over a period of 36 h (Fewtrell et al. 2011). This resulted in an estimated 1 in 200-yr flood event in Carlisle (Mason et al. 2007b) with approximately 1800 homes inundated (Day 2005). The peak discharge on the Rivers Caldew and Petteril preceded that of the Eden, and it was these tributaries that first reached out of bank conditions at approximately 0200 UTC 8 January in the Denton Home, Caldewgate, and Botcherby Bridge areas of Carlisle, the flooding exacerbated by blockages from debris (Day 2005; Environment Agency 2005). The River Eden overtopped its defenses several hours later at approximate 0830 UTC, and there followed significant backwatering effects along the main tributaries (Day 2005; Environment
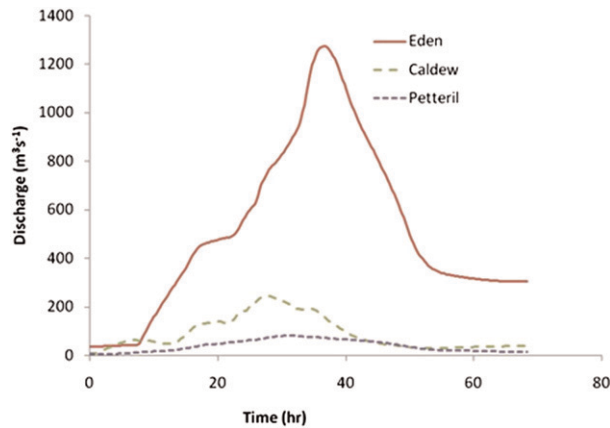
Figure 3. Hydrograph of the three main water courses in Carlisle at the time of the January 2005 flood. The start time of the hydrograph is 0045 UTC 7 Jan 2005.

Agency 2006). In the weeks following the flood, two postevent survey teams from the University of Bristol and the Environment Agency identified visible evidence of the maximum extent and/or depth of the floodwaters and measured and recorded their locations using dGPS. The dGPS locations were recorded relative to a temporary base station within the study area, which had been located relative to a nearby ordnance survey maintained reference point (Horritt et al. 2010). Horritt et al. (Horritt et al. 2010) estimate the accuracy of the devices to be within ± 0.01 m; although this has not been verified by the authors, it is consistent with figures quoted for comparable surveys (e.g., Rayburg et al. 2009). Six river gauges are in or immediately upstream from the study area: the locations of three of these gauges are marked in Figure 2. Ideally the inflow boundary conditions for the three rivers would be provided by data from the upstream gauges. However, the upstream gauges on the Eden and the Caldew were known to be unreliable, so new ratings had to be derived for these gauges from a separate modeling exercise (described in Horritt et al. 2010). It is these new ratings that were used here to derive the upstream boundary conditions for the Eden and Caldew; the resulting hydrographs from the three rivers are shown in Figure 3. Some of the most destructive flooding occurred alongside the Caldew, where initial flooding was likely to have been caused by partial channel blockage by debris at a footbridge. This was followed by significant backwatering up the Caldew channel as the main flood wave on the Eden arrived (Environment Agency 2005).

3. Methods

3.1. Analysis of wrack and water marks in Carlisle

As mentioned in the introduction, both Neal et al. (Neal et al. 2009) and Fewtrell et al. (Fewtrell et al. 2011) provided estimates of the accuracy of the wrack and water marks surveyed in Carlisle after the January 2005 flood. Neal et al. (Neal et al. 2009)

restricted their analysis to a comparison of the only area where multiple wrack and water marks were located within 100 m of each other. From this limited subset, two wrack marks were compared to four water marks giving a mean discrepancy of 0.51 m (Neal et al. 2009). Fewtrell et al. (Fewtrell et al. 2011) carried out a more extensive analysis comparing all wrack and water marks that are within 500 m of each other and also making comparisons against readings from the river gauge at Botcherby Bridge, which is located within a dense cluster of wrack and water mark observations (see Figure 2). They conclude the mean difference between true peak water levels and wrack/water mark measurements to be approximately 0.1 m. Here we build on the work by Fewtrell et al. (Fewtrell et al. 2011) by first comparing observations not just with other, individual observations but also with the average of a set of nearby observations. This allows us to not only estimate the internal uncertainty in the set of observations but also pinpoint individual measurements that are particularly inconsistent and later, in section 3.2, propose a smoothing algorithm for reducing those inconsistencies.

We analyze each observation by identifying the 10 nearest neighbors (10nn) to that point and comparing the recorded height of the point with the mean height of the 10nn. The results can be seen in Figure 4.

Figure 4a shows data for all 263 observational data points. Little information can be gained from the points that are clearly isolated from their neighbors (i.e., where the mean distance to the 10nn is greater than 200–300 m). For these points, the height difference shows considerable variation here, as is expected because no account is taken of water slope between neighboring observations. When the observations are in closer proximity there is less variation, but several points stand out by showing discrepancy in height of over 0.5 m with a mean distance of less than 100 m to their 10nn. This difference is unlikely to be accounted for by water slope alone, which would be expected to be between 0.0001 and 0.001 $m\,m^{-1}$ [0.01–0.1 m $(100\,m)^{-1}$] (Mason et al. 2007a); indeed Horritt et al. (Horritt et al. 2010) estimate the slope at Carlisle to vary between 0.000 25 and 0.0006 $m\,m^{-1}$. However, it cannot necessarily be concluded that any larger discrepancy shown by wrack and water marks must be due to inaccuracy in the reading: in this event, peak water heights are likely to be influenced by small-scale hydraulic effects in the built-up areas or where waterways become blocked by debris, as was the case for both the Petteril and the Caldew (Environment Agency 2005). To provide a simple visual summary, Figure 4b shows the mean absolute height differences collected into 50-m bins. The mean absolute height difference for all measurements where the mean distance of less than 100 m to their 10nn is 0.105 m, an estimate of uncertainty that is consistent with the figure of 0.1 m estimated by Fewtrell et al. (Fewtrell et al. 2011). The mean height difference can be seen to increase in areas where the observations are more sparse (greater mean distance to 10nn) and the effect of the slope of the water is likely to have more influence.

