Anonymous Referee #1 -

GENERAL COMMENTS

I found many drawbacks mainly due to the paper organization and presentation that have to be solved by the authors before it can be considered adequate for a publication in the HESS journal. Moreover there are some key points that I would like to underline.

One is the justification of the small sensitivity of the roughness parameter with respect to the channel bathymetry that to me seems reasonable but I found some difficulty in understanding if this is supported by a robust analysis or it is not adequately shown in the manuscript. To this end, it would be interesting to see the DYNIA analysis carried out also for the channel roughness in order to recognize its information content and its value of identifiability. If this was already done, but not shown, some comments or an explaining figure would be very welcome.

We did indeed carry out the DYNIA analysis on the channel roughness *'n_c'* parameter for all the SAR images used in this study but we found the results for this parameter were not as sensitive as those for the depth parameter '*r*'. Variations in channel roughness value did not produce a strong enough response in the model to match observed flood extent. For this reason these '*r*' results did not feature in the final version of the paper. The authors have amended the text in section 3.1 to explain this:

> *"Consequently an important result of this paper is that - in this particular experimental set up with channel roughness parameter 'n_c' examined simultaneously with the channel depth parameter 'r' for the available ENVISAT SAR data - 'n_c' has a much reduced sensitivity compared with the 'r' depth parameter response. It is observed that 'n_c' will yield optimal results for as long as 'r' is also unknown. This lack of sensitivity of channel roughness in this and all subsequent results meant that 'n_c' could not be identified with any real confidence with this methodology (while 'r' is also unknown). So while 'n_c' analysis was carried out we present here onwards only those results from the more identifiable 'r' parameter. 'n_c' results are now omitted (but can be provided upon request if of interest)".*

To illustrate we show some of the channel roughness results below. Here we show firstly the results for the *'r'* parameter (red), then 'n_c' (blue) CSI results, against parameter value for a sample of the SAR images. These CSI plots generally reflect the results we observed in all SAR data and illustrate the greater sensitivity of the depth parameter over the channel roughness parameter in our experimental results:

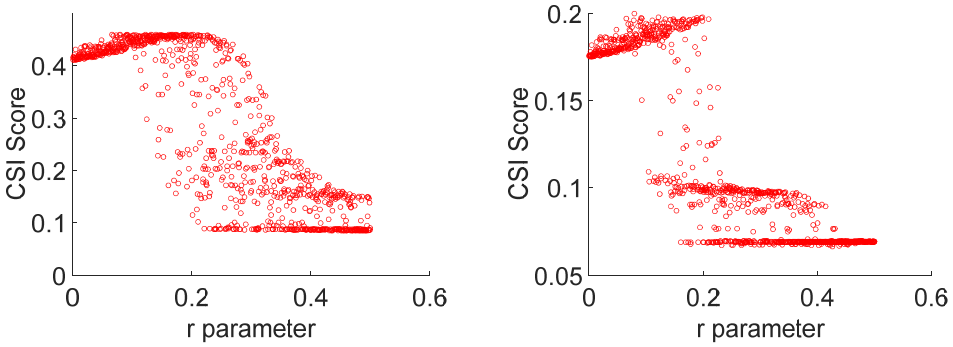


Figure 1 Plot of 'r' parameter against Critical Success Index score for left: 23rd July 2007 at 10:27 and right: 24th January 2008 at 10:12.

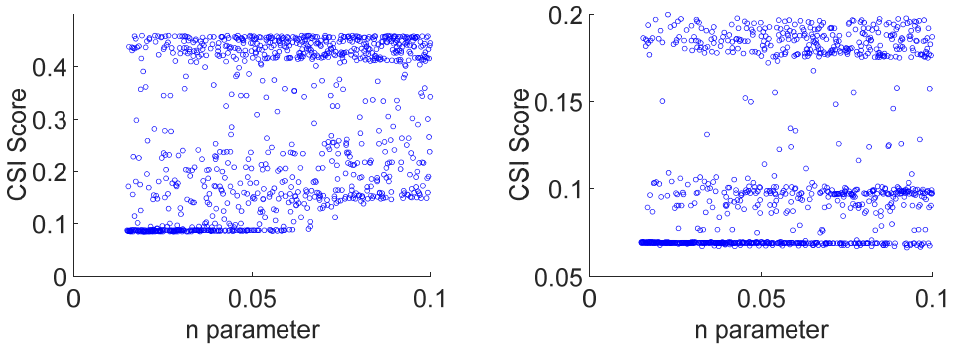


Figure 2 Plot of 'n_c' parameter against Critical Success Index score for left: 23rd July 2007 at 10:27 and right: 24th January 2008 at 10:12.

Additionally, Figure 3 below shows the cdf plot of the gradient of the cumulative distribution of rescaled support values for groupings of SAR data, against 'n_c' value. This illustrates how different/lower the results are for 'n_c' identifiability when compared with 'r' results (paper fig 7, inset).

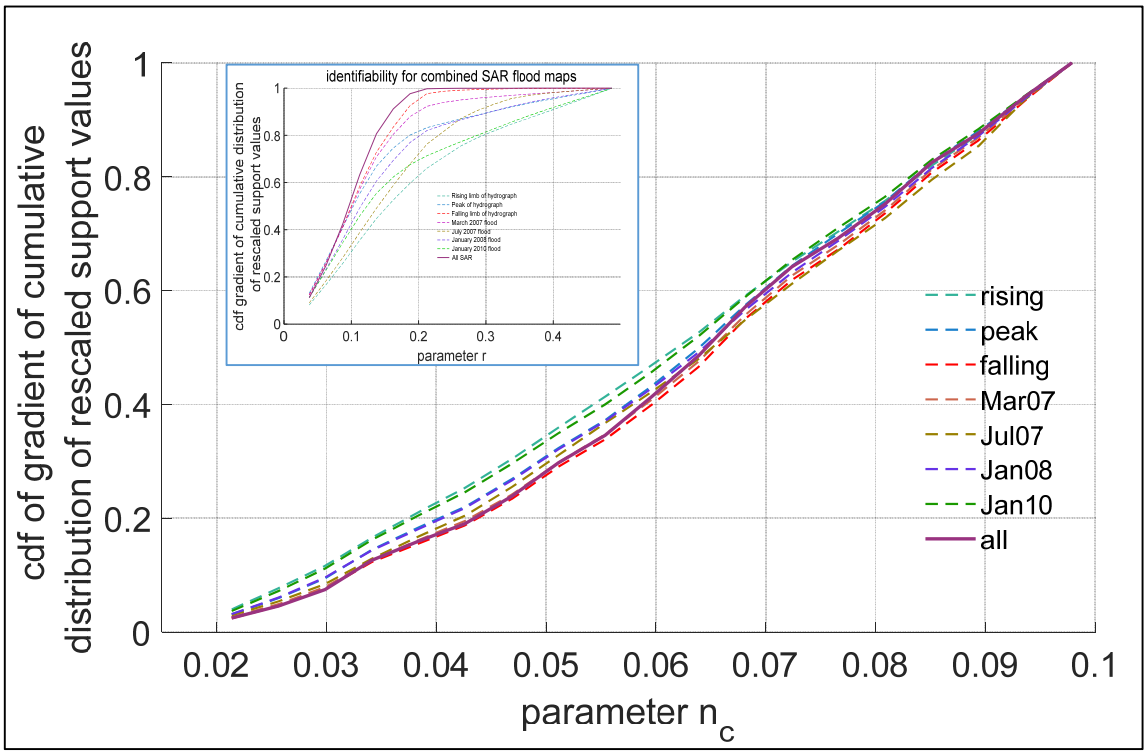


Figure 3 cdf of the gradient of the cumulative distribution of rescaled support values for each individual SAR image, against 'n_c' parameter value.

Lastly, the Information Content (IC) results for 'groupings' of SAR images are also shown here for 'n_c' and 'r' in Table 1. These illustrate how much lower the 'n_c' information content can be compared with the IC of the depth parameter, for the same combination of data.

Table 1 Information Content (IC) for groupings of SAR images, showing results for r and n_c.

'grouping'	Parameter 'r'	Parameter 'n_c'
Rising limb	0.13	0.11
Peak of hydrograph	0.23	0.12
Falling limb	0.64	0.16
March 07 event	0.50	0.14
July 07 event	0.37	0.14
January 08 event	0.25	0.12
January 10 event	0.13	0.11
All SAR [1-11]	0.68	0.16

A second point is the assumption that the error related to the processing of the SAR image will not affect the results (not considered for simplicity). I think that these errors are part of the procedure of identifiability and are able to affect the information content of the different images. For this point my question is: due to the different acquisition, times of the SAR images under different atmospheric and land conditions can be considered the error related to the image processing stationary? My opinion is that this error varies from image to image. This at least deserves some discussion. The authors could consider that aerial flood maps for analyzing this point or, since the area is very well instrumented, doing the same type of analysis with stage data.

This is a good point and certainly each SAR image will have a unique processing error associated with it and indeed the errors inherent in the processed SAR data will be passed on to the final identifiability and IC score for single images. These will not be stationary errors, they will vary between images.

For example, while different atmospheric conditions will not significantly affect the radar signal, different incident angles can have an effect. In the paper these were considered to be so small compared with the errors associated with the assumptions around parameter identifiability that they were thought to be overshadowed. Furthermore the use of moderate resolution flood imagery for hydraulic model calibration may lead to inaccuracies but it was deliberately chosen because we want to understand the usefulness of this data for global locations where other data may quite simply not be available. And while the magnitude of flooding and the land surfaces affected can cause specific errors/uncertainties, this is somewhat mitigated by the larger spatial scales employed in this analysis.

Other errors in the processing of the SAR image can be evident, such as from bias - where in some areas radar data does not inform on the extent of flooding. Ideally, such non-informative areas would be masked out but this requires more comprehensive analysis and is currently an active area of research (e.g. Giustarini *et al.,* submitted: puts forward the idea of flood 'probability' maps to illustrate confidence in the detected flood extent). There is now more discussion about this around Figure 3 in the paper which illustrates the CSI scores for the aerial flood map image that was acquired in July 2007. However since the authors already have some concerns about the length of the document, and the comments of reviewer #2 on the length of the paper also, we attempt to keep these discussions brief.

A final point is the quality of all the figures in the manuscript that I found very poor and such that to impede a proper understanding of the manuscript.

Thank you for this feedback, we have replaced the figures mentioned with new marker-line versions (greyscale) and increased the image resolution to improve the quality.

Based on that I recommend publication after major revisions. In the following, the authors can find a list of comments with the associated relevance listed in order of appearance in the manuscript.

COMMENTS

Pag. 2 Lines 53-76 MINOR: the authors may also cite the work of Moramarco et al. (2013) which uses an interesting method for identifying the flow depth distribution in natural channels.

Moramarco, T., Corato, G., Melone, F., Singh, V.P., 2013. An entropy-based method for determining the flow depth distribution in natural channels. Journal of Hydrology, 497,176-188.

This reference has been added to the paper.

Pag. 3 Lines 88-91 MINOR: Can you rephrase this sentence more clearly?

Lines 88-91 have been rephrased to:

> "*In particular the methodology uses flood extent with an accuracy-scoring method that disregards the correct detection of 'no water' pixels*"

Pag. 4 Lines 120-128 MAJOR: If I understand correctly the channel depth is expressed as H=r*B where H is the channel depth, and B is the width of the channel. Since the hypothesis of linear scaling is central in the study, I thinks this part deserves more profound discussion about: 1) How much it will affect the results of the study. 2) Which are the expected problematics associated with the uniform channel depth.

Extra discussion has been inserted at the end of the manuscript around the limitations of our assumptions and what affect these might have on the results.

See also the paper of Yan et al. (2014) where H is a free parameter of the model uniform along the river reach.

Yan, Kun, et al. "Exploring the potential of SRTM topography and radar altimetry to support flood propagation modeling: Danube case study." Journal of Hydrologic Engineering 20.2 (2014): 04014048.

This reference has been added to the paper.

Pag 3 Section 1.2. MINOR: A scheme or figure of the method would significantly help to understand the image-processing algorithm.

A good point, the authors have inserted a new figure to illustrate the steps of the methodology: Figure 1.

Pag. 6 Lines 211-213: MODERATE: it is not clear how the procedure is used with multiple images. Please provide more details.

A description of the procedure for combining image results has been updated in section 1.4:

> *"These group scores are determined by multiplying each single model/SAR flood map CSI result with the CSI score of the next SAR flood map until all members of the particular group have been added. The unique combinations which comprise these groups are described in Table 3 below. This combining of CSI scores is done for results from each of the 1000 models/parameter*

scenarios. The next step is the same as for single CSI scores as described above – i.e. to rescale the objective function and compute the cumulative sup

Pag 7 Figure 1 MINOR: The quality of this figure is very poor. Please provide a larger and cleared picture where the identification of the study area and the boundary conditions are more clearly visible.

Thank you for pointing this out, Figure 1 (now Figure 2) has now been amended to be clearer with boundary locations highlighted.

Pag 11 Figure 2 MODERATE: the quality and the description of this figure is very poor. Also, ENVISAT and Aerial data seem to be a bit different although with this picture is very difficult to compare the results. I understand that the processing of the SAR image inherently contain errors, I am wondering if the results of the paper might be affected by these errors. The authors could test the procedure also on the aerial photograph to understand the effect of the errors in the processing of the image or on the observed stages.

The observed model which is expected to behave better than the test model seems to be worse than the test model? Do you have a justification for that? Does this depend on the calibration?

A well observed point for Figure 2 (now Figure 3) in the paper. The description of this figure in the text has been updated so it is clearer why it was inserted. Also the CSI scores embedded in the figure have been moved to a table for easier comparison/interpretation for the reader. The figures themselves (particularly the modelled and ENVISAT flood extents) have a coarse resolution that does not reproduce very nicely on the page unfortunately.

This particular aerial flood extent image was derived from a single aerial photograph of the flood of July 2007 on the Severn by manual delineation and so is restricted both by the limits of the photographs (cutting off the upper River Avon for example) and interpretation of the image in terms of the flood boundary through vegetation. In this case we used the aerial data here more for validation purposes rather than explicitly to test the calibration methodology. The aim of the paper was to test a series of more 'moderate' resolution imagery that is more extensively and frequently available.