As can be seen from Figure 2, many of the wrack marks were collected beyond the north and eastern banks of the Eden away from the urban areas affected by the flood. In these areas the water level is likely to vary smoothly with little small-scale variation, and there are many wrack marks in close proximity. To explore this effect, Figure 5 splits the data point into two zones: the northeast, where all the observations are in rural locations, and the southwest, where many of the observations are in built-up urban areas (see regions in Figure 2). It is clear that the
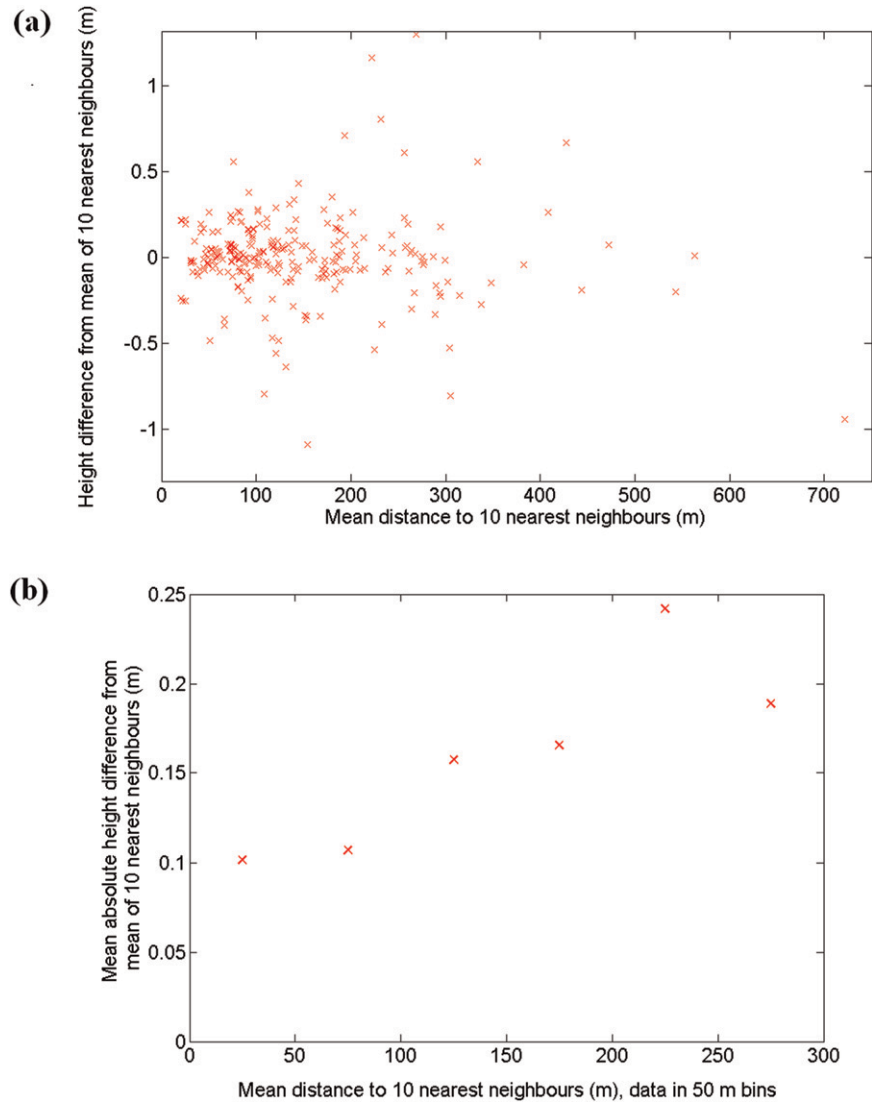
Figure 4. (a) For each observational data point, difference between recorded height of data point and mean height of the 10nn plotted against mean distance to the 10nn. (b) Mean absolute height difference between data points and their 10nn, collated into 50-m bins.

variations in observations are greatest in the urban areas. The explanation is less clear: there may well be more highly localized spatial heterogeneity in maximum water depth; alternatively, measurement errors may be higher in the urban area [Horritt et al. (Horritt et al. 2010) point out that difficulties of obtaining GPS measurements next to buildings due to poor satellite coverage]; and the observations in northeast are, with one exception, all wrack marks, whereas the observations in the southwest are a more heterogeneous mix of wrack and water marks, and any bias between the two types of observation will manifest itself as a greater variation observation heights.
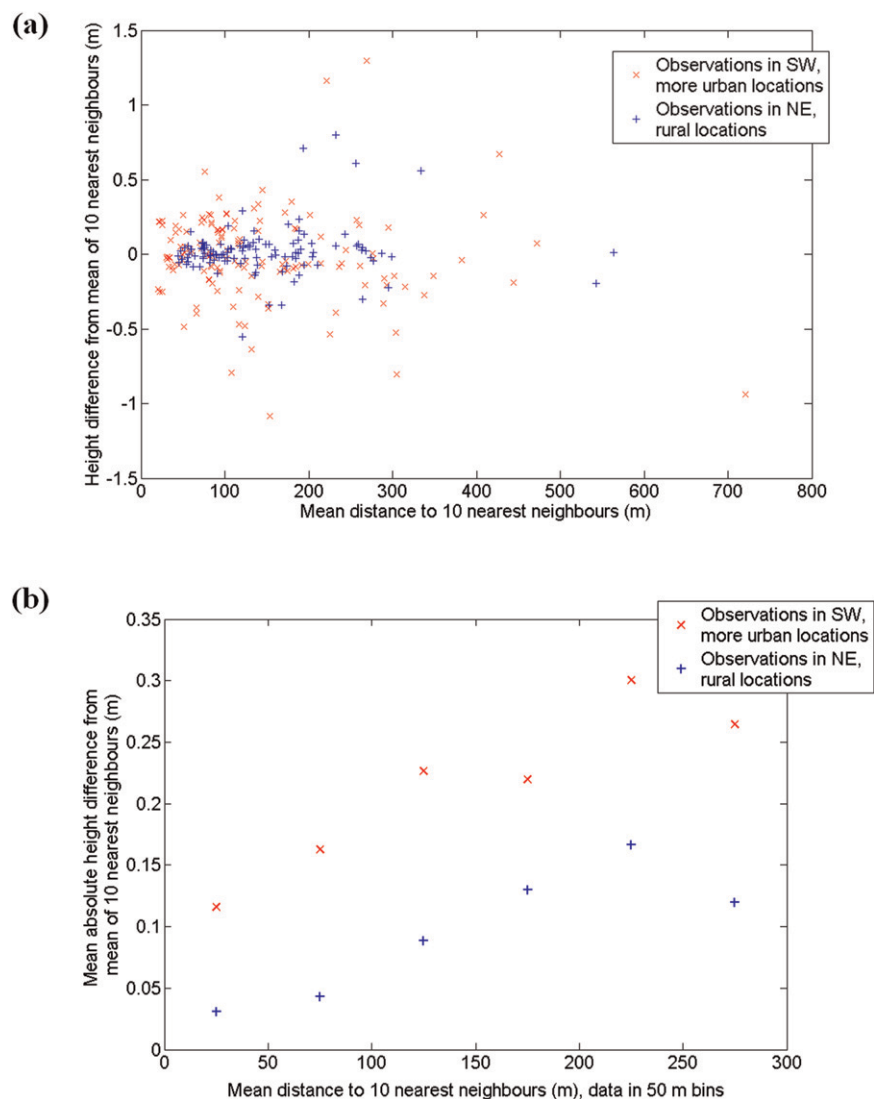
(a)

(b)

Figure 5. (a) For each observational data point, difference between recorded height of data point and mean height of the 10nn plotted against mean distance to the 10nn. (b) Mean absolute height difference between data points and their 10nn, collated into 50-m bins. Observations are split between those to the north and east of the study area (in rural locations) and those in the south and west, in more urban areas.

3.2. Proposed smoothing algorithm for improving the accuracy of water and wrack mark readings

In this paper, we propose a smoothing algorithm that can either be applied to the raw dataset to attempt to reduce inconsistencies between neighboring observations or can be used to identify the most inconsistent data points, which the researcher may then decide to exclude from the dataset. The proposed smoothing algorithm

Table 1. Parameters used by the smoothing algorithm.

Parameter	Description	Default value
Nearest neighbors (nn)	The maximum number of neighbors that can influence each data point	10
Validity cutoff (v_h)	Determines the influence each point has on its nearest neighbors	0.5 m
Alpha (α)	Determines the rate at which the influence of a point decreases with increasing distance	100

differs somewhat from a suggestion made previously by Neal et al. (Neal et al. 2009) to run multiple interpolations, missing out one measurement each time to identify outliers that may have large errors. The analysis above suggests that there are not many instances where an obvious outlier could be identified, particularly in the urban areas where most of the variation in observation height occurs. Instead an algorithm has been developed that adjusts each reading by an amount that is influenced by a number of its geographically nearest neighbors. The influence that any one reading has on another is affected by the geographical separation of the readings and the variance between the influencing reading and the mean height of its nearest neighbors. The physical basis for attempting to smooth the observations in this way is the assumption that water level slopes at this stage of a river are low (Mason et al. 2007a), so the peak water level should not vary on a short spatial scale as much as the raw data suggest. It is acknowledged that the hypothesis that the maximum flood level will vary smoothly is more likely to be true in open, rural topography, whereas maximum flood levels in urban areas or near infrastructure such as bridges and manmade culverts may well show greater variation because of the hydraulic effects of the constructions. This point is addressed in the results section by comparing the effects of the smoothing algorithm in urban and rural areas and by analyzing the effects of the algorithm on individual observations near a large bridge.