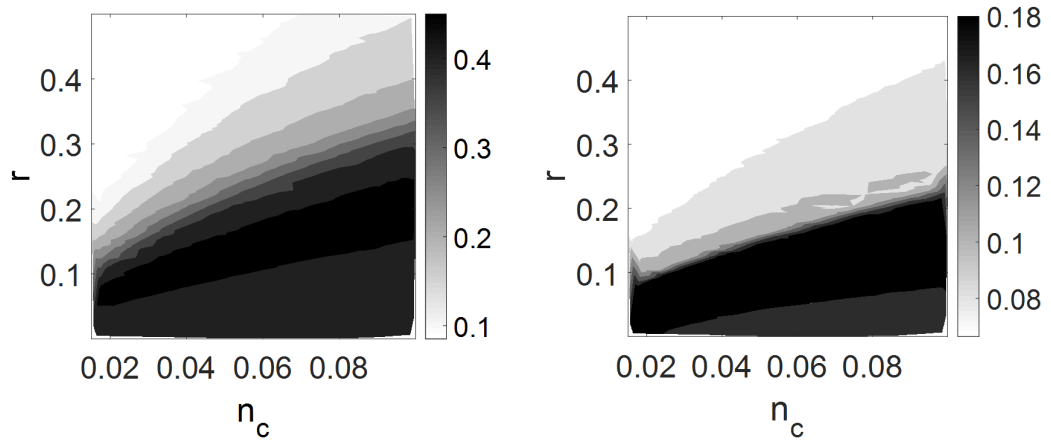
However the authors have added some additional text around the new table of aerial versus ENVISAT CSI scores to explain further the test results we found when applying the methodology on the single aerial flood extent data. A good comparison could be made between the methodology applied to the ENVISAT observation (acquired 23rd July 2007, 10:27am) and the subsequent aerial image (acquired 24th July 2007, 11:30am) as they were observing the flood merely 24hrs apart. While it is quite likely that the impact of SAR processing errors would be manifest in the final results, and be most obvious when compared with better resolution data such as can be obtained from a gauge record or aerial imagery, the authors felt that an explanation of the impact of these errors was beyond the scope of the paper (which is rather long already).

In Figure 2 (now Figure 3), it is obvious that the observed model (constructed using surveyed cross sections) has not represented observed flood extent as well as the test model. This is most evident in the tributaries to the main River Severn. We updated the text in section 3.1 of the paper to explain this:

"The scores and flood extent for the observed model are not better than the test model results as might be expected. This may be explained by the fact that while the bathymetry of the observed model does come from survey data, the (domain-average) channel roughness value is not calibrated in either model. While the test model had 1000 parameter-varying depth and roughness values, the observed model had a best estimate of domain-average channel roughness parameter (of 0.038). While appropriate for the main rivers, it is evident that the channel roughness value is not suitable for the narrower tributaries."

Pag 11 Figure 3 MAJOR: Please provide a better figure with colors. It is very difficult (with this figure) to follow the authors' statements.

This contour plot has now been updated in greyscale so it is easier to read with two single opposing colours than the original colour spectrum which made it difficult to see where CSI was highest and lowest, even in full colour. Plus this version should be more printer friendly in black and white.



From the manuscript: figure 4 : Single SAR acquisitions are compared with LISFLOOD-FP modelled flood maps for the July 2007 flood event. Left: results from the SAR acquisition on 23rd July 2007 at 10:27, right: result from the SAR acquisition 24th January 2008 at 10:12.

Pag 11 Section 3.1 MODERATE: I found this section very difficult to follow and to read. I suggest to try to present it better.

A re-write of this section has been carried out and the authors hope it is now easier to read and follow.

Pag 12 lines 346-349, MAJOR: the authors concluded that nc is insensitive when estimated simultaneously with the channel depth. However, it seems that this was concluded based only on two images (23rd July 10:27, and 17 January 2008 21:55). Do the authors exclude that this is true in any case and there are not effects of the time of acquisition and the magnitude of the flood event? I think the authors should provide more proofs for this statement. Overall, I find this assumption reasonable however I think that including the DYNIA also for the parameter nc would add a lot of value to the paper.

This paragraph has been rewritten within the context of the above comment on section 3.1 and hopefully now better explains why 'n_c' results were excluded from the final version of the paper. All

available SAR data were analysed for '*r*' and 'n_c' though only two '*r*' plots featured in the paper for simplicity. As the authors hope has been explained sufficiently within the general comments section above, the 'n_c' parameter did indeed undergo the same DYNIA analysis as the depth parameter but results were not exceptional enough to be included within an already long paper. The same lack of responsiveness of this parameter in the results were observed for all SAR data analysed, and even when the SAR data were 'grouped' the identifiability and IC results did not greatly improve for 'n_c' .

With this available dataset of ENVISAT SAR images and hydraulic model set up, the authors have concluded that 'n_c' is insensitive when estimated simultaneously with the channel depth parameter '*r*'. However additional testing would need to be carried out to conclusively say that this insensitivity is true with other SAR data and magnitudes of flooding.

Pag. 13 line 361. MODERATE: It is not clear how the IC score is calculated for multiple images.

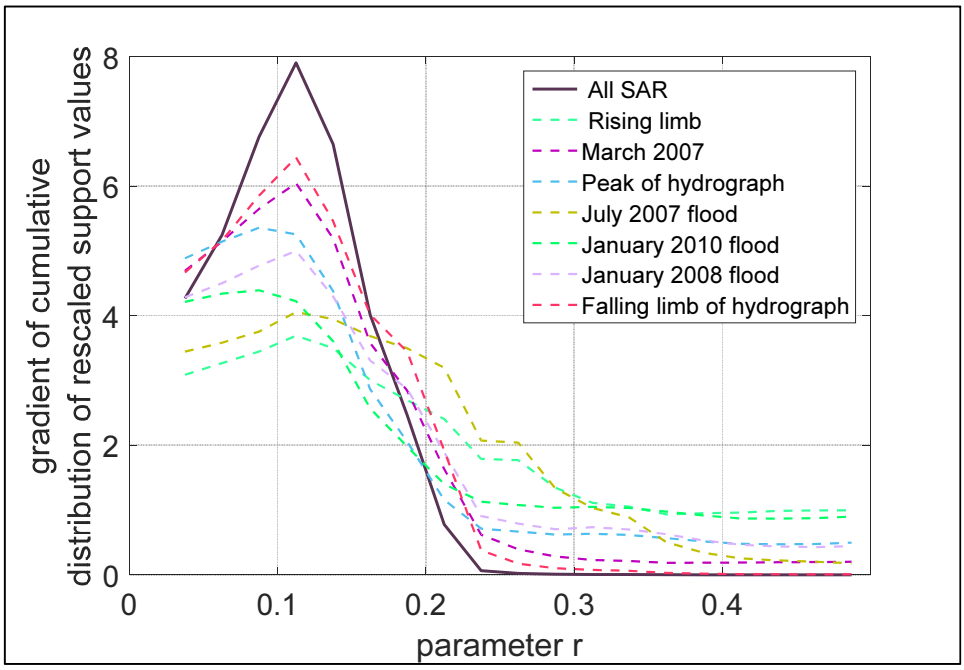
The grouping of SAR data occurs after each single model/SAR flood map is assessed and given a CSI score. It is the CSI scores which are multiplied together in particular combinations (as described in Table 3) before the DYNIA methodology is applied and identifiability and IC determined. A group IC score is derived from the grouped SAR data and CSI scores multiplied together. This explanation has been inserted in to the paper in section 3.2 and 1.4 for greater clarity.

Figures 4, 5, 6, 7 and 8 MAJOR: The interpretation of these figures must be described in the method section. From the text it is very difficult to follow the authors' statements. Please provide a better quality figures as well. It is impossible to discriminate between the different lines. If I understand correctly these are the cumulative distribution of the rescaled support values and not the gradient. The gradient should refer to their slopes. Isn't? If so, I expect a figure like the one in the paper of Wagener et al. (2003), (see FIGURE 8 in their paper).

Thank you for pointing out how difficult these figures are to read. The original figure coloured lines have been replaced with marker-line figures (greyscale) in this revised version and an explanation of how the cdf plots should be interpreted is now inserted into the method section.

It is correct that Figures 4 – 8 in the paper are representative of the (gradient of the) cumulative distribution of rescaled support values. The gradient plots were converted to cumulative distribution function (cdf) plots, and this is what was shown in the final version of the paper. This was a change introduced to make the plots easier to read together and normalised, when the original histogram plots were numerous and previously side-by-side for each grouping.

For this experiment, the SAR data acquisitions are not regular or plentiful through time (as in the experiment of Wagener e*t al.*, 2003 which used 6 years of continuous daily flow data) and so it was more difficult to represent in a line plot. However for purposes of explanation, the last cdf plot in the paper is converted into a 'gradient' plot for parameter '*r*' below (top figure) to mimic Figure 8 in the paper of Wagener *et al*., 2003. We have amended the Figures 4-8 (now Figures 5-9) within the paper so they more closely resemble the Wagener identifiability plots.



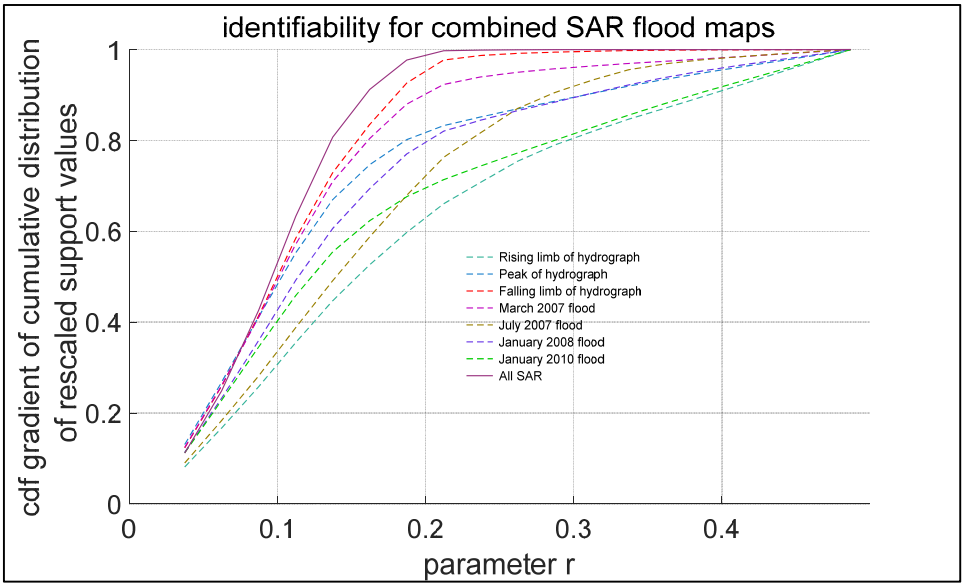


Figure 4 top: a new gradient of the cumulative distribution of rescaled support values for each SAR grouping, and bottom: the original cdf of the gradient of the cumulative distribution of rescaled support values for each SAR grouping, as featured in the paper.

Pag 18 lines 503. MODERATE: No colors can be seen in the figures.

Figures have been redone and text updated accordingly.

Pag 19 lines 545-556. MODERATE: I expect here some discussions about the possible consequences of the assumptions made in the paper.

This consequences of our assumptions in the paper have been expanded within the results and discussion sections, as described above.

Reviewer Guy Schumann (Reviewer #2)

The IC approach is really nice and gives an objective assessment of the value of a flood image for calibration. However, what we are still missing in the literature is to find a way that gives an objective IC of a SAR flood map without the need to calibrate first. In other words, in this paper, which I think has a lot of merit, IC is built up based on parameter identifiability rather than for instance inter-comparing each SAR image and applying the score and identifiability that way, so without the need of a model and its parameter but I understand that this is outside the scope of this paper.

It is an appealing idea to attribute IC to individual SAR data, without need of using models and parameters for calibration. At the moment the authors are not sure how to inter-compare SAR data to reveal information content but it could be an interesting topic of further study.

I also think that what is innovative here is the analysis of IC and identifiability in relation to what stage in the hydrograph we are looking at and what type of data we use (single image, combined images, gauge data). I wonder if the title and the introduction should better reflect that since to me this is one of the first papers to try and answer these questions using real data.

Thank you for raising this point. The authors have updated the title of the paper and reworded the abstract, so that these reflect more accurately the unique points within the paper.

My biggest reservation in this study lies with the choice of performance metric used, which may explain in my opinion why the greatest information content is in the SAR images closest to peak flow. Stephens et al. (2014) showed that the performance measure used here is particularly biased towards largest flooded area (in other words, it always gives the highest score to the biggest flooded area). This is significant in this study and could lead to an unwanted "bias" in the calibration. I suggest the authors repeat the exercise offline with the "F2" measure for instance ((A-B)/(A+B+C)) or an area in error index ((B+C)/(A+B+C+D)) to see if the same SAR images give the highest sensitivity still.

A valid point is raised here. In the preparation for this paper the authors did indeed prepare a number of 'skill score metrics' before deciding on the CSI skill score in preference over 'F' measures and other promising metrics such as Percentage Correct. A sample of these initial plots are shown here in Figure 5 to explain why in the end we decided to use the CSI results:

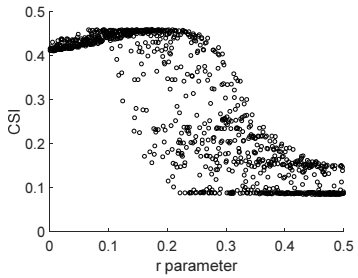
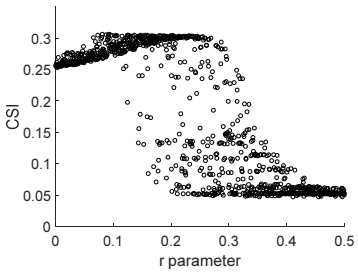
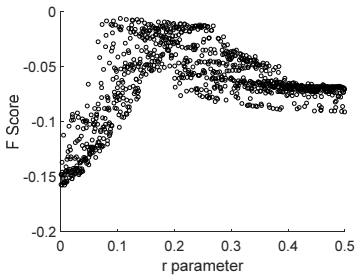
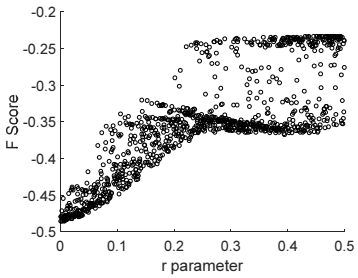
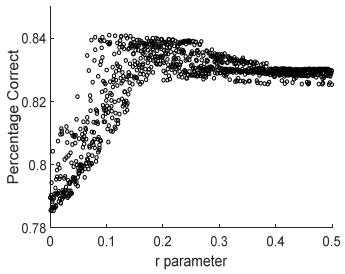
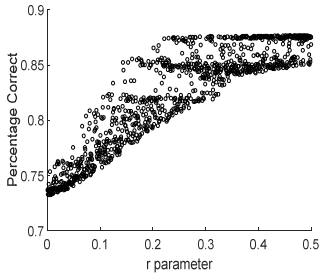
Metric	Jul 2007 # 1 at time 10:27	Jul 2007 #2 at time 21:53
CSI A/(A+B+C)		
F4 / F2 (A-B)/(A+B+C)		
PC (A+D)/(A+B+C+D)		

Figure 5 Skill Scores using a range of performance metrics. Top: CSI, Middle: F4 or F2 Score, Bottom: Percentage Correct

These are just a sample of the results available and show the results from analysis of 2 SAR data from the flood of July 2007, but they are fairly representative of all the results for the full range of SAR-derived flood maps which were analysed. The F2 (aka F4) and CSI plots in particular gave a more sensitive response through the changing parameter value '*r*'. This was no doubt due to the 'white space' of no-water cells being absent from the skill score equation. Considering all SAR results together, the CSI scores were observed to provide a more responsive and consistent result with changing parameter value than the other metrics which were assessed.

Concerning the comment regarding CSI skill scoring usually providing a better result for fuller flood extents due to the dominance of 'water' pixels, as pointed out by Stephens *et al.* (2014). This point is well raised but the authors would respond that the CSI skill score is secondary to the shape of the CSI peak itself in this particular case. This identifiability methodology looks at how sensitive the model is for different parameters around this CSI peak and gives little significance to the CSI scores themselves.

To illustrate also the greater sensitivity/information for '*r*' seen in the images when using CSI, the IC scores using PC and F4/F2 as the central metrics are shown in the table below, against the original scores obtained using CSI:

Table 2 Information Content (IC) for parameter *r*. Top row: single SAR images from July 2007 flood event, and bottom row: the same 2 data, grouped into 'flood event'.

IC for	CSI	F2/F4	PC
Single SAR data (left: SAR1 at time 10:27, right: SAR2 at time 21:53)	0.165 / 0.188	0.066 / 0.102	0.102 / 0.101
July 2007 'flood event'	0.37	0.079	0.105

Abstract

Single satellite Synthetic Aperture Radar (SAR) data are now regularly used to estimate hydraulic model parameters such as channel roughness, depth and water slope. However despite channel geometry being critical to the application of hydraulic models and poorly known *a priori*, it is not frequently the object of calibration. This paper presents a unique method to calibrate simultaneously the bankfull channel depth and channel roughness parameters within a 2D LISFLOOD-FP hydraulic model using an archive of moderate (150m) resolution ENVISAT satellite SAR-derived flood extent maps and a binary performance measure for a 30x50km domain covering the confluence of the rivers Severn and Avon in the UK. The unknown channel parameters are located by a novel technique utilising the Information Content and DYNIA identifiability (Wagener *et al.* 2003) of single and combinations of SAR flood extent maps to find the optimum satellite images for model calibration. Highest Information Content is found in those SAR flood maps acquired near to the peak of the flood hydrograph, and improves when more images are combined. We found model sensitivity to variation in channel depth is greater than for channel roughness and a successful calibration for depth could only be obtained when channel roughness values were confined to a plausible range. The calibrated reach-average channel depth was within 0.9m (16% error) of the equivalent value determined from river cross section survey data, demonstrating that a series of moderate resolution SAR data can be used to successfully calibrate the depth parameters of a 2D hydraulic model.

Introduction

Flooding of over one third of the world's land area affected more than 2 billion people - 38% of the world's population – between 1985 and 2003 (Dilley *et al.*, 2005). Climate change forecasts also indicate that in the future there may be an increase in the frequency and pattern of flooding (European Environment Agency, 2012, European Commission, 2014, IPPC, 2014). One response to this global hazard has been an increasing demand for better flood forecasts (Schumann *et al.*, 2009a). Flood inundation models have an important role in flood forecasting and there has been scientific interest in combining direct observations of flooding from remote sources with these inundation models to improve predictions because of the persistent decline in the number of operational gauging stations (Biancamaria *et al.*, 2010), and the reality that many river basins are inaccessible for ground measurement. Synthetic Aperture Radar (SAR) satellites have particular importance in this respect as they can discriminate between land and smooth open water surfaces over large scales. These microwave (radar) frequency satellites are capable of all-weather day/night observations and this makes them a particularly attractive option for observing floods. Currently active SAR satellites include RADARSAT-2, ALSOS-2/PALSAR-2, TerraSAR-X, TanDEM-X, Sentinel 1 and the COSMO SkyMed constellation. Historic data are also available from SAR satellites now out of operation such as ENVISAT, ERS1 and 2 and RADARSAT-1.

By processing SAR data it is possible to produce binary maps of flood extent that can then be used, either on their own, or intersected with a Digital Elevation Model (DEM) to produce shoreline water levels, for model calibration and validation. Integration of SAR data with models is an established technique for reducing uncertainty in model predictions as it updates/calibrates the model states/parameters with observed data (e.g. Andreadis *et al.*, 2007, Biancamaria *et al.*, 2011b, Domeneghetti *et al.*, 2014, Giustarini *et al.*, 2011, Garcia-Pintado *et al.*, 2013 and 2015, Hostache *et*

al., 2009, Matgen *et al.,* 2010, Mason *et al.*, 2009 and 2012, Tarpanelli *et al.,* 2013, Yan *et al.* 2015), with the aim of improving flood forecasts. Naturally, calibration of these hydraulic models is essential for accurate results, and calibration studies to date have largely focussed on roughness. Aronica *et al.* (2002), Tarpanelli *et al.,* 2013 , Hall *et al.,* 2005 and Di Baldassarre *et al.* (2009a, 2010 & 2011) have used flood extent maps to successfully find best fit roughness parameter values. Mason *et al.* (2003) point to roughness being a dominant factor for shallow reaches in particular and Di Baldassarre *et al.* (2009b) found that the optimal roughness parameters depend on the timing of the SAR image and the magnitude of the flood event. Given this prior research, historic observations of flooding should have a particular role in model calibration and sensitivity testing.

The provision of good bathymetric data is also critical to the application of hydraulic models (Trigg *et al.*, 2009, Legleiter *et al.* 2009~~).~~, Yan *et al.*, 2015). Yet generally there are few ways to obtain bathymetry information for hydraulic models where no ground data measurements exist. River depth may be estimated (e.g. Durand *et al.,* 2010 employed an algorithm based on the Manning equation or Moramarco *et al.* 2013 who created an entropy depth distribution using surface flow velocity data) or measured with optical satellites using reflectance as Legleiter *et al.* showed (though the method is best suited to clear and shallow streams). Hostache *et al.*, (2015) also proposed a drifting GPS buoy to assimilate water elevation and slope data into a hydraulic model to define riverbed bathymetry, but overall passive and remote mechanisms are scarce. Spatially distributed river depths are rarely available and there is a strong argument that where channel geometry is *a priori* unknown it should also be estimated through calibration.

It has commonly been thought that channel geometry and roughness traded off against each other (e.g. as in the well-known Manning equation) and therefore that they could not be uniquely identified at the same time. However, Garcia-Pintado *et al.* (2015) estimated channel friction and spatially-variable channel bathymetry together using water levels derived from a sequence of real SAR overpasses (3m resolution data from the COSMO-SkyMed constellation of satellites) and the Ensemble Transform Kalman Filter. Durand *et al.* (2008) demonstrated that estimates of depth and water (i.e. friction) slope could be derived simultaneously from synthetic observations of water surface elevation integrated with a hydraulic model, though this research related more specifically to depth of flow, rather than depth of channel. Yoon *et al.* (2012) were also able to derive bed elevations from similar synthetic data. Mersel *et al.* (2013) progressed this further by proposing a slope-break method to locate optimal locations to measure flow depth, through low to high flows over time, using synthetic data. Durand *et al.,* Yoon *et al.* and Mersel *et al.* used synthetic altimetry data which was created within the context of the upcoming Surface Water & Ocean Topography (SWOT) mission that will be able to resolve rivers over 100m wide only.

Research to date has therefore demonstrated the feasibility of calibrating hydraulic model parameters governing channel depth and channel roughness simultaneously. This has been achieved using the higher spectrum resolution (up to 50m resolution) SAR images of flood extent. But because pixel size is inversely proportional to orbit revisit time, high resolution data are available only infrequently. There is thus some benefit to also exploring the use of existing moderate (50m to 300m) resolution SAR data (such as the archive of 150m resolution ENVISAT Wide Swath Mode) to understand more about how channel depth and friction can be identified concurrently using coarser resolution SARs, and whether a single SAR flood map is sufficient to achieve this or a sequence of flood maps are more beneficial.

Therefore the objective of this paper is to draw on this prior research for simultaneous channel roughness and depth calibration and extend it to determine whether medium resolution SAR data can be used to concurrently estimate channel friction and geometry parameters in a hydraulic model. If it can be used; to determine if a single SAR derived flood map is sufficient to do this, or if a sequence of flood maps is more useful. For this the identifiability technique presented by Wagener et al. (2003), namely Dynamic Identifiability Analysis (DYNIA) is utilised. A secondary aim of this paper is therefore to test the utility of the DYNIA identifiability technique in this specific context to find the SAR images with high parameter information and locate the likely optimum parameter values. This methodology particularly uses flood extent with an accuracy-scoring method that disregards the correct detection of 'no water' pixels.

In section 1 we describe the methodology with information on the hydraulic model, the data needed to run it and the methods used to select the range of model parameters. There is also an introduction to the procedure used to process the satellite data and create flood extent maps. Section 2 describes the study area and data used, whilst Section 3 presents and discusses the results (including whether SAR observations at particular times during a flood or particular combinations of images are more successful). Conclusions are presented in Section 4.

1 Method

1.1 Hydraulic model

We use the LISFLOOD-FP hydraulic model with the Sub-Grid formulation of Neal *et al.* (2012) to simulate flood flows. LISFLOOD-FP (Bates and De Roo, 2000) is a 2D hydraulic model for subcritical flow that solves the local inertial form of the shallow water equations using a finite difference method on a staggered grid. As input the model requires ground elevation data describing the floodplain topography, channel bathymetry information (river width, depth and shape), boundary condition data consisting of discharge time series at all inflow points to the domain, water surface elevation time series at all outflow points and friction parameters which typically distinguish different values for the channel and floodplain. Of these data floodplain topography information is readily available from airborne and satellite Digital Elevation Models, boundary condition data can be taken from ground gauges, hydrologic models or statistical distributions, and friction parameters are typically estimated from lookup tables or calibrated. Channel bathymetry can be taken from ground surveyed cross sections, however for much of the planet no such measurements exist and are impossible to obtain remotely. In this situation channel bathymetry is *a priori* unknown and it is therefore sensible to also treat it as a parameter that must be calibrated along with the friction.

In order to describe bathymetry as a calibrated variable in this experiment, river channel depth was parameterised as a linear scaling of reach-average width. In general, this linear approach will not be appropriate over an entire river network where the reach-averaged width to depth relationship would be expected to change with bankfull discharge. However, the width of the river chosen as a test case for this paper is constant along the simulated reach, while we assume the depth of tributaries has an insignificant impact on the flooding on the main stem. In effect the optimisation problem therefore simplifies to estimating reach-averaged bankfull depth and Manning's 'n_c' for a channel of reach-average width. In width-varying river systems a dual parameterisation approach for depth and width

could be adopted but would substantially complicate the parameter estimation problem. The floodplain Manning's roughness coefficient was assumed constant in these experiments as previous tests have shown that the model was less sensitive to floodplain friction than channel friction .

We used Latin Hypercube Sampling (LHS) to take 1000 samples of the two uncertain LISFLOOD-FP parameters '*r*' and channel Manning's roughness 'n_c'. LHS is a useful sampling scheme for multiple variables as the method can sample parameter values within a prior distribution in more than one dimension (Huntington, 1998). We used LHS here as it is an efficient scheme that statistically represents the parameter space without repetitions (Beven, 2009~~).~~ and Pianosi *et al.* 2016).

1.2 SAR image processing algorithm

Because SAR satellites are capable of all-weather day and night observations and can distinguish the differences between land and open water signal returns they are particularly useful for observations of flooding. To derive flood extent maps from the SAR images, we adopted the method proposed by Matgen *et al.* (2011) and developed by Giustarini *et al.* (2013) and Chini *et al.* (under review). This method has three steps as illustrated in Figure 1 below. Firstly the probability density function (pdf) of the open water backscatter values in the SAR data is estimated. This requires identification of the bimodal aspect to a histogram of backscatter values so that 'open water' values can be recognized from other backscatter values. A theoretical pdf of water backscatter is then fitted to this histogram using nonlinear regression techniques. The backscatter threshold value (Th_{seeds}) where this pdf starts to diverge from the histogram is identified. Then isolating those pixels with backscatter values lower than this threshold produces a preliminary flood map (region growing seeds). The second step is to apply a region growing approach to grow the flooded areas within the preliminary flood map until a tolerance threshold level is reached ($Th_{tolerance}$). For the SAR image this step refines the extent of pixels with an open water value.

In the last step a reference image is used to remove pixels from the flood map that do not change between the flood and non-flood images (Hostache *et al.,* 2012) – i.e. pixels which have 'water surface like' radar responses and could be either bodies of permanent water or smooth surfaces such as car parks or flat roofs. This third step creates the final binary map of flood extent. Errors inherent in the SAR processing are, for simplicity, not considered in this paper.