The smoothing algorithm is a two-stage process involving three parameters described in Table 1, which can be tuned using a process described in section 3.3. The algorithm works by first calculating a "validity weighting" for all the observations in the set of wrack and water marks. Each mark M is given a validity weighting V between 0 and 1, which is calculated from the mean of the difference between the water height as recorded at M and the recorded water heights of the set of nn to M ($\overline{\Delta h}$). If the absolute value of $\overline{\Delta h}$ is greater than the validity cutoff v_h, the reading is not deemed to have an influence on neighboring points and is assigned a validity weighting of zero. Otherwise, the validity weighting of point M (V_M) is given by

$$V_M = 1 - \left(\frac{1}{v_h} |\overline{\Delta h}| \right). \tag{1}$$

It should be noted that the validity weighting is not necessarily a measure of the validity or accuracy of the reading, just a measure of the validity of using the reading as an influence on its neighbors.

Next, the proposed adjustment C is calculated for each observation M from its nearest neighbors such that

Table 2. Sensitivity indices of the model to the four parameters varied on the Latin hypercube sample.

Parameter	First-order sensitivity index (S_i)
Nearest neighbors (nn)	0.250
Validity cutoff (v_h)	0.082
Alpha (α)	0.202
Total first-order sensitivity	0.534

$$C_M = \frac{\sum_{i=1}^{nn} \Delta h_i V_i \left[\frac{1}{(1 + d_i/\alpha)} \right]}{\sum_{i=1}^{nn} \sqrt{V_i \left[\frac{1}{(1 + d_i/\alpha)} \right]}}, \tag{2}$$

where d_i is the distance from M to its ith nearest neighbor and α is a scaling constant that determines how the influence of one point on another decreases with distance. A figure of 100 for α means that a separation of 100 m between two points will reduce the influence the points have on each other by 50%. Section 3.3 describes how the parameter can be tuned to select values suitable for the specific dataset.

3.3 Tuning process for smoothing algorithm

The process for selecting suitable parameter values for the three parameters used by the smoothing algorithm involves multiple simulations of sets of point observations of a known, ideal water surface. For each simulation, a number of data points are randomly selected from the ideal water surface and their heights are given a random perturbation away from the height of the water surface.

Several aspects of the simulation can be chosen to match the actual study area. For example, in order to select algorithm parameters appropriate for the observational dataset from the January 2005 Carlisle flood, 263 data points were simulated across a rectangular area of 4760 m by 3060 m. The simulated water slope was 0.001 m m^{-1}, which is the upper end of the range suggested in section 3.1; this was considered suitable because of the presence of smaller tributaries and urban inundation, which are likely to give rise to steeper water gradients in some areas. The simulated data points were randomly perturbed by an amount following a normal distribution, mean of 0, and standard deviation of 0.105 m to match the estimate of observational uncertainty calculated in section 3.1. Each simulation was scored on the basis of how much the root-mean-square error (RMSE) of the perturbed, simulated data points was reduced after applying the smoothing algorithm. After running a Monte Carlo simulation, the default values for the parameters given in Table 1 were deemed suitable for the Carlisle 2005 observational dataset. A first-order sensitivity analysis using the method described by Saltelli et al. (Saltelli et al. 2008) suggest the performance of the algorithm is most sensitive to nn (see Table 2).

4. Results

When this algorithm is applied to the dataset of 263 wrack and water mark readings collected after the 2005 Carlisle flood, the mean absolute change to the points was 0.0823 m with a maximum increase in height of 0.5951 m and a maximum decrease of 0.5938 m. To a certain extent, the assessment of the value of these results is a subjective choice to be made by the researcher. It may be that the results are just used as guide to identify the most incongruous data points that should be considered for removal. Here the results of the algorithm are used to create an alternative, smoothed dataset, which is initially assessed for internal consistency (section 4.1), and then a limited objective assessment is made against river gauge data (section 4.2). Finally, a comparison is made with the raw dataset by using both datasets to calibrate a 2D hydraulic model of the flood event (section 4.3).

4.1. Internal consistency

Figure 6 shows the height difference between each point and the average of its 10nn for both the original ''unsmoothed'' data and after applying the smoothing algorithm. As expected, Figure 6 shows the algorithm has a marked smoothing effect on the data, reducing the height variation among proximate readings; the mean absolute height difference for all measurements where the mean distance of less than 100 m to their 10nn is reduced from 0.105 to 0.038 m for the smoothed data. The mean change was 0.0043 m suggesting no significant overall shift in the readings up or down. Figure 7 shows the separate effects of the algorithm on urban and rural areas, where, as expected, the variation between proximate readings is seen to be greatest in urban areas.

4.2. Comparison with river gauge data

As remarked previously, the wrack and water marks form the bulk of the validation data for simulations of this event, so there is not a significant body of data against which to validate the effect of the smoothing algorithm. However, there are recordings of maximum flood levels recorded by the Botcherby Bridge and Denton Holme river gauges (locations marked on Figure 2), which are both close to a mixture of wrack and water mark observations. Water level readings from river gauges at times of flood are less problematic than the corresponding discharge estimates and the accuracy of the peak water levels is thought to be around 1–2 cm (Di Baldassarre and Montanari 2010), so comparing the water height observations in the vicinity of the river gauges will help provide an objective indication of their accuracy. Figure 8 shows the peak water heights of all wrack and water marks within 250 m of the Botcherby Bridge and Denton Holme gauges. Table 3 shows how the mean difference between the peak water height recorded by the river gauge and the proximate observations is reduced once the smoothing algorithm has been applied. The reduction in height difference between observations and gauge is more marked for the Botcherby Bridge, possibly because there are more observations very close to this gauge.
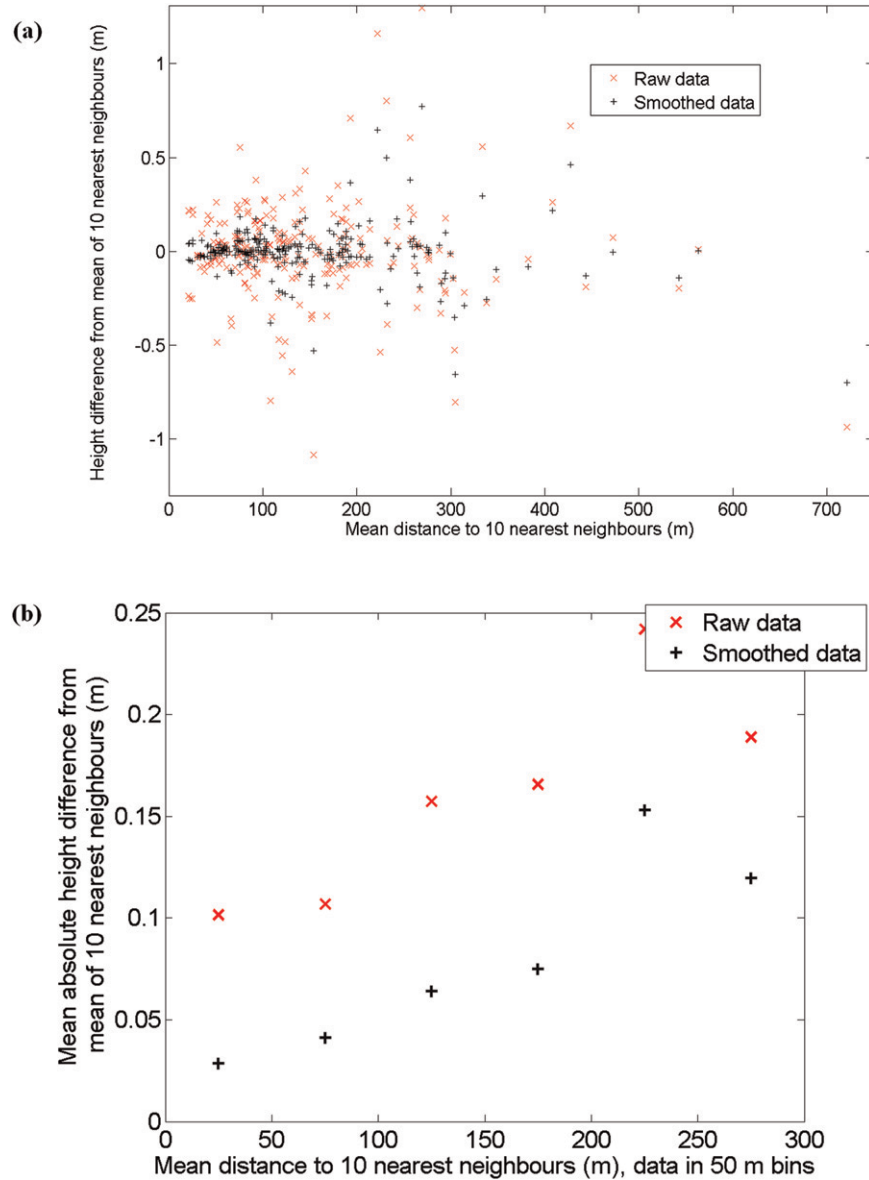
Figure 6. (a) For each observational data point, difference between recorded height of data point and mean height of the 10nn plotted against mean distance to the 10nn. (b) Mean absolute height difference between data points and their 10nn, collated into 50-m bins. Raw (red x's) and smoothed (black cross) data are denoted.