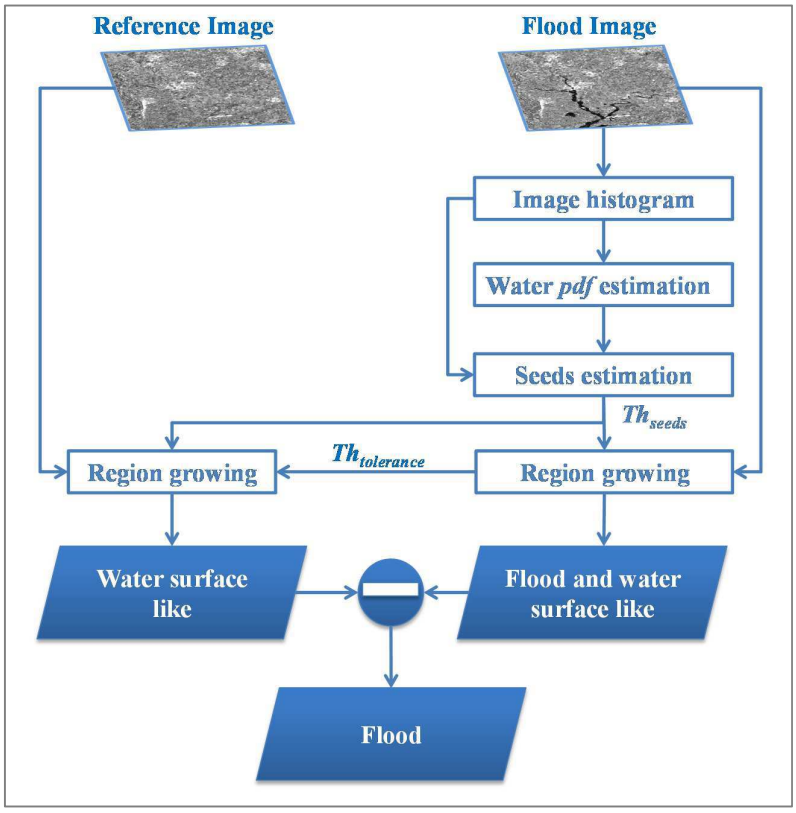


Figure 1 - General scheme of the three processing steps of the flood detection algorithm.

1.3 Performance measures

We compare these SAR derived flood maps against the simulated flood maps generated from LISFLOOD-FP output at the equivalent time step by using a contingency matrix shown in Table 1. Flood maps are compared pixel to pixel to determine if there is agreement or disagreement between the two paired maps on whether there is surface water present or not.

Table 1 - Contingency table (after Stephens et al, 2014 and Mason, 2003).

		Modelled	
		Water	**No Water**
Observed	**Water**	A) Correct Water (Hits)	B) Under-prediction (Misses)
	No Water	C) Over-prediction (False Alarms)	D) Correct No Water (Correct Rejections)

From this a binary pattern performance measure is used to give a deterministic indication of how well each LISFLOOD-FP simulated flood map has represented the observed data (Mason, 2003 and Stephens *et al.*, 2014). We chose to use the Critical Success Index (CSI, equation 1 below) as this measure does not consider 'correct rejections' (*D* in Table 1) in the calculation (Bates and De Roo, 2000, Horritt *et al.,* 2001a, Aronica *et al.,* 2002) and it weights over- and under-prediction equally (*C*

and *B* respectively). CSI scales between 1 (indicating perfect skill in the model) and 0 (indicating no skill in the model).

$$CSI = \frac{A}{A+B+C} \quad (1)$$

If 'correct rejections' were included by the use of a different performance measure the result would be overly optimistic scores, given the large areas of 'no water' normally observed in a SAR image. All LISFLOOD-FP simulated flood maps would seem to perform exceptionally well with little to help differentiate between each simulation.

Before comparing SAR and LISFLOOD-FP model results an independent remote dataset is used to illustrate the impact of observation errors and gaps inherent in the SAR data from processing. This validation step makes use of a very high-resolution (0.2m) aerial photograph taken by the EA on 24 July 2007 from an aircraft passing over at 11:30 GMT (details within Giustarini *et al.,* 2013). A flood map shapefile was created from this imagery by manual definition of the flood boundary. This was then converted and upscaled to a raster with the same spatial resolution (75m) of the LISFLOOD-FP model results. Both the ENVISAT data and the LISFLOOD-FP results (the highest scoring models) are compared with this aerial data. A figure showing these flood extents and the CSI results from this comparison are given in section 3.1 below.

1.4 Parameter identifiability

To determine most likely values for '*r*' and 'n_c' we follow the technique of Wagener *et al.* (2003) in applying a dynamic identifiability analysis (DYNIA) to the ensemble of CSI score results. Since the original DYNIA method was applied to continuous data and not discrete observations some changes are needed which are described at the end of this section.

The first stage in the DYNIA method is to rescale the 'objective function' (i.e. CSI scores) so that they add up to one, which is done by dividing each model result by the sum of all scores. Next, computing the cumulative distribution of the rescaled objective function transforms the objective function into a support measure which sums to unity - the 'cumulative support' – so that each support measure may be comparable. To obtain the information content (IC) a confidence limit is applied to the rescaled objective functions to exclude outliers. The width of the confidence limit depends on how the best performing parameters are spread within the parameter space: a wide confidence limit suggests that the parameters are distributed within the parameter space evenly and IC is low, whereas a narrow confidence limit suggests that the best performing parameters are located within a smaller range and IC is higher. To normalise results for this data a transformation measure was used (1 minus the width of the confidence limits over the parameter range, normalised to run from zero to one): so a value close to 1 is equivalent to a high IC. The IC can have any value between 0 (no information in that observation for parameter identification purposes) and 1 (observation is most informative for the parameter). The IC results are shown in section 3.2 below.

The second stage in DYNIA is to find the identifiability by locating where in the parameter-time space most parameter information can be found. This is achieved by examining a plot of 'cumulative support' against a parameter value. Any deviation from a straight line gradient of this cumulative support indicates whether the parameter is conditioned by the objective function or not. The stronger the

deviation, the stronger is the conditioning/identifiability of the parameter variable. This is done using the marginal parameter distributions – interactions are therefore only implicitly accounted for. The final stage is to organise the data into bins and calculate the gradient of the cumulative support between them. The results from this examination are shown in section 3.3 below. These results are represented using plots of the gradient of the cumulative support value versus the parameter of interest to indicate the strength of the identifiability in each case. The IC and identifiability for all single SAR acquisitions are shown along with particular SAR combinations/groupings: by flood event and by position in the flood hydrograph as detailed in section 2.2 and Table 3. ~~The identifiability plots have been converted to cumulative distribution function (cdf) plots for easier cross- comparisons.~~

The original method proposed by Wagener *et al.* (2003) recommends a pre-selection of models before stage 1 by using only the top 10% performing models. We deviate from this original method by using the complete sample of 1000 sets of CSI scores since we found this gave a clearer overview picture of identifiability with our data.

The objective of this paper is to determine if a grouping of SAR data provides more information than single data. Here the method of obtaining the CSI 'group' score is also a small departure from the original DYNIA method. These group scores are determined by multiplying each single model/SAR flood map CSI result with the CSI score of the next SAR flood map until all members of the particular group have been added. The unique combinations which comprise these groups are described in Table 3 below. This combining of CSI scores is done for results from each of the 1000 models/parameter scenarios. The next step is the same as for single CSI scores as described above – i.e. to rescale the objective function and compute the cumulative support. So although multiplying CSI values will reduce the grouped score, it has no bearing as it is the changes to the gradient of the cumulative support value that indicates parameter identifiability, not the CSI scores themselves. The group IC and identifiability results shown in sections 3.2 and 3.3 result from SAR data that was grouped by this multiplication of CSI scores.

2 Study area and data used

The area around Tewkesbury (UK), located at the confluence of the Rivers Severn and Avon is our test location. Figure 2 illustrates the 30.5 km by 52.4 km model domain, showing the two main rivers and their tributaries.

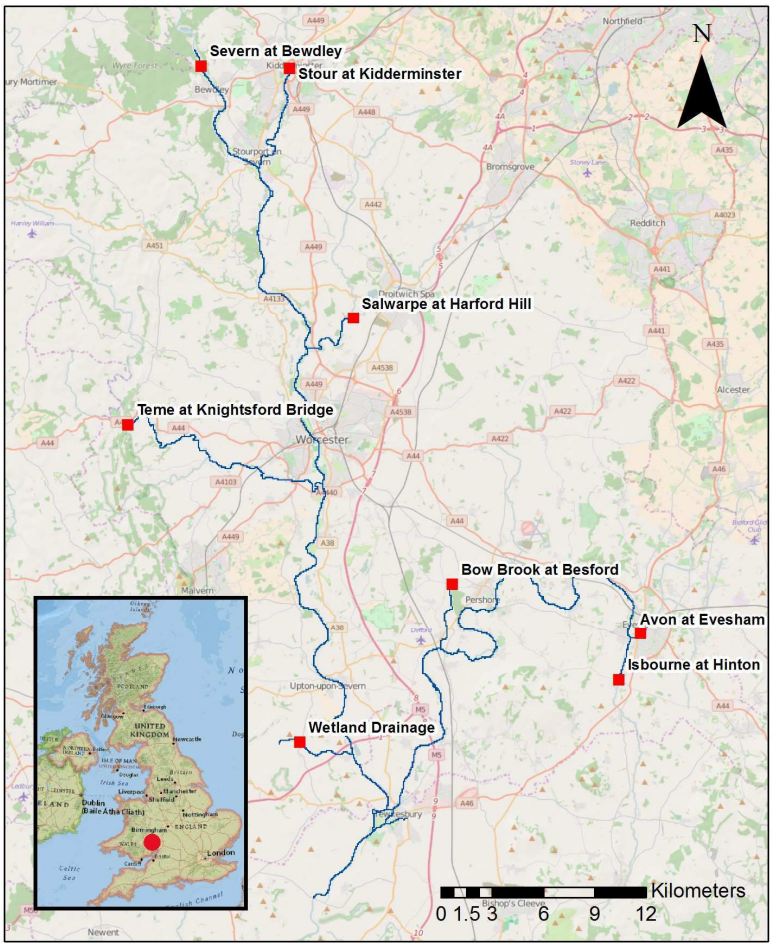


Figure 2 - Extent of the River Severn model.

2.1 River Severn model set up

Two separate LISFLOOD-FP models were created to test the methodology. Both models are at 75m spatial resolution and use the same background DEM. Additionally, both models use the same gauged inflows and have a rectangular shaped channel. At the lower end of the model a 'free' downstream boundary condition was applied with a fixed energy slope of 0.00007, based on the average valley slope.

The differences between the two separate models are in how bankfull channel depth and Manning's channel roughness values are obtained. First, an 'observed' model was created using surveyed cross sections of the main rivers to determine channel width and depth with a fixed Manning's channel roughness parameter of 0.038 (a value representing a main channel which is clear with some winding

and presence of stones/vegetation - from Ven Te Chow, 1959). The cross section survey data were provided by the Environment Agency of England and Wales (EA). Second, a 'test' model was created in which the depth parameter '*r*' and Manning's channel roughness parameter 'n_c' are determined using the DYNIA identifiability analysis as described in the previous section. The depth parameter *'r'* was sampled between 0.0 and 0.5 so that the modelled river depth would never exceed half of the river width. This is a reasonable assumption for this site where the Severn is on average around 75m wide (estimated from LiDAR data) with surveyed bankfull depth varying between 6m and 11m. The range of Manning channel roughness values for the sampling was set between 0.015 and 0.100 (Ven Te Chow, 1959). A low 'n_c' of 0.015 would represent a channel which is clear and straight whereas and a high 'n_c' value of 0.100 would represent a channel with very thick vegetation/submerged branches present. This range widely encompasses recommended roughness values for the rivers present within the study domain.

For both the test and observed models the Manning's floodplain roughness value was set at a standard 0.06 for the entire domain. This is a reasonable average for the floodplain which is mainly crop and grassland (0.03-0.04) but with presence of some trees (0.12) and brush (0.07). The Manning's values for the floodplain and the river channel ('n_c') are assumed to be spatially and also temporally invariant. The floodplain topography was taken from a 2m resolution LiDAR based Digital Surface Model (DSM) with vertical RMSE of 0.10m taken on 9 December 2005 by EA. The EA treated the DSM to remove structures and vegetation and we then spatially averaged this Digital Terrain Model (DTM) to 75m resolution as this is an appropriate compromise between model fidelity and computational cost for rural river reaches (Horritt and Bates, 2001b). The 75m DTM was further processed to reinsert the maximum height of the flood embankments along the reach in order to preserve normal flood behaviour along the river banks. No bridges or weirs are included in the model. Neal *et al.*, 2011 and Garcia-Pintado *et al.*, 2013 provide additional details of the model set up for the River Severn around Tewkesbury.

Observed flows obtained from the EA were used as inflow to both models. Forcing flows come principally from the gauging station on the River Severn at Bewdley but with additional inputs from three tributaries of the River Severn: River Stour (at Kidderminster), River Salwarpe (at Harford Hill near Droitwich Spa) and River Teme (at Knightsford Bridge near Knightwick). For the River Avon flows from the Evesham gauging station were used, with two additional flow contributions from the Avon tributaries Bow Brook (at Besford) and the River Isbourne (at Hinton). A smaller input from a wetland area west of Tewkesbury was also included, with flows scaled by area from the Salwarpe gauged flows.

The River Severn floods events of March 2007 (simulation period: 19 February 2007 - 29 April 2007), July 2007 (simulation period: 5 June 2007 - 12 August 2007), January 2008 (simulation period: 26 November 2007 - 25 February 2008) and January 2010 (simulation period: 4 January 2010 – 18 February 2010) were modelled. The dates were chosen so the model would start at least 10 days before the start of the flood and end after flows had returned to within bank.

2.2 SAR observations of the River Severn

Historic ENVISAT Wide Swath Mode ('WSM', 150m resolution) data are available from the European Space Agency's ENVISAT catalogue. These were resized to 75m resolution data. Previous research at this site has largely focused on the July 2007 flood event observations (Mason *et al.,* 2012 and 2014,

Durand *et al.,* 2014, Garcia-Pintado *et al.,* 2013, Schumann *et al.,* 2011). The present work makes use of other historic flood observations in this area – namely the floods of March 2007, January 2008 and January 2010. Details of the satellite acquisition times are shown in Table 2, along with hydrologic information on the flood taken from the gauging station at Saxons Lode in the middle of the model domain. Time to peak describes the number of hours between the start of the event and the peak of the flood. Flooding from sequential events or with high contributions from other sources such as groundwater will therefore have a greater time to peak.

Table 2 - The ESA sourced ENVISAT ASAR WSM acquisitions used with equivalent flow and return period data for Rivers Avon and Severn: gauged data was obtained from the EA.

SAR ID	Date	Time	Time to flood peak (approx., hrs.)	Gauged flow (m3/s)	Event return period (approx.)	Gauged flow (m3/s)	Event return period (approx.)
			At Saxons Lode (Severn)			At Evesham (Avon)	
1 (March 2007, 1)	05/03/2007	10:27	268	388	<5	188	<3
2 (March 2007, 2)	05/03/2007	21:53	268	405		87	
3 (March 2007, 3)	08/03/2007	10:34	268	419		55	
4 (March 2007, 4)	08/03/2007	21:58	268	400		45	
5 (July 2007, 1)	23/07/2007	10:27	132	532	30-40	196	110-150
6 (July 2007, 2)	23/07/2007	21:53	132	512		167	
7 (January 2008, 1)	17/01/2008	21:55	228	432	<5	64	<3
8 (January 2008, 2)	24/01/2008	10:12	228	440		28	
9 (January 2008, 3)	24/01/2008	21:38	228	433		26	
10 (January 2010, 1)	18/01/2010	10:30	73	407	<3	107	2
11 (January 2010, 2)	18/01/2010	21:53	73	403		37	

We separated these 11 SAR observations into different categories by particular flood event (section 3.3.2) or where the acquisition occurs on the flood hydrograph (section 3.3.3). Table 3 shows how this segmentation of the 11 acquisitions into categories was devised.

Table 3 - Description of SAR groupings.

Description	SAR ID										
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
By flood 'event'	March 2007				July 2007		January 2008			January 2010	
By point in hydrograph [r =rising limb, p = peak, f = falling limb]	r	r	f	f	f	f	p	f	f	p	p

3 Results and discussion

3.1 CSI scores

In this paper we compare the results of hydraulic model-generated flood maps with the SAR observations of flood extent in order to determine if the satellite data has information in terms of calibrating the model. However with inherent errors in the SAR data from processing it is worthwhile first to compare the SAR data with those from other available remote data to illustrate the impact of observation errors. For validation, the CSI score is calculated between the ENVISAT data and an aerial photograph of the River Severn taken on 24th July 2007.

Figure 3 illustrates the derived flood extent from this aerial data (far left) with the ENVISAT WSM SAR derived flood map (centre left) from the previous day. Highest scoring LISFLOOD-FP simulation flood maps from the 'observed' model (centre right) and 'test' model (far right) at the same time step as the ENVISAT data are included for comparison. The CSI results from this SAR-aerial and SAR-LISFLOOD-FP model comparison are shown in Table 4.

Table 4 - CSI scores, for July 2007 flood extent maps. Comparing results obtained using ENVISAT WSM SAR and aerial derived flood extents, with hydraulic model generated flood extent.

Flood map:	Aerial photograph derived	ENVISAT derived
Aerial photograph derived	-	0.47
'Observed' Model (not calibrated)	0.74	0.43
'Test' Model	0.75	0.46

It is clear that the observed and test LISFLOOD-FP models produce lower CSI scores with the SAR data than with the aerial data. This is to be expected and other studies, which have used higher resolution SAR imagery for validation (e.g. Bates *et al.* 2006, Di Baldassarre *et al*. 2009a and 2010), have observed the same result. The aerial photograph-derived flood map was delineated manually and therefore has improved representation of flooding because there are no detection gaps in the flood extent, whereas SAR-derived flood extents rely on the correct detection of areas of water using a procedure which is vulnerable to issues of detection and processing. So while we may conclude that aerial imagery has the best level of detail in flood extent available here, this data can also be limited by observation extent and processing (i.e. manual delineation of the flood edge is still interpretive) and as a resource is not as frequently available as SAR data for observing flood events. It is also worth pointing out that for the ENVISAT SAR data, describing flood extent using the semi-automated algorithm can be a faster solution than manually delineating flood extent from new photographs.

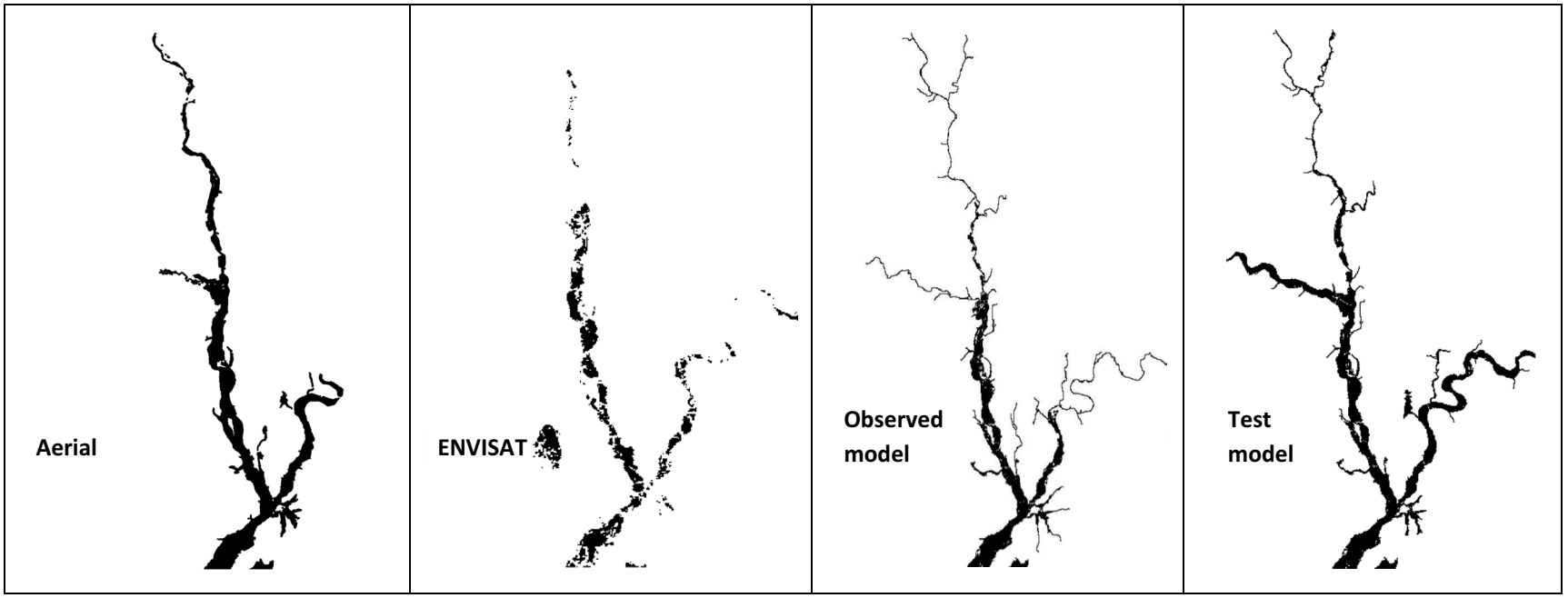


Figure 3 - The July 2007 flood extents as observed by aerial photography (on 24rd July 2007 at 11:30 , left) and ENVISAT ASAR instruments in WSM (on 23rd at 10:27, centre left). The same flood event simulated in LISFLOOD-FP with surveyed cross sections (centre right, with Manning's channel roughness fixed at 0.038) and the test model with optimally calibrated parameters (right).

The scores and flood extent for the observed model are not better than the test model results as might be expected. This may be explained by the fact that while the bathymetry of the observed model does come from survey data, the (domain-average) channel roughness value is not calibrated in either model. While the test model had 1000 parameter-varying depth and roughness values, the observed model had a best estimate of domain-average channel roughness parameter (of 0.038). While appropriate for the main rivers, it is evident that the channel roughness value is not suitable for the narrower tributaries.

Of interest also, when the aerial data is compared with the ENVISAT WSM SAR derived flood maps (row 1, last column), CSI scores are similar to those obtained from the best hydraulic model results. This indicates that the hydraulic models are representing the observed flood extent for this flood accurately, within the limits of the available data. While sections of the flood are missing in the SAR data (for example upper River Avon and Severn) bias can be introduced. Ideally these non-informative areas of the SAR data would be masked out to limit the impact, but with series of data each differently capturing a flood event this requires a more comprehensive analysis than available here. It is currently is an active area of research; for example Giustarini *et al.* (submitted) propose flood probability maps from sequences of SAR data. These maps could be used to mask out 'low probability of flooding' areas. Also Schlaffer *et al.*, 2015 makes use of harmonic analysis to refine flood extent mapping – a mask could be created to obscure pixels with low signal to noise ratios.

As explained in section 1 the first step in the methodology is to examine the accuracy of the test model with changing parameter value, using CSI. The ENVISAT WSM SAR and LISFLOOD-FP CSI results were plotted against the *'r'* and 'n_c' parameter variables and are presented in Figure 4. This figure includes only two plots: one for an ENVISAT WSM acquisition taken on 23rd July 2007 (10:27am) and one taken on 24th January 2008 (10:12am), but these CSI results represent typical results for the entire SAR data available.

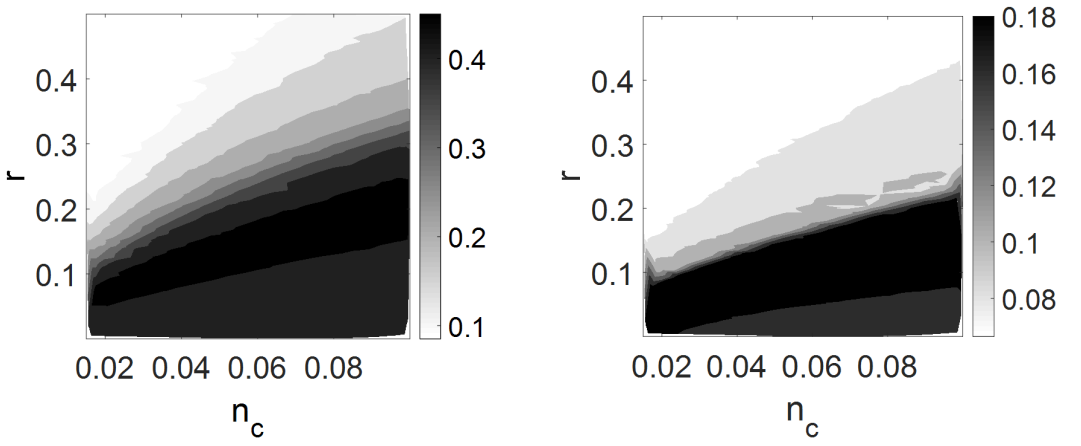


Figure 4 - Single SAR acquisitions are compared with LISFLOOD-FP modelled flood maps. Left: results from the SAR acquisition on 23rd July 2007 at 10:27, right: result from the SAR acquisition 24th January 2008 at 10:12.

The black areas in Figure 4 show that a number of *'r'* and *'n_c'* parameter combinations/models are able to produce a good result (i.e. equifinality as described by Beven, 2009). The optimal *'r'* parameter range varies slightly depending on the image considered. Here test models with the best reproduction of the SAR flood map have '*r*' parameters between approximately 0.10 and 0.30 (July 2007) and

between 0.07 and 0.25 (January 2008). Generally, the best reproduction of the SAR flood maps is obtained with models that have an '*r*' value in the smaller parameter range which translates to a wide and shallow river channel.

Figure 4 also illustrates the co-variance and a linear dependency between the two parameters. This was observed in all the SAR data. Although the choice of parameter range emphasizes it, there is a slightly greater skill score sensitivity to changes in *'r'* than for *'n_c'*. This is to be expected since changes in channel depth would have an immediate and local impact on flood level and flood extent. It is logical therefore to see changes in '*r*' producing a marked change in flood extent. Channel roughness changes by contrast have an impact more on flow velocities, consequently impacting on the timing of flood wave propagation through the channel (as discussed in Neal *et al.,* 2015). This would have a more spatially diffuse impact on flood extent that is barely perceptible here.

Previous SAR based assimilation studies (Hostache *et al.,* 2009, Mason *et al.* 2009, Di Baldassarre *et al.* 2009a) show that with a known and fixed channel bathymetry there is sufficient sensitivity in the roughness parameter to enable calibration. The above findings indicate that the sensitivity of 'n_c' is less obvious when '*r*' is also unknown. There are previous studies also where, as here, channel friction appears less sensitive when other parameters are simultaneously calibrated. Roux *et al*. (2008) for example found sensitivity in hydraulic model response to channel roughness to be weaker than sensitivity to geometry parameters and boundary conditions within a Generalised Sensitivity Analysis framework. Additionally Garcia-Pintado *et al.* (2015) found that sensitivity to bathymetry parameters dominated when using the Ensemble Transform Kalman Filter to simultaneously estimate bathymetry and channel friction. The sensitivity in channel friction may therefore be not as obvious when other parameters are simultaneously calibrated because the model is no longer compensating for previously unrepresented uncertainties. It could be suggested that channel friction is reverting to its true sensitivity and so when channel friction is combined with more dominant parameters such as channel bathymetry it is rendered less useful for model calibration.

Consequently an important result of this paper is that - in this particular experimental set up with channel roughness parameter *'n_c'* examined simultaneously with the channel depth parameter '*r*' for the available ENVISAT SAR data - 'n_c' has a much reduced sensitivity compared with the '*r*' depth parameter response. It is observed that 'n_c' will yield optimal results for as long as '*r*' is also unknown. This lack of sensitivity of channel roughness in this and all subsequent results meant that 'n_c' could not be identified with any real confidence with this methodology (while '*r*' is also unknown). So while 'n_c' analysis was carried out, from this point onwards only those results from the more identifiable '*r*' parameter~~.~~ are shown. 'n_c' results are now omitted (but can be provided upon request if of interest).