4.3. Evaluation using flood simulation data

In this section, we evaluate the effects of the smoothing algorithm by comparing the observations against the results of simulations of the flood. This is not an ideal evaluation method and is somewhat self-referential given that flood models are calibrated using the same raw observational dataset. It is, however, a worthwhile
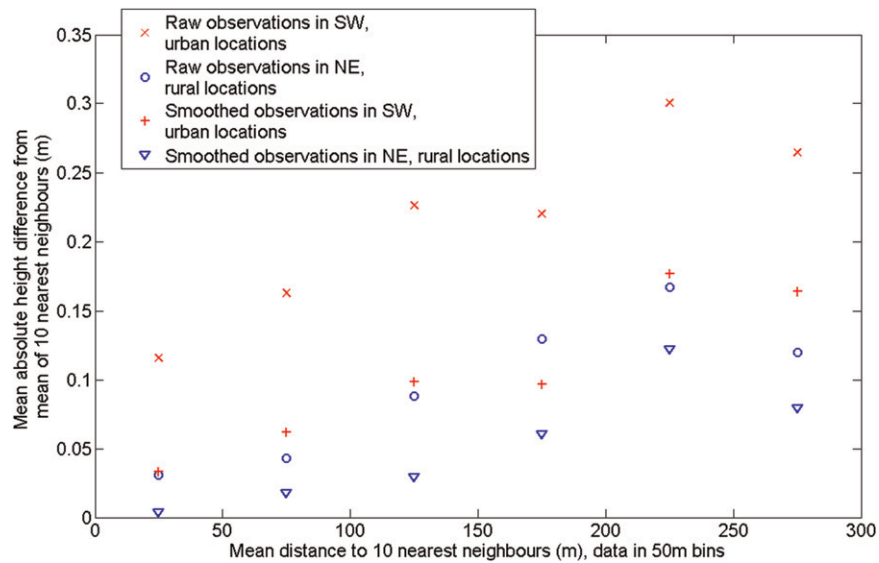
Figure 7. Mean absolute height difference between data points and their 10nn, collated into 50-m bins. Observations are split between those to the north and east of the study area (in rural locations) and those in the south and west, in more urban areas. Raw (red x's and crosses) and smoothed (blue circles and triangles) data are denoted.

exercise if it highlights limitations in the smoothing method and/or the model calibration process. While simple models based on interpolating surface elevations in the channel sometimes provide useful benchmarks for more sophisticated 2D models (see, e.g., Mason et al. 2009), in this case the topography and the complexity of three separate flood waves in three water courses mean the simple interpolation of water surface elevations does not perform well. As such, the output of a 2D hydrodynamic flood model of the event is used to evaluate the smoothing algorithm. Here we use the LISFLOOD-FP model developed by Bates and De Roo (Bates and De Roo 2000) with the inertial formulation of the shallow-water equations described in Bates et al. (Bates et al. 2010) run at 10-m resolution. The simulations use a DEM resampled to 10 m from a 1-m horizontal-resolution lidar dataset of the site. Buildings and vegetation were removed from the lidar topology to give the ''bare earth'' digital terrain model (DTM) and then the buildings reinserted using digital map data as described in Mason et al. (Mason et al. 2007b). The lidar dataset has an estimated vertical accuracy of between 0.05 and 0.15 m (Geomatics Group 2011). However, the postprocessing required to create the DEM can be problematic in heavily vegetated areas. As a result the RMSE of the DEM is estimated to be 0.18 m (Mason et al. 2007b). Neal et al. (Neal et al. 2013) resampled (using the nearest-neighbor technique) the DEM to 5- and 10-m resolutions and found only minor improvements in model performance (<0.02 m RMSE) when running at the higher resolution. This is consistent with a point made by Bates (Bates 2012) that correct representation of dynamic wetting and drying tends to require a higher resolution than modeling maximum flood extent, as is the
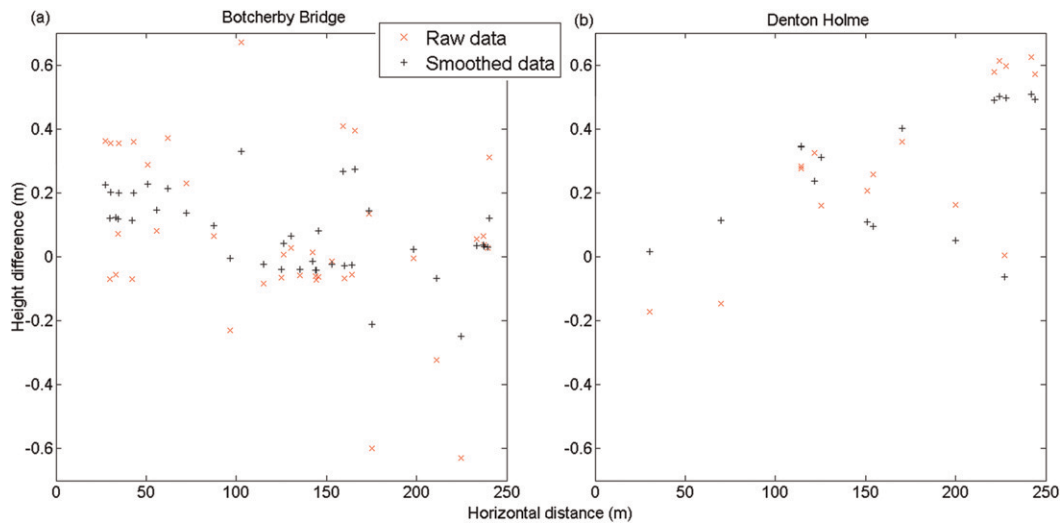
Figure 8. Comparison of observed maximum water heights to maximum hater height recorded by (a) Botcherby Bridge and (b) Denton Holme river gauges. Raw (red x's) and smoothed (black crosses) data are denoted.

case here. For that reason the 10-m-resolution model was considered sufficient for this study. Here we use the output from a Monte Carlo simulation of 999 model runs to assess the effect the correction algorithm might have when evaluating the flood model. Because of the unreliability of the river gauges discussed in section 2, the discharge figures for the three rivers are varied as part of the Monte Carlo simulation to account for the uncertainty in the upstream boundary conditions. Consequently, the following five model parameters were varied using a uniform Latin hypercube sampling method to generate the 999 parameter sets:

1) Channel roughness (Manning's n): A single global value is used across the whole model domain, varied between and 0.03 and 0.09 $m^{1/3} s^{-1}$.
2) Floodplain roughness (Manning's n): A single global value is used across the whole model domain, varied between 0.03 and 0.09 $m^{1/3} s^{-1}$.
3) Eden discharge multiplier: A multiplier was applied to all upstream Eden discharge values above a certain base flow level (200 $m^3 s^{-1}$). The portion of the discharge above the minimum was multiplied by the discharge multiplier. The discharge multiplier was varied between 0.5 and 1.5.
4) Caldew discharge multiplier: A multiplier is applied to discharge on the Caldew above a base value (15 $m^3 s^{-1}$). The discharge multiplier was varied between 0.5 and 1.5.