3.2 Information content (IC)

Table 5 presents IC results for depth parameter '*r*'. For single SAR observations (left column) there is clearly greater information content in the July 2007 flood event images. The inundation during this higher magnitude event extended well into the floodplain and the flood detection algorithm was able to detect a large number of flooded cells. The lower IC scores for the March 2007, January 2008 and January 2010 events show that these observations contain less information to help estimate parameter '*r*'.

Table 5 - Information content for '*r*' from SAR observations and groups of SAR observation with a 90% confidence limit applied.

Sequence	Information Content	Sequence	Information Content
1 - Mar07_1	0.10	**Rising limb**	0.13
2 - Mar07_2	0.11	**Peak of hydrograph**	0.23
3 - Mar07_3	0.11	**Falling limb**	0.64
4 - Mar07_4	0.11	**March 07 event**	0.50
5 - Jul07_1	0.16	**July 07 event**	0.37
6 - Jul07_2	0.19	**January 08 event**	0.25
7 - Jan08_1	0.10	**January 10 event**	0.14
8 - Jan08_2	0.11	**All SAR [1-11]**	0.68
9 - Jan08_3	0.11		
10 - Jan10_1	0.10		
11 - Jan10_2	0.10		

Grouping SAR data boosts the IC scores considerably as can be seen in the right hand side columns of Table 5. Group IC scores are estimated after the SAR data have been grouped together and CSI scores combined as described in section 1.4. Different SAR groupings were tested as illustrated in Table 3 including combinations according to flood event, position on the hydrograph as well as 'all SAR' data.

For IC the July 2007 flood now no longer outperforms the rest and instead combinations of images, like the March 2007 flood event, have greater information on '*r*'. The March 2007 flood combination combines observations either side of the hydrograph peak and the January 2008 flood combination observes flooding 'at peak' and soon after in the falling limb. By contrast the reduced-scoring January 2010 and July 2007 combinations acquired images at a single stage in the hydrograph only. We might conclude that the detection quality of the SAR flood maps and timing of acquisition must influence the final IC score and this is supported also by the observation that the early 'falling limb' grouping has one of the largest IC scores here.

Nevertheless, the number of SAR flood maps combined appears to be important also since the 'all SAR' and early 'falling limb' (just over half of these SAR images, Table 3) groupings emerge as providing the highest IC. The March 2007 flood grouping also contains twice as many members as the July 2007 or January 2010 flood groupings and outperforms both. Clearly, incorporating data from multiple observations improves IC since combining SAR images (and CSI scores) improves the likelihood of extracting information on the unknown parameters. However it is not simply a question of numbers otherwise 'falling limb' (combining 6 SAR flood maps for an IC score of 0.64) would not be approaching the success of 'all SAR' (combining 11 SAR flood maps for an IC score of 0.68). Nor is greater information necessarily revealed by removing poor scorers ('all SAR' IC score reduces from 0.68 to 0.64 when the 4 lowest scoring flood maps are removed from this grouping). Instead the solution may lie in using SAR flood maps around the peak and falling limb of the flood since combining 'falling limb' and 'rising limb' observations together yields an IC score of 0.65 but combining 'falling limb' and 'peak' observations together provides an IC score of 0.67. Further work and data is necessary to draw any firm conclusions for the '*r*' model parameter.

3.3 Identifiability

The identifiability of '*r*' within single images and combinations of images is assessed in this section. This shows where the parameter is most easily identified in the ensemble of model results. A strong identifiability response would be marked out by having a sharper peak in the following plots. The steeper the gradient, the stronger is the identifiability of the parameter. A sharper peak indicates that the best performing parameters are concentrated in a small area of the parameter space. Conversely a wider, shallower peak would indicate lower identifiability and that the best performing models are widely distributed within the parameter range.

From the CSI contour plots as illustrated in Figure 4 we see that the best performing model parameter combinations are distributed fairly evenly within the parameter space so a 90% confidence limit was also applied to the data prior to measuring the gradient of cumulative distribution of rescaled support values and creation of these following plots.

3.3.1 Individual SAR observations

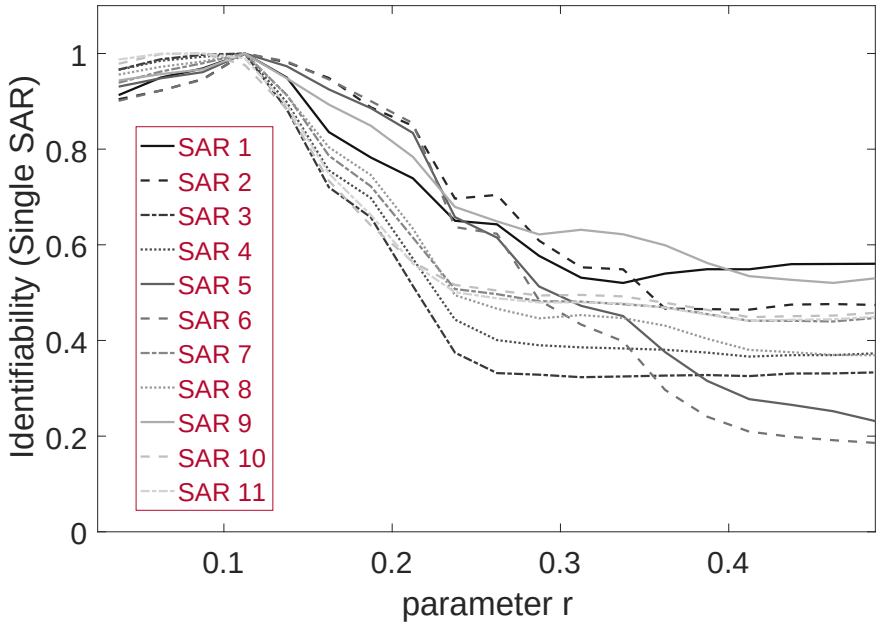


Figure 5 - Identifiability against *'r'* parameter, for each ENVISAT SAR observation in archive.

Figure 5 shows the identifiability plots for all single SAR data, numbered as in Table 2. Because these plots do not generally have a strong peak, identifiability is relatively weak for the individual SAR observations. The strongest response here occurs for '*r*' between 0.05 and 0.15. The peaks are shaped differently for each SAR observation; SAR 4 and SAR 3 both have stronger identifiability (narrower peaks than the rest) whereas SAR 6 and SAR 2 are relatively weak in this ensemble by having wider peaks.

Taken collectively these data provide inconclusive results. This generally weaker identifiability suggests that parameter '*r*' would be difficult to identify within this data individually. The SAR data

were acquired during different flood events (see Table 3) and their peaks occur at different *'r'* parameter values. This variation may be due to differences in the size of flood extent (magnitude of flooding), the processing of the image or simply how the flood developed and that the location of flooded pixels is important.

3.3.2 Flood event

This section illustrates identifiability when data from individual SAR images are combined into 'flood events' as indicated in Table 3. An important characteristic of the 'flood event' identifiability plots is that the SAR acquisitions are taken together in close sequence. Garcia-Pintado *et al.* (2013) found that a tight sequence of images could improve model predictions. Combining observations in this way appears to focus the location of the '*r*' parameter more clearly than is possible using single images.

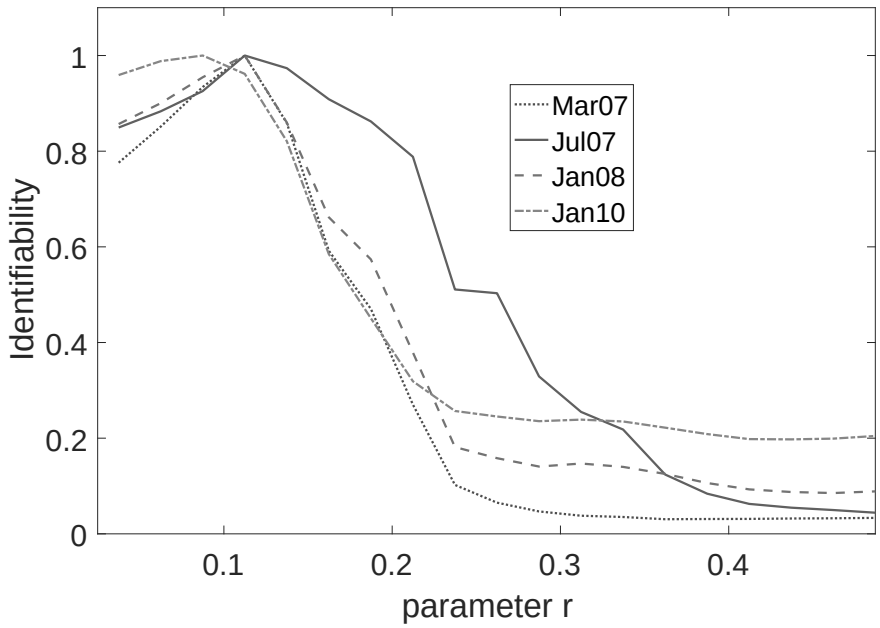


Figure 6 -Identifiability against parameter *'r'*, for flood events.

This plot shows that the March 2007 and January 2008 events produce a stronger identifiability, between '*r*' parameter values 0.07 and 0.15. However, the optimum '*r*' value varies between 0.07 to 0.1 and 0.1 to 0.15 depending on which of these floods is examined. It is entirely reasonable that identifiability of channel depth parameter in the data would vary with flood event as each flood is unique in magnitude and mechanism. Based on Figure 6, the March 2007 and January 2008 SAR images might therefore be best utilised to locate the value of parameter *'r'*. These two events have approximately the same peak discharge flows at Saxons Lode (see Table 2). However, the IC results point towards the March 2007 data combination alone as having more parameter information and the reason for this becomes clear when looking at the individual SAR maps of flood extent. The group of SAR images acquired in March 2007 combine to yield a more complete representation of the flood development than the combination from January 2008. So although Figure 6this identifiability plot shows that both March 2007 and January 2008 flood events would be useful to locate the parameter '*r*', IC shows the information contained in the March 2007 flood maps to be of most value.

3.3.3 Through the flood hydrograph

Figure 7 looks at identifiability at three stages of a flood hydrograph for the '*r*' parameter, namely from observations at the (late) rising limb, the peak and the (early) falling limbs (with reference to the stage hydrograph at Saxon's Lode in the central portion of the model domain). The SAR data used for 'through the hydrograph' groupings is described in Table 3.

Previous studies have found that the scheduling of SAR images is important for calibration of models. Di Baldassarre *et al*. (2009b) found that identification of the optimal model parameters depended on the timing of the SAR image acquisition and the magnitude of the flood event. Garcia-Pintado *et al.*'s (2013) paper established that to improve forecasting of water levels in a model, regular observations during the rising limb and then less frequent observations during the falling limb gave most success. Additionally, Schumann *et al.* (2009b) cautioned that SAR images acquired during the wetting and drying phases of a flood could be showing floodplain connections and dewatering processes unconnected with the hydraulics represented by the model.

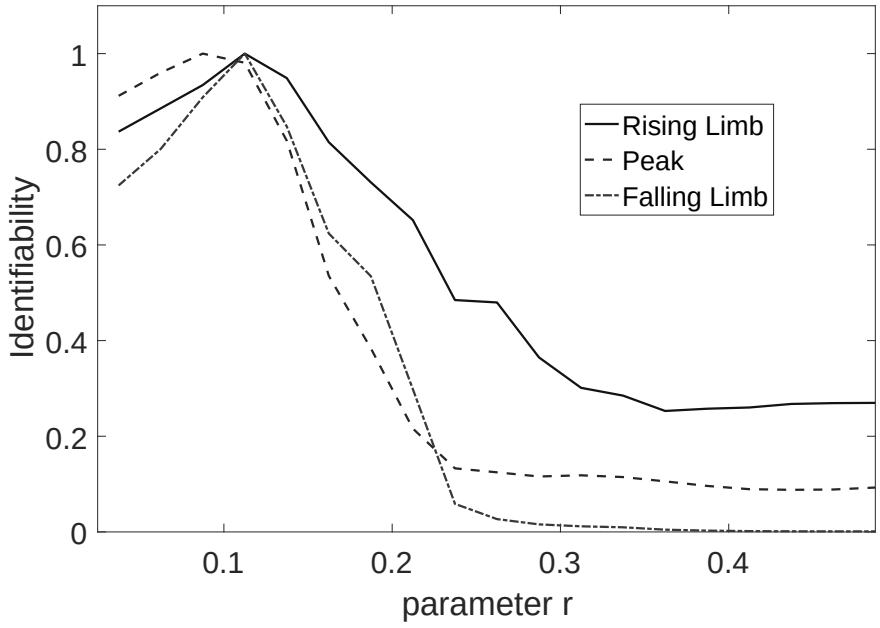


Figure 7 - Identifiability against *'r'* parameter, for different stages in hydrograph.

While here the number of SAR data within each category is limited, Figure 7 shows there is a still difference in identifiability for these separate phases. The strongest '*r*' parameter identifiability occurs for those images taken around the flood peak and falling limb of the hydrograph. These lines have the steepest gradients and narrower peaks. Parameter '*r*' is most identifiable between 0.1 and 0.2 in this data.

The weakest identifiability for the '*r*' parameter occurs for the images taken during the rising limb as evidenced by the wider peak. Yet this result is in contrast to previous studies (e.g. Garcia-Pintado *et al.,* 2013). The reasons for this disagreement with earlier research may simply lie with the way that 'through the hydrograph' images were categorised. The method makes use of only a single

independent gauge (at Saxons Lode) to define the phases and as such it could be an oversimplification of the flood dynamics in a river catchment such as this where the 'rising', 'peak' and 'falling limb' of the flood occur at different times depending on where you measure within the model domain. It might be more accurate to state that these flood extents observed around the peak and early falling limb capture the average moment of transition of flows over banks into the floodplain and these are better conditions for identifying channel depth parameters.

Alternatively this divergence of findings for the optimum image time could be explained by the different experimental set up and goals. Garcia-Pintado *et al.* (2013) made use of distributed and derived water levels to correct model inflow errors and improve model predictions with assimilation, whereas identifiability here makes use of SAR derived flood extent to calibrate reach-averaged bathymetry and roughness parameters for the entire river network. Information obtained during the rising limb was the most useful time to correct inflows because the water level and channel volumes are most changing during this time. Whereas this experiment, in locating the optimum bathymetry and roughness parameters, relies on mapping of flood extent (i.e. at bankfull and overbank). This is seen most usually in the so-named 'peak' and 'falling limb' images where there is indeed flood extent but also where flows (at some locations within the model domain) are transitioning between channel and floodplain.

3.3.4 All data

Figure 8 shows the identifiability result for all 11 SAR flood maps combined and compares it with all the previous group results so far. As for the IC results, this 'All SAR' arrangement produces an observable improvement in identifiability compared with the single SAR or 'flood event' plots. Although section 3.3.1 shows that a single image does provide the information needed to locate parameter '*r*', these results show that a grouping of similarly conditioned images can locate '*r*' more distinctly and thus with greater confidence. Here the strongest identifiability is for those models with '*r*' between approximately 0.10 and 0.12. Identifiability is particularly strong for the 'All SAR' results.

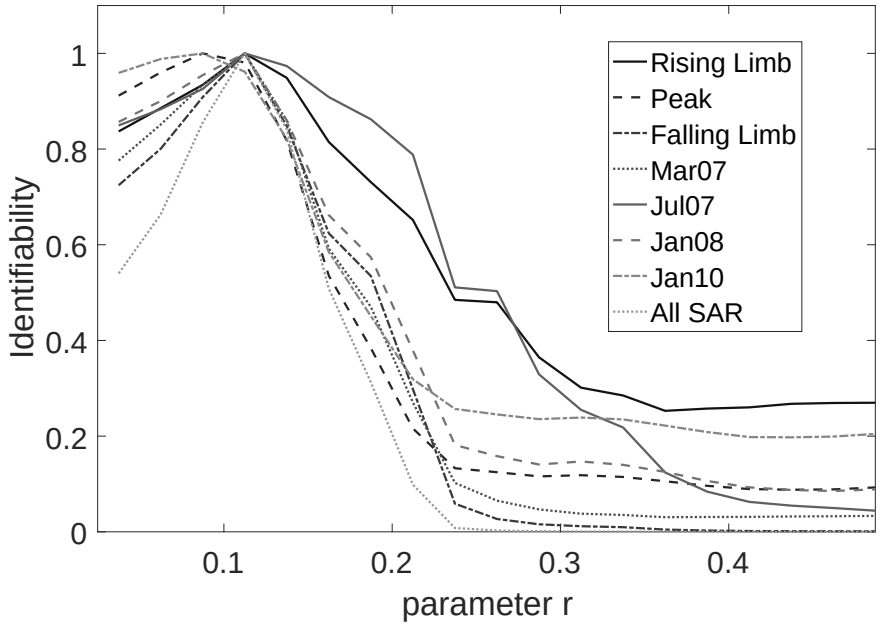


Figure 8 - Identifiability against r parameter, for all hydrographs.

These results suggest that greatest information for parameter '*r*' can be obtained by making use of as much data as is available: in other words that by simply making use of all available images the depth parameter '*r*' becomes more identifiable. Moreover 'All SAR' data mixes flood magnitudes and therefore the model is therefore likely to be more robustly calibrated for a range of event scenarios. In this instance including even relatively poor flood maps does not negatively impact the result. However, this might not always be true and situations may arise where particular flood maps (or sets of flood maps) would be disinformative.

3.4 Constraining the channel roughness parameter 'n_c'

The results above show that calibration is possible for the more dominant depth parameter but that roughness is less easily located in this simultaneous calibration methodology. So far it is assumed that no ground data are available to give prior information on either parameter and so the ranges are deliberately broad. However one or both parameters could be constrained further with some knowledge of the catchment and standard look up tables (e.g. Phillips *et al.,* 2007, Ven Te Chow, 1959). Given that even a cursory examination of Google Earth imagery shows regions of meander and channel alteration, obstructions and changing vegetation along the River Severn reach, the Manning's channel roughness values are more likely to lie between 0.035 and 0.055. This section shows that if we constrain the 'n_c' parameter to a narrower range based on physical principles and expert judgement it is possible to improve on first results. We focus here on just on the top performing models (the maximum CSI score or within 2% of it) to remove outlying model results.

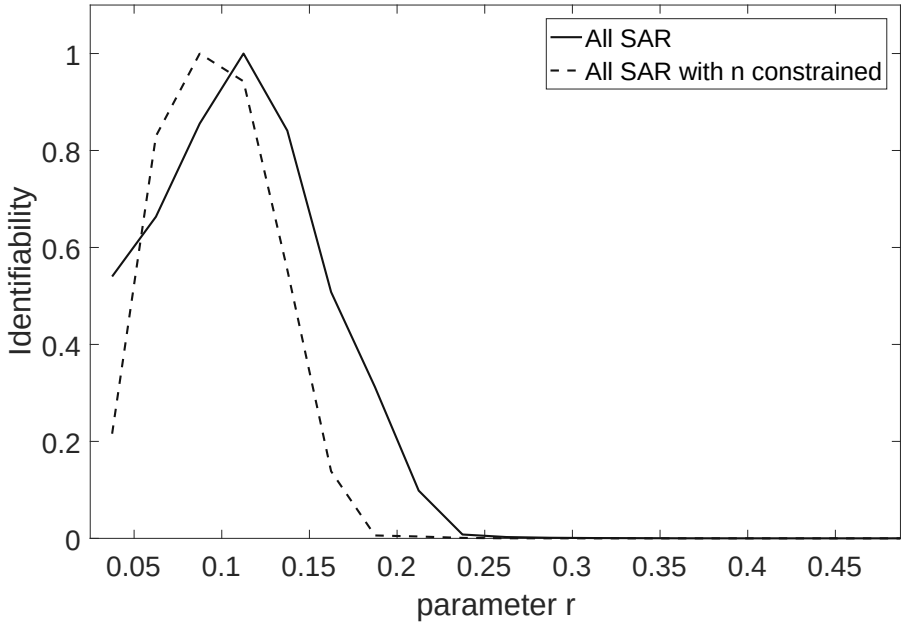


Figure 9 – Identifiability for 23rd July 2007 at 10:27 showing 'all data' (solid line) and with 'n_c' restricted to between 0.035 and 0.055 (dashed line).

Figure 9 compares the identifiability for 'All SAR' data for the full range of models (roughness is not constrained, solid line) and for 236 models which satisfy the constraint of having 'n_c' between 0.035 and 0.055 (dashed line). Where there is no constraint on 'n_c' the location of '*r*' is most identifiable between approximately 0.10 and 0.12 in 'All SAR' groupings. With 'n_c' constrained the '*r*' value moves to a lower depth range of between approximately 0.08 and 0.10. This translates to a reach-average model depth of between 6m and 7.2m and is reasonably close to the observed data. In this constrained group of models, the single highest scoring model has '*r*' of 0.086 ('n_c' of 0.036) and so indicating the optimum reach-average model depth is around 6.51m. The equivalent rectangular depth from the EA survey is 5.63m (assuming a reach median width of 76m) using bank-full cross sectional area. The difference therefore between the calibrated value and the observed equivalent is approximately 0.88m (an error of 16%).

The model responds to changes in channel friction by altering the speed of the flood wave and flow velocities. These results highlight the important reasons for calibrating this second parameter concurrently. If channel roughness were set too high the flood wave would be delayed. Set too low and the flood wave would be too advanced.

4 Conclusion

This paper presents a methodology for dual calibration of bankfull depth and channel roughness parameters of the LISFLOOD-FP Sub-Grid hydraulic model using SAR data and a binary pattern classification measure based on flood extent. Multiple models performed well initially, but by employing an identifiability methodology we located the area of the parameter space with highest information for the depth parameter '*r*'. The location narrows with the use of more SAR images.

The methodology provides some information on which single and combinations of SAR flood maps would be most useful for calibration purposes. Single SAR flood maps would be sufficient to calibrate the depth parameter but the identifiability is much improved when multiple maps are combined. Combinations aligned according to particular flood events/magnitudes are not conclusively different, but using many or all available SAR images does offer a real improvement in identifiability. There are indications that combining maps with similar flood duration, or stage of flood (i.e. SAR images acquired close to peak or just after) would be beneficial for calibrating the reach-average depth parameter, but further work is needed with more targeted observations than the 11 used here. For robustness, a good range of flood magnitudes should be used for calibration.

The channel roughness parameter 'n_c' was less sensitive to variations in flood extent and we failed to locate a representative value for this parameter when '*r*' was also varied. The likely cause probably due to the initial range selected being too broad and the suggestion that depth/bathymetry is the more dominant parameter in the model which largely overrides, at this model scale at least, the significance of channel friction. By constraining 'n_c' to a more plausible range it was possible to improve the calibration method and further improve the global estimate for the depth parameter. Under this constraint the models with top CSI and identifiability results show that the reach-averaged depth parameter is calibrated to 0.086, translating roughly to a reach-average depth of approximately 6.51m. This is an error of 0.88m compared with an equivalent measure from observed cross section data, where channel depth is approximated as 5.63m.

A benefit of this methodology is that although we used gauged inflows within the model, in theory the calibration methodology should work also with no recourse to ground data if good inflows can be simulated and a good DEM is available. The method also does not require a step to obtain water levels from the flood data. It does however make some simplifications and assumptions. First, the method assumes that as there are no errors in the return signals or processing of the ENVISAT WSM images and the derived flood maps therefore represent the true and full flood extent, however in reality all data have some error~~.~~ and this would likely impact on the identifiability and IC results here. This is particularly true for single SAR data which are compared against each other but perhaps less easily isolated in grouped SAR data as the combining of data smooths out errors and by accumulation compensates for perceived detection errors in the remotely sensed data. Understanding the impact of these individual errors on the final result would be an interesting follow-on experiment. Neither has the importance of the SAR resolution been tested here.

There is also error likely in the assumptions behind the model set up. For example, we assume that the channel depth can be approximated with a parameter '*r*' which is the ratio between channel depth to width at bankfull flow (i.e. '*r*' is a linear scaling so as width varies, so directly does depth in order to conserve water volumes). There is also the assumption that there is no rate of change between width and depth, so in essence depth and width do not vary along the modelled reach and are therefore uniform within the domain. This fixes '*r*', width and depth to a single value per model, which is applied throughout the domain. This assumption cannot truly represent the reality of channel bankfull flows at particular points in the model, so it can only be used if there is an assumption that results represent a 'reach-average' depth value for the entire modelled domain, based on a reach-average width. In this way, local variations in width, depth and flow can be smoothed out. Straight uniform channels are observed in natural systems only for short stretches of river and so the methodology may be more appropriate within smaller sub-reaches (i.e. 'sub-regions' or tributaries) where hydraulics and

hydrology are similar. or within medium sized catchments with ostensibly negligible variation in domain channel width. Future work will investigate the applicability of the methodology under these conditions.

5 Acknowledgements

We thank the Environment Agency of England and Wales for providing the river cross-section data, DEM and gauging station data. The authors would also like to thank the editor and two reviewers for their valuable comments which helped us to improve the manuscript

M. Wood's contribution was supported by the National Research Fund of Luxembourg through the PAPARAZZI project (CORE C11/SR/1277979).

References

Aronica G., Bates P.D. and Horritt M.S. 2002. *Assessing the uncertainty in distributed model predictions using observed binary pattern information with GLUE*. Hydrol. Process. **16**. 2001-2016.

Andreadis K.M., Clark E.A., Lettenmaier D.P. and Alsdorf D.E. 2007. *Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster based hydrodynamics model*. Geophys. Res. Lett. **34**. L10403.

Bates P.D. and De Roo A.P.J. 2000. *A simple Raster-based Model for Flood Inundation Simulation. Journal of Hydrology*. **236**. 54-77.

Bates P.D., Wilson M.D., Horritt M.S., Mason D., Holden N. and Currie A. 2006. *Reach scale floodplain inundation dynamics observed using airborne Synthetic Aperture Radar imagery: data analysis and modelling.* Journal of Hydrology. **328**. 306–318.

Bates P., Horritt MS. and Fewtrell T. 2010. *A Simple Inertial Formulation of the Shallow Water Equations for Efficient Two-Dimensional Flood Inundation Modelling. Journal of Hydrology*. **387**. 33-45.

Biancamaria S., Durand M., Andreadis K.M., Bates P.D., Boone A., Mognard N.M., Rodríguez E., Alsdorf D.E., Lettenmaier D.P. and Clark E.A. 2011a. *Assimilation of virtual wide swath altimetry to improve Arctic river modelling.* Remote Sensing of Environment. **115**. 373-381.

Biancamaria S., Hossain, F. and Lettenmaier D.P. 2011b. Geophysical Research Letters. **38**, l11401, doi:10.1029/2011GL047290.

Beven, K. J. 2009. *Environmental modelling: an uncertain future*? pp60. Routledge Press. ISBN 0-415-46302-9.

Bjerklie D. M., Dingman S. L., Vorosmarty C. J., Bolster C. H., & Congalton R. G. 2003. *Evaluating the potential for measuring river discharge from space.* Journal of Hydrology, **278(1)**, 17-38.

Brakenridge G. R., Nghiem S. V., Anderson E., & Chien S. 2005. *Space-based measurement of river runoff. Eos, Transactions American Geophysical Union.* **86(19),** 185-188.

Chini M., Hostache R., Giustarini L. & Matgen P. Submitted. *A Hierarchical Split-Based Approach (HSBA) for automatically mapping flooded areas using SAR images of variable size and resolution*. IEEE Transactions on Geoscience and Remote Sensing.

Corato G., Moramarco T., & Tucciarelli T. 2011. *Discharge estimation combining flow routing and occasional measurements of velocity*. Hydrology and Earth System Sciences. **15(9).** 2979-2994.

Di Baldassarre G., Schumann G. and Bates P.D. 2009a. *A technique for the calibration of hydraulic models using uncertain satellite observations of flood extent*. Journal of Hydrology. **367**. 276-282.

Di Baldassarre G., Schumann G. and Bates P. 2009b. *Near real time satellite imagery to support and verify timely flood modelling*. Hydrological Processes.**23**. 799-803.