Table 3. Mean difference between peak water level recorded by river gauges and observations near to the gauges.

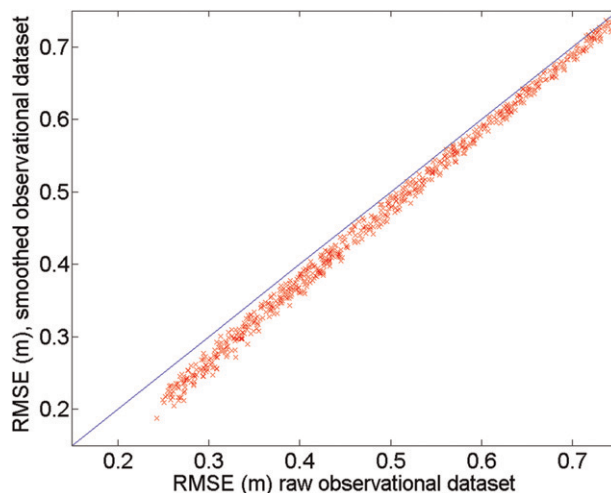| | Mean absolute height difference (m) | | |
Gauge name	Raw data	Smoothed data	No. of observations
Botcherby Bridge	0.18	0.11	40
Denton Holme	0.33	0.29	16

Figure 9. RMSE scores for model simulations: raw observational dataset vs smoothed observational dataset.

    5) Petteril discharge multiplier: A multiplier applied to discharge on the Petteril above a base value ($10 \, m^3 \, s^{-1}$). The discharge multiplier was varied between 0.5 and 1.5.

The maximum water extent from each simulation was then compared to the water extent as recorded by the observations of wrack and water marks. The location and height of the wrack and water marks are compared to the maximum simulated water levels at the locations of the wrack and water marks. The RMSEs between the simulated and observed water levels are calculated to give a single score for each simulation. If the simulated water extent does not reach the location of the wrack or water mark, the water level is taken from the nearest point that is inundated in the simulation. This process is repeated for the smoothed observational dataset. Figure 9 compares the RMSE using the uncorrected data with the RMSE from the smoothed dataset; it shows that all simulations performing well improve when scored against the smoothed observational dataset. While it cannot be claimed that an improvement in simulations scores necessarily proves that the smoothed observational dataset is a closer representation of reality than the original or "raw" observational dataset, the universal improvement in RMSE for every simulation with an RMSE of less than 0.6 m strongly suggests some improvement in the validation dataset. The RMSE for the best performing simulation improves from 0.243 to 0.187 m, a value very close to the estimated uncertainty in the DEM.

The correction algorithm is designed to reduce the error in anomalous data points if they are in the proximity of several other points that are roughly consistent, and it will also smooth a series of nearby points if they vary randomly around a mean value. However, it is unlikely to identify or correct errors such as any systematic bias that may be present and, as mentioned previously, situations can be envisaged where considerable differences in maximum water height are conceivable over short distances, in which case the smoothing algorithm may be making the observations less accurate. These scenarios are addressed below.

Figure 10. A water mark and a wrack mark deposited on a tree trunk by a flood in Tewkesbury, United Kingdom, May 2012. The water mark appears to exceed the wrack mark by some 5 cm; this could be due to capillary action of the tree bark. The top of the line of debris is approximately 15 cm above the water. Photo taken 3 May 2012.

4.4. Apparent bias in observations

The physical processes involved in creating a wrack mark or a water mark have not been exhaustively studied, and there may be ways that evidence of a high water mark may mislead the observer. Neal et al. (Neal et al. 2009) point out that a wrack mark can be left whenever debris-laden water sits unchanging for a period, so an observation of a wrack mark may not always correspond to the maximum water extent. Similarly, wind and wave action could push both wrack and water marks above the true maximum water level and, specifically for water marks, capillary effects of the material on which the mark is left could raise the apparent water level (Neal et al. 2009): an example of this effect can be seen in Figure 10. Occasional, random errors introduced in this way should be effectively corrected by the algorithm proposed above, but there is a suggestion of a nonrandom systematic discrepancy between the height of water marks and wrack marks, which may indicate a bias in either (or indeed both) types of observation. Figure 11 shows that the total mean error for simulations measured against wrack marks tends to be roughly 0.25 m higher than the corresponding error for water marks, suggesting that water marks may be, on average, 0.25 m higher than wrack marks.

The numerical dominance of wrack marks (217) over water marks (43) might suggest that the bias between the two types of observation would be hidden by the smoothing algorithm, and this might well be the case if the water marks were geographically interspersed among the wrack marks. However, as seen in Figure 2, the water marks are almost all located in the urban areas south of the Eden, with a particularly dense cluster west of the Petteril. This close geographical cluster in particular suggests that a systematic difference would not be hidden by the smoothing algorithm, and this can be seen in Figure 12, where the difference in
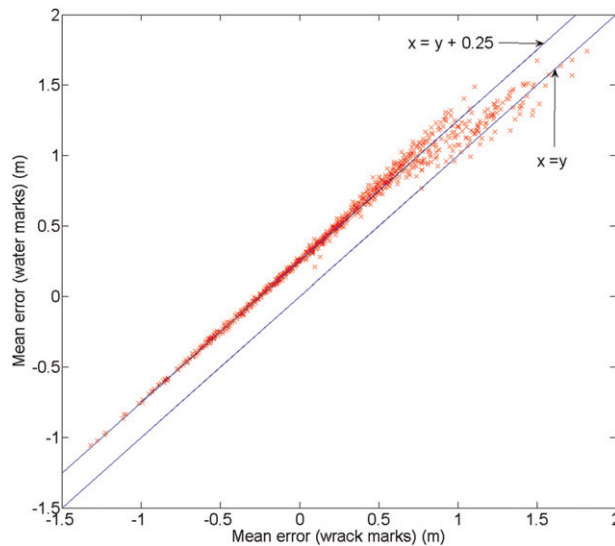
Figure 11. Mean error for all 999 model simulations measured using wrack mark observations (abscissa) and water mark observations (ordinate).

maximum water height is shown for the best performing simulation against both raw and smoothed observational data. Figure 12 shows that many of the isolated larger errors are considerably reduced by the smoothing process, but where the errors are clustered together they largely remain intact. For example, the cluster of water marks to the west of the Petteril (see highlighted area 1 in Figure 12a), which are observed to be above the simulated water level and the nearby wrack marks, are not significantly reduced. Indeed, it is this cluster of points that account for most of the apparent bias between the wrack and water marks. No obvious reason for this cluster of "raised" water marks presents itself, but many factors may have contributed:

- The water marks are located in a densely urban area where there may be localized hydraulic effects not represented by the simulation.
- As reported by the Environment Agency (Environment Agency 2005), there were reports of multiple debris blockages on the nearby River Petteril, which can cause significant variation in water height over small distances.
- The flooding near the Petteril seemed to have been caused by the high discharge from the Petteril itself but also from backwatering effects from the Eden when the main flood wave arrived. As Figure 3 shows, the timing of peak discharge on the Petteril and Caldew preceded the peak discharge on the Eden by some 10–12 h; it could be that we are attempting to compare marks of maximum water levels deposited at different times.
- The observed water marks are measured to be higher than most nearby observations. It could be that the floodwater depth was locally increased from upsurge from a drainage system.
- Finally, it is likely that there are errors in identifying and recording the location of the marks, as previously discussed.
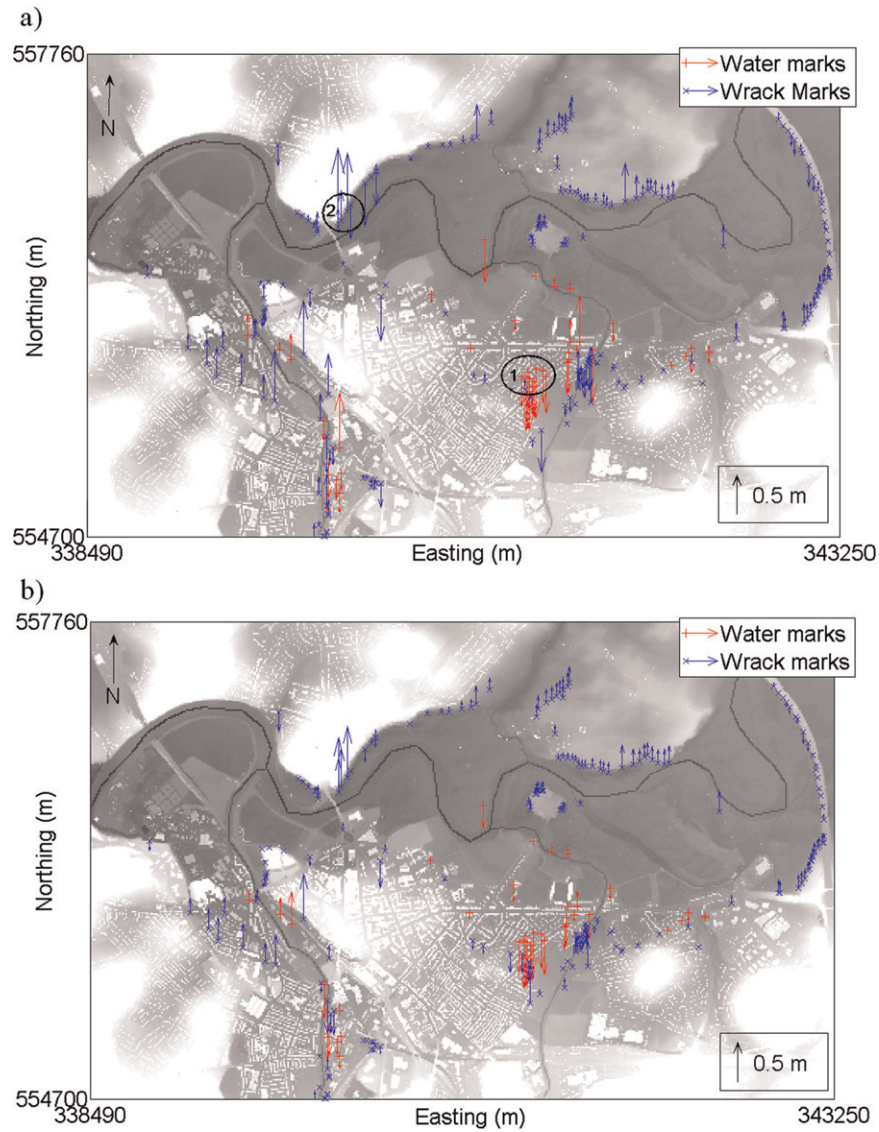
Figure 12. DEM of Carlisle showing how the best performing simulation compares against wrack (x's) and water mark (crosses) observations. Up arrows indicate the simulation water depth exceeds the observation. (a) Raw observational data and (b) smoothed observational data. Highlighted in (a) is 1) the cluster of water marks to the West of the Petteril (discussed in section 4.4) and 2) the water marks upstream of the A7 bridge (discussed in section 4.5).

4.5. Localized hydraulic effects

Figure 12b shows how the differences between simulated water heights and observations of wrack marks located in rural areas to the north and east of the study area are noticeably reduced by the correction algorithm. This suggests that the
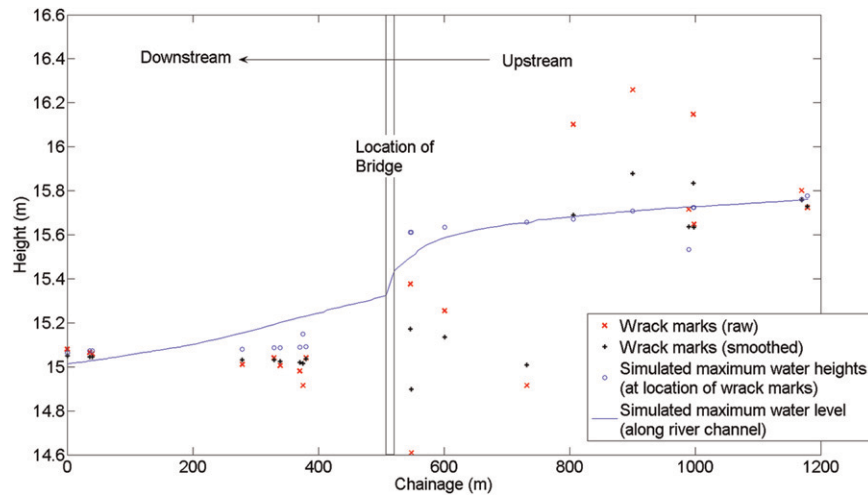
Figure 13. Height of uncorrected (x's) and smoothed (crosses) wrack mark observations near the A7 bridge on the River Eden plotted against river chainage. Also shown are the simulated maximum water levels from the best performing simulation.

smoothing algorithm may provide the most benefit in rural areas where there are densely populated series of homogeneous measurement types unaffected by the complicating factors of urban flooding. The notable exceptions are the four observations immediately to the east of the large bridge over the River Eden (see highlighted area 2 in Figure 12a). These four observations are well below both the observations immediately upstream and the simulated water level at this point (see Figure 13).

Figure 13 clearly shows the effect of the bridge on the simulated water levels. It can be seen that the observations downstream are consistent with each other and the simulation. However, this is not the case for the observations immediately upstream of the bridge. There seems to be considerable error present in some of the observations upstream from the bridge where recorded water levels between adjacent observations differ by up to 1.1 m. Although the smoothing algorithm appears to bring some of the observations closer to the simulated levels, it did not improve the four observations immediately upstream from the bridge.

From a physical perspective the expected hydraulic effect of the bridge should be to reduce maximum water level moving from upstream to downstream because of the narrowing of the channel (as the simulated water level shows), but this effect would not be manifest in the smoothing algorithm. Instead, the algorithm would use the observations downstream of the bridge to adjust the upstream points downward. Since the smoothing algorithm is removing information from the observational dataset, there are likely to be particular contexts where the effect will be to reduce, rather than increase, the accuracy of the measurements. This appears to be happening for at least two of the observations above the bridge, but the rather high inconsistency between neighboring observations near the bridge makes it hard to isolate or quantify a negative effect. No obvious reason for the inconsistency in the observations above the bridge presents itself; the observations are located on a

sloping bank sparsely populated with trees and shrubs, so it could be that the slope of the bank and the vegetation made it difficult to identify the wrack marks correctly. Alternatively, the presence of the bridge itself could be the cause by slowing down the dewatering process upstream such that debris was more likely to have been deposited by the descending floodwater below the maximum water level. By ascertaining that, for these particular readings, the smoothing algorithm appears to have little or no beneficial effect it may be decided, after consideration of the characteristics and location of these readings, that the value of the dataset is improved by discounting these readings entirely.

4.6. Possible impact on model uncertainty

Much of the analysis in sections 4.3 and 4.4 has focused on the single best performing model simulation as scored using the RMSE of all observations. However, it has long been recognized that the uncertainties inherent in environmental modeling mean there are likely to be a number of parameter sets equally acceptable in simulating the observed behavior (Beven 2006), a concept now known as equifinality (Beven and Freer 2001). The generalized likelihood uncertainty estimation (GLUE) technique proposed by Beven and Binley (Beven and Binley 1992) is a widely used method for identifying the acceptable subset of the parameter space (Beven 2006). Aronica et al. (Aronica et al. 2002) extended the GLUE technique by proposing a way of weighting model parameter sets based on the ability of the simulations to match flood extent data, then producing 2D probability maps to give spatially distributed estimates of uncertainty. These probabilistic flood maps are a powerful way of visualizing the uncertainty in the predictive precision of flood models. Here we establish the likely effect the smoothing algorithm would have on probabilistic flood maps produced as a result of the simulations performed in this study.

Beven (Beven 2006) points out that the choice of weighting scheme and the decision of whether a parameter set is considered behavioral are often subjective. They are also intrinsically linked to the uncertainty in the observational datasets (Schumann et al. 2007b), so if the proposed smoothing algorithm does indeed reduce the uncertainty in the observational dataset it may also impact which model parameter sets are considered behavioral. To establish if this is the case, we took an approach commented on by Refsgaard et al. (Refsgaard et al. 2006), which is to compare the RMSE of the simulation with the observational uncertainty. In this case, if the model is to have skill at predicting inundation depths for future flood events in the domain, the uncertainty in the DEM must be also be considered and combined with the observational uncertainty of the wrack and water mark readings. Mason et al. (Mason et al. 2007b) estimate the RMSE of the DEM U_{DEM} to be 0.18 m. In section 3.1 we compared wrack and water mark height readings with nearest neighbors to give an estimate of observational uncertainty U_{obs} of 0.105 m for the raw readings which is reduced to 0.038 m by the smoothing algorithm. Since these errors are statistically independent, the total observational uncertainty U is given by the root sum square,

$$U = \sqrt{U_{\mathrm{DEM}}^2 + U_{\mathrm{obs}}^2}. \tag{3}$$

Equation (3) gives values for U of 0.206 m using the raw readings and 0.184 m for the smoothed readings. Scoring the parameter sets of the 999 model runs in the Monte Carlo simulation and rejecting those where the RMSE is greater than twice the uncertainty results in a smaller number of parameter sets considered behavioral (235) when using the smoothed data than the raw data (267). The possibility, shown by these results, of using the smoothing algorithm to narrow the selection of behavioral parameter sets demonstrates its potential value in increasing the precision of predictions from flood models.

5. Discussion and conclusions

This paper has examined the accuracy of a set of point observations of high water marks left after a severe flood in Carlisle, United Kingdom, in January 2005. While it is clear there are errors in the observations, the lack of reference data makes it difficult to quantify and identify the inaccuracies. A smoothing algorithm is described that highlights the most inconsistent observations and can be applied to the dataset to reduce the inconsistency between observations and their nearest neighbors. The efficacy of the smoothing algorithm is evaluated by comparing the "raw" and "corrected" measurements against the peak water level recorded by two river gauges in the study area, and the results suggest the correction algorithm is broadly successful at reducing the error in the observational dataset. The algorithm was assessed through the use of a set of Monte Carlo simulations of a hydraulic model of the event showing how the smoothing algorithm can lead to improvements in model predictions by reducing model uncertainty. Furthermore, two areas were highlighted that showed distinct local inconsistency that would not be removed by applying the smoothing algorithm, providing the additional benefit of better informing the researcher to make the necessary subjective inferences on the data. Ideally, flood modelers would not need to introduce such subjectivity; however, in reality, the limitations of the models and data often necessitate it, as discussed by Pappenberger et al. (Pappenberger et al. 2007) and Schumann et al. (Schumann et al. 2009).

While we accept that if the smoothing algorithm is applied without due consideration of the characteristics of the particular study event, there is the possibility that it may reduce the overall information content in the dataset. However, given the inherent scarcity of and unavoidable uncertainty in observational data of extreme floods, any technique that either improves data quality or, at minimum, assists in highlighting and quantifying errors is potentially of great benefit to researchers. Furthermore, the spread of GPS-enabled camera phones and other consumer devices is enabling the collation of point observations from the public (see, e.g., Met Office 2012; Parkin 2010). The resulting datasets are likely to be prone to stochastic uncertainty, thus increasing the necessity of data cleansing techniques such as the smoothing algorithm described here. Further examination of the method in other case studies would be welcomed, especially where extensive validation from an alternative source such as aerial photos or SAR images is available.

Influence on water resources infrastructure and managing risks

This paper highlights the hydraulic complexity of modeling floods in built-up, high-risk areas where the hazard of flooding is greatest. The difficulties do not just

arise from the need to run simulations at a high enough resolution to reflect the small-scale hydraulic effects caused by buildings, bridges, and existing flood mitigation infrastructure. Other factors to consider include the following: the practice of using a single, global figure for the floodplain roughness means that, although concreted urban areas would have a very low friction in contrast to areas under vegetation, these differences are not adequately represented by a global friction parameter; subsurface features such as culverts and drains that may not be included in the model will affect localized water levels; remotely sensed observations of water extent or indeed in situ observations can be impeded by tall structures (Horritt et al. 2010; Schumann et al. 2011); and the relative abundance of bridges and artificial channel constraint in urban areas means channel blockages are more likely to occur during a flood, and this will influence the dynamics of the flood but is currently beyond the scope of hydraulic modeling codes. In the case of the 2005 Carlisle flood, it is likely that some combination of these factors is the reason the model uncertainty is higher in the urban than the rural areas. From the perspective of water resource infrastructure planning and flood risk management, it is certainly a priority to reduce uncertainty in high-risk areas by continuing the recent advances in monitoring, understanding, and simulating urban flooding.

References

Aronica, G., P. Bates, and M. Horritt, 2002: Assessing the uncertainty in distributed model predictions using observed binary pattern information within GLUE. *Hydrol. Processes,* **16,** 2001–2016.

Barker, P. A., R. Wilby, and J. Borrows, 2004: A 200-year precipitation index for the central English lake district. *Hydrol. Sci. J.,* **49,** 769–785.

Bates, P. D., 2012: Integrating remote sensing data with flood inundation models: How far have we got? *Hydrol. Processes,* **26,** 2515–2521.

——, and A. De Roo, 2000: A simple raster-based model for flood inundation simulation. *J. Hydrol.,* **236,** 54–77.

——, M. S. Horritt, and T. J. Fewtrell, 2010: A simple inertial formulation of the shallow water equations for efficient two-dimensional flood inundation modelling. *J. Hydrol.,* **387,** 33–45.

Beven, K., 2006: A manifesto for the equifinality thesis. *J. Hydrol.,* **320,** 18–36.

——, 2008: *Environmental Modelling: An Uncertain Future?* Taylor & Francis, 328 pp.

——, and A. Binley, 1992: Future of distributed models: Model calibration and uncertainty prediction. *Hydrol. Processes,* **6,** 279–298.

——, and J. Freer, 2001: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology. *J. Hydrol.,* **249,** 11–29.

Bhavsar, P., 1984: Review of remote sensing applications in hydrology and water resources management in India. *Adv. Space Res.,* **4,** 193–200.

Biggin, D., 1996: A comparison of ERS 1 satellite radar and aerial photography for river flood mapping. *Water Environ. J.,* **10,** 59–64.

Brakenridge, G., B. Tracy, and J. Knox, 1998: Orbital SAR remote sensing of a river flood wave. *Int. J. Remote Sens.,* **19,** 1439–1445.

Connell, R., C. Beffa, and D. Painter, 1998: Comparison of observations by flood plain residents with results from a two-dimensional flood plain model: Waihao River, New Zealand. *J. Hydrol.,* **37,** 55–79.

Dawson, R., J. Hall, P. Sayers, P. Bates, and C. Rosu, 2005: Sampling-based flood risk analysis for fluvial dike systems. *Stochastic Environ. Res. Risk Assess.,* **19,** 388–402.

Day, A.-L., 2005: Carlisle storms and associated flooding: Multi-agency debrief report. U.K. Resilience Tech. Rep., 163 pp.

Di Baldassarre, G., and A. Montanari, 2010: Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. Earth Syst. Sci.,* **13,** 913.

Dottori, F., and E. Todini, 2013: Testing a simple 2D hydraulic model in an urban flood experiment. *Hydrol. Processes,* **27,** 1301–1320, doi:10.1002/hyp.9370.