Di Baldassarre G., Schumann G., Bates P.D., Freer J.E. and Beven K. J. 2010. *Flood-plain mapping: a critical discussion of deterministic and probabilistic approaches*. Hydrological Sciences Journal. **55(3)**.

Di Baldassarre G., Schumann G. Brandimarte L. and Bates P. 2011. *Timely low resolution SAR imagery to support floodplain modelling: a case study review*. Surv Geophys.**32**. 255-269.

Dilley M., Chen R.S., Deichmann U., Lerner-Lam A.L. and Arnold M. 2005*. Natural Disaster Hotspots: A Global Risk Analysis.* The World Bank, US, 150 pp.

Dixon J., Hannaford J. and Fry M. 2013. *The Effective management of national hydrometric data: experiences from the United Kingdom*. Hydrological Sciences Journal. **58(7).** http://dx.doi.org/10.1080/02626667.2013.787486.

Domeneghetti A., Tarpanelli A., Brocca L., Barbetta S.,Moramarco T., Castellarin A., Brath A. 2014. The use of remote sensing-derived water surface data for hydraulic model calibration. Remote Sensing of Environment. **149**. 130–141

Durand M., Andreadis K. M., Alsdorf D. E., Lettenmaier D. P., Moller D. and Wilson M. 2008. *Estimation of bathymetric depth and slope from data assimilation of swath altimetry into hydrodynamic model*. Geophysical Research Letters. 35. L20401. doi 10.1029/2008GL034150

Durand M., Neal J., Rodríguez E, Andreadis K. M., Smith L. C., and Yoon Y. 2014. *Estimating reach-averaged discharge for the River Severn from measurements of river water surface elevation and slope*. Journal of Hydrology. **511**. 92-104.

European Space Agency. 2007. ENVISAT ASAR product Handbook. Issue 2.2. Earthnet online: https://earth.esa.int/handbooks/asar/toc.html

European Environment Agency. 2012. http://www.eea.europa.eu/media/newsreleases/climate-change-evident-across-europe

European Commission. 2014. http://ec.europa.eu/clima/policies/adaptation/how/index_en.htm

Garcia-Pintado J., Neal J.C., Mason D.C., Dance S.L. and Bates P.D. 2013. *Scheduling satellite-based SAR acquisition for sequential assimilation of water level observations into flood modelling*. Journal of Hydrology. **495**. 252-266.

Garcia-Pintado J., Mason D.C., Dance S.L, Cloke H.K., Neal J.C. and Bates P.D. 2015. *Satellite-supported flood forecasting in river networks: a real case study.* Journal of Hydrology. **523**. 706-724. DOI: 10.1016/j.jhydrol.2015.01.084.

Giustarini, L., Hostache, R., Kavetski, D. Chini, M. Corato, G., Schlaffer, S., and Matgen, P., (submitted). *Probabilistic Flood Mapping using Synthetic Aperture Radar Data*. Transactions on Geoscience and Remote Sensing. Submitted.

Giustarini L., Matgen P., Hostache R., Montanari M., Plaza D., Pauwels V.R.N., De Lannoy G.J.M., De Keyser R., Pfister L., Hoffmann L., and Savenije H.H.G. 2011. *Assimilating SAR-derived water level data into a hydraulic model: a case study*. Hydrol. Earth Syst. Sci. **15**. 2349-2365.

Giustarini L., Hostache R., Matgen P., Schumann GJP., Bates P. and Mason DC. 2013. *A Change Detection Approach to Flood Mapping in Urban Areas Using TerraSAR-X*. IEE Transactions on Geoscience and Remote Sensing. **51(4).**

Gleason C.J. and Smith L.C. 2014. *Towards global mapping of river discharge using satellite images and at-many-stations hydraulic geometry*. PNAS. **111(13).** 4788–4791. doi: 10.1073/pnas.1317606111

Hall, J.W., Tarantola, S., Bates, P.D. and Horritt, M.S., 2005. *Distributed sensitivity analysis of flood inundation model calibration.* Journal of Hydraulic Engineering. **131(2).**

Horritt M.S. 2000. *Calibration of a two-dimensional finite element flood flow model using satellite radar imagery*. Water Resources Research. **36** . 3279-3291.

Horritt M S., Mason D C. and Luckman A J. 2001a. *Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model*. Int. J. Remote Sens. **22(13).** 2489–2507.

Horritt M.S. and Bates P.D.. 2001b. Effects of spatial resolution on a raster based model of flood flow. Journal of Hydrology, 253, 239-249. (10.1016/S0022-1694(01)00490-5).

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

Hostache R., Matgen P., Schumann G., Puech C, Hoffmann L. and Pfister L. 2009. *Water Level Estimation and Reduction of Hydraulic Model Calibration Uncertainties Using Satellite SAR Images of Floods*. IEEE Transations on Geoscience and Remote Sensing. **47(2)**.

Hostache R., Matgen P. and Wagner W. 2012. *Change detection approaches for flood extent mapping: How to select the most adequate reference image from online archives?* International Journal of Applied Earth Observation and Geoinformation, **19**:205-213.

Hostache R., Matgen P., Giustarini L., Teferle F.N., Tailliez C., Iffly J.-F., G. Corato. 2015. *A drifting GPS buoy for retrieving effective riverbed bathymetry.* Journal of Hydrology. **520**. 397-406.

Huntington D.E. and Lyrintzis CS. 1998. *Improvements to and limitations of Latin Hypercube Sampling.* Prob. Engng. Mech. **13(4)**.245-253.

Intergovernmental Panel on Climate Change. 2014. *Climate Change 2014. Synthesis Report.* https://www.ipcc.ch/pdf/assessment-report/ar5/syr/SYR_AR5_SPMcorr2.pdf

Intergovernmental Panel on Climate Change. 2012. *Managing the risks of extreme events and disasters to advance climate change adaptation. A special report of working groups i and ii of the Intergovernmental Panel on Climate Change*. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp

Irons J.R. and Petersen G.W. 1981. *Texture transforms of remote sensing data*. Remote Sensing of Environment. **11**. 359–370.

Legleiter, C.J. and Roberts, D.A., 2009. *A forward image model for passive optical remote sensing of river bathymetry*. Remote Sensing of Environment. **113(5),** pp.1025-1045.

Leopold L B. and Maddock T J. 1953. *The Hydraulic Geometry of stream channels and some physiographic Implications.* U.S. Geol. Surv. Prof. Pap. **252**. 56.

Marsh T. J. and Hannaford J. (Eds). 2008. *UK Hydrometric Register. Hydrological data UK series*. Centre for Ecology & Hydrology. 210 pp.

Mason D.C., Cobby D.M., Horritt M.S., Bates P.D. 2003. *Floodplain friction parameterization in two-dimensional river flood models using vegetation heights derived from airborne scanning laser altimetry. Hydrological Processes*. **17**. 1711–1732.

Mason D.C., Bates, P.D. and Dall'Amico, J.T. 2009. *Calibration of uncertain flood inundation models using remotely sensed water levels*. Journal of Hydrology. **368**. 224-236.

Mason D.C., Speck R., Devereux B., Schumann G., Neal J., and Bates P D. 2010. *Flood Detection in Urban Areas Using TerraSAR*-X. IEEE Trans. Geosci. Remote Sens. 48(2). 882–894.

Mason D.C., Schumann G.J-P., Neal J.C. Garcia-Pintado J. and Bates P.D. 2012. *Automatic near real-time selection of flood water levels from high resolution synthetic aperture radar images for assimilation into hydraulic models: a case study*. Remote Sens. Environ. **124**. 705-716.

Mason D., Garcia-Pintado J and Dance S. 2014. *Improving flood inundation monitoring and modelling using remotely sensed data.* Chartered Institution of Civil Engineering Surveyors. 34-37.

Mason IB. 2003. Binary Events. *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley and Sons: Chichester.

Matgen P., Montanari M., Hostache R., Pfister L., Hoffmann L., Plaza D., Pauwels V. R. N., De Lannoy G.J.M., Keyser R.D. and Savenije H.H.G. 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(9)**. 1773–1785.

Matgen P., Hostache R., Schumann G., Pfister L., Hoffmann L. and Savenije H.H.G.. 2011. *Towards an automatic SAR-based flood monitoring system. Lessons learned from two case studies*. Phys. Chem. Earth. **36**(7/8). 241–252.

Mersel M., Smith L.C., Andreadis K.M. and Durand M. T. 2013. *Estimation of river depth from remotely sensed hydraulic relationships* Water Resources Research. **49**. p3165–3179, doi:10.1002/wrcr.20176.

Moramarco, T., Corato, G., Melone, F. and Singh, V.P., 2013. An entropy-based method for determining the flow depth distribution in natural channels. Journal of Hydrology, 497, pp.176-188.

Montanari M., Hostache R., Matgen P., Schumann G., Pfister L. and Hoffmann L. 2009. *Calibration and sequential updating of a coupled hydrologic-hydraulic model using remote sensing-derived water stages.* Hydrology and Earth System Sciences. **13(3)**. 367-380.

Neal J., Schumann GJP., Fewtrell T., Budimir M., Bates P and Mason D. 2011. *Evaluating a New LISFLOOD-FP Formulation with Data from the Summer 2007 Floods In Tewkesbury, UK*. Journal of Flood Risk Management. 1-8. The Chartered Institution of Water and Environmental Management.

Neal J., Schumann G., and Bates P. 2012. *A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas*, Water Resources Research. **48**, W11506.

Neal J. C., Odoni N. A., Trigg M. A., Freer J. E., Garcia-Pintado J., Mason, D. C., Wood, M. & Bates, P. D.. 2015. *Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models*. Journal of Hydrology. **529**. 169-183

NRFA. 2015. http://nrfa.ceh.ac.uk/

Pappenberger F., Cloke H. L., Balsamo G., Ngo-Duc T. and Oki T. . 2010. *Global Runoff Routing with the Hydrological Component of the ECMWF NWP System*. Int. J. Climatol., **30(14),** 2155–2174.

Phillips J.V., and Tadayon S., 2006, *Selection of Manning's roughness coefficient for natural and constructed vegetated and non-vegetated channels, and vegetation maintenance plan guidelines for vegetated channels in central Arizona:* U.S. Geological Survey Scientific Investigations Report 2006–5108. 41 p.

Pianosi, F., Beven, K., Freer, J., Hall, J.W., Rougier, J., Stephenson, D.B. and Wagener, T., 2016. Sensitivity analysis of environmental models: A systematic review with practical workflow. Environmental Modelling & Software, 79, pp.214-232.

Roux H. and Dartus D. 2008. *Sensitivity Analysis and Predictive Uncertainty Using Inundation Observations for Parameter Estimation in Open-Channel Inverse Problem.* J. Hydraul.Eng. **134**. 541-549.

Schumann G, Matgen P. , Hoffmann L. , Hostache R., Pappenberger F., Pfister L.. 2007. *Deriving distributed roughness values from satellite radar data for flood inundation modelling*. Journal of Hydrology. **344**. 96-111.

Schumann G., Pappenberger F. and Matgen P. 2009a. *Estimating uncertainty associated with water stages from a single SAR image.* Advances in Water Resources. **31**. 1038-1047.

Schumann G., Bates P., Horritt MS. and Matgen P. 2009b. *Progress in Integration of Remote Sensing-Derived Flood Extent and Stage Data and Hydraulic Models*. Reviews of Geophysics. **47**. RG4001/2009. American Geophysical Union.

Schumann G.J.P., Neal J.C., Mason D.C. and Bates P.D. 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 Sensing of Environment. **115**. 2536-2546.

Schlaffer, S., Matgen, P., Hollaus, M. and Wagner, W., 2015. *Flood detection from multi-temporal SAR data using harmonic analysis and change detection*. International Journal of Applied Earth Observation and Geoinformation. **38**. 15-24.

Smith L.C., Pavelsky T.M.. 2008. *Estimation of river discharge, propagation speed, and hydraulic geometry from space: Lena River, Siberia*. Water Resources Research. **44(3).**

Stephens E., Schumann G. and Bates P. 2014. *Problems with Binary Pattern Measures for Flood Model Evaluation.* Hydrol. Process. **28(18)**. 4928-4937.

Stuart-Menteth, A. 2007. *UK Summer 2007 Floods*. Newark, CA: Risk Management Solutions. http://riskinc.com/Publications/UK_Summer_2007_Floods.pdf

Tarpanelli A., Brocca L., Melone F. and Moramarco T. 2013. *Hydraulic modelling calibration in small rivers by using coarse resolution synthetic aperture radar imagery*. Hydrological Processes. **27**. 1321-1330.

Trigg M.A., Wilson M.D., Bates P.D., Horritt M.S. Alsdorf D.E. Forsberg B.R. Vega M.C. *Amazon flood wave hydraulics*. Journal of Hydrology. **374(1)**. 92-105.

Ven Te Chow. 1959. Open Channel Hydraulics. McGraw-Hill: New York. ISBN 978-1932846188.

Vrugt J. A., Bouten W., Gupta H. V., & Sorooshian S. 2002. *Toward improved identifiability of hydrologic model parameters: The information content of experimental data*. Water Resources Research. **38(12)**. 48-1.

Wagener T., McIntyre N., Lees M, Wheater H S. and Gupta HV. 2003. *Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis*. Hydrological Processes. **17**. 455-476.

Wagener T., Camacho L.A. and Wheater H.S. 2002. *Dynamic identifiability analysis of the transient storage model for solute transport in rivers. Journal of Hydroinformatics.* **04.3,** 199-211.

Werner M., Blazkova S., Petr J. 2005. *Spatially distributed observations in constraining inundation modelling uncertainties*. Hydrological Processes. **19**. 3081–3096.

Yan, K., Tarpanelli, A., Balint, G., Moramarco, T. and Baldassarre, G.D., 2014. Exploring the potential of SRTM topography and radar altimetry to support flood propagation modeling: Danube case study. Journal of Hydrologic Engineering, 20(2), p.04014048.

Yoon Y., Durand M., Merry C. J., Clark,E. A., Andreadis K. M., & Alsdorf D. E. 2012. *Estimating river bathymetry from data assimilation of synthetic SWOT measurements*. Journal of hydrology, **464.** 363-375.