Dung, N., B. Merz, A. Bárdossy, T. Thang, and H. Apel, 2011: Multi-objective automatic calibration of hydrodynamic models utilizing inundation maps and gauge data. *Hydrol. Earth Syst. Sci.,* **15,** 1339.

Environment Agency, 2005: Dealing with flooding: A review of the floods in northern England and north Wales January 2005. Environment Agency Rep., 20 pp.

——, 2006: Cumbria floods technical report: Factual report on the meteorology, hydrology, and impacts of the January 2005 flooding in Cumbria. Environment Agency Rep., 182 pp.

Fewtrell, T. J., J. C. Neal, P. D. Bates, and P. J. Harrison, 2011: Geometric and structural river channel complexity and the prediction of urban inundation. *Hydrol. Processes,* **25,** 3173–3186.

Geomatics Group, cited 2011: Lidar product description. [Available online at http://www.geomatics-group.co.uk/GeoCMS/Products/LIDAR.aspx.]

Horritt, M., and P. Bates, 2001: Predicting floodplain inundation: Raster based modelling versus the finite element approach. *Hydrol. Processes,* **15,** 825–842.

——, and ——, 2002: Evaluation of 1D and 2D numerical models for predicting river flood inundation. *J. Hydrol.,* **268,** 87–99.

——, D. Mason, and A. Luckman, 2001: Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *Int. J. Remote Sens.,* **22,** 2489–2507.

——, P. Bates, T. Fewtrell, D. C. Mason, and M. Wilson, 2010: Modelling the hydraulics of the Carlisle 2005 flood event. *Proc. ICE Water Manage.,* **163,** 273–281.

Hunter, N. M., P. D. Bates, M. S. Horritt, A. De Roo, and M. G. F. Werner, 2005: Utility of different data types for calibrating flood inundation models within a GLUE framework. *Hydrol. Earth Syst. Sci.,* **9,** 412–430.

——, ——, ——, and M. D. Wilson, 2006: Improved simulation of flood flows using storage cell models. *Water Manage.,* **159,** 9–18.

——, and Coauthors, 2008: Benchmarking 2D hydraulic models for urban flooding. *Water Manage.,* **161,** 13–30.

Lane, S. N., T. D. James, H. Pritchard, and M. Saunders, 2003: Photogrammetric and laser altimetric reconstruction of water levels for extreme flood event analysis. *Photogramm. Rec.,* **18,** 293–307.

Leedal, D., J. Neal, K. Beven, P. Young, and P. Bates, 2010: Visualization approaches for communicating real time flood forecasting level and inundation information. *J. Flood Risk Manage.,* **3,** 140–150.

Lhomme, J., J. Gutierrez-Andres, A. Weisgerber, M. Davison, J. Mulet-Marti, A. Cooper, and B. Gouldby, 2010: Testing a new two-dimensional flood modelling system: analytical tests and application to a flood event. *J. Flood Risk Manage.,* **3,** 33–51.

Marcus, W. A., and M. A. Fonstad, 2008: Optical remote mapping of rivers at sub meter resolutions and watershed extents. *Earth Surf. Processes Landforms,* **33,** 4–24.

Mason, D. C., M. S. Horritt, J. T. Dall'Amico, T. R. Scott, and P. D. Bates, 2007a: Improving river flood extent delineation from synthetic aperture radar using airborne laser altimetry. *IEEE Trans. Geosci. Remote Sens.,* **45,** 3932–3943.

——, ——, N. M. Hunter, and P. D. Bates, 2007b: Use of fused airborne scanning laser altimetry and digital map data for urban flood modelling. *Hydrol. Processes,* **21,** 1436–1447.

——, P. Bates, and J. Dall'Amico, 2009: Calibration of uncertain flood inundation models using remotely sensed water levels. *J. Hydrol.,* **368,** 224–236.

Matgen, P., and Coauthors, 2010: Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the particle filter: Proof of concept. *Hydrol. Earth Syst. Sci.,* **14,** 1773–1785.

Met Office, cited 2012: Met Office Weather Observations website WOW. [Available online at http://wow.metoffice.gov.uk/home.]

Mignot, E., A. Paquier, and S. Haider, 2006: Modeling floods in a dense urban area using 2D shallow water equations. *J. Hydrol.,* **327,** 186–199.

Neal, J. C., P. D. Bates, T. J. Fewtrell, N. M. Hunter, M. D. Wilson, and M. S. Horritt, 2009: Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. *J. Hydrol.,* **368,** 42–55.

——, G. Schumann, T. Fewtrell, M. Budimir, P. Bates, and D. Mason, 2011: Evaluating a new LISFLOOD FP formulation with data from the summer 2007 floods in Tewkesbury, UK. *J. Flood Risk Manage.,* **4,** 88–95.

——, C. Keef, B. C. Bates, K. Beven, and D. Leedal, 2013: Probabilistic flood risk mapping including spatial dependence. *Hydrol. Processes,* **27,** 1349–1363. doi:10.1002/hyp.9572.

Pappenberger, F., P. Matgen, K. J. Beven, J. B. Henry, and L. Pfister, 2006: Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Adv. Water Resour.,* **29,** 1430–1449.

——, K. Beven, K. Frodsham, R. Romanowicz, and P. Matgen, 2007: Grasping the unavoidable subjectivity in calibration of flood inundation models: A vulnerability weighted approach. *J. Hydrol.,* **333,** 275–287.

Parkin, G., 2010: The September 2008 Morpeth flood: Information gathering for dynamic model reconstruction. Natural Environment Research Council Summary Rep., 25 pp.

Pender, G., and S. Néelz, 2010: Flood inundation modelling to support flood risk management. *Flood Risk Science and Management,* G. Pender and H. Faulkner, Eds., Blackwell, 234–257.

Porter, J., 2010: *The Extreme Flood Outline Map: Co-Producing Flood Risk Mapping and Spatial Planning in England.* King's College London Department of Geography, 520 pp.

——, and D. Demeritt, 2012: Flood risk management, mapping and planning: The institutional politics of decision support in England. *Environ. Plann.,* **44,** 2359–2378.

Rayburg, S., M. Thoms, and M. Neave, 2009: A comparison of digital elevation models generated from different data sources. *Geomorphology,* **106,** 261–270.

Refsgaard, J. C., J. P. Van der Sluijs, J. Brown, and P. Van der Keur, 2006: A framework for dealing with uncertainty due to model structure error. *Adv. Water Resour.,* **29,** 1586–1597.

Saint-Venant, B., 1871: Theory of unsteady water flow, with application to river floods and to propagation of tides in river channels. *French Acad. Sci.,* **73,** 237–240.

Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, 2008: *Global Sensitivity Analysis: The Primer.* Wiley, 304 pp.

Schumann, G., R. Hostache, C. Puech, L. Hoffmann, P. Matgen, F. Pappenberger, and L. Pfister, 2007a: High-resolution 3-D flood information from radar imagery for flood hazard management. *IEEE Trans. Geosci. Remote Sens.,* **45,** 1715–1725.

——, P. Matgen, L. Hoffmann, R. Hostache, F. Pappenberger, and L. Pfister, 2007b: Deriving distributed roughness values from satellite radar data for flood inundation modelling. *J. Hydrol.,* **344,** 96–111.

——, P. D. Bates, M. S. Horritt, P. Matgen, and F. Pappenberger, 2009: Progress in integration of remote sensing derived flood extent and stage data and hydraulic models. *Rev. Geophys.,* **47,** RG4001, doi:10.1029/2008RG000274.

——, J. C. Neal, D. C. Mason, and P. D. Bates, 2011: The accuracy of sequential aerial photography and SAR data for observing urban flood dynamics, a case study of the UK summer 2007 floods. *Remote Sens. Environ.,* **115,** 2536–2546.

Smith, M., E. Edwards, G. Priestnall, and P. Bates, 2006: Exploitation of new data types to create digital surface models for flood inundation modelling. FRMRC Research Rep. UR3, 86 pp.