Hydrol. Earth Syst. Sci. Discuss., 3, S1289–S1300, 2006
www.hydrol-earth-syst-sci-discuss.net/3/S1289/2006/

Hydrology and Earth System Sciences Discussions

HESSD

3, S1289–S1300, 2006

Interactive Comment

Interactive comment on "Fuzzy set approach to calibrating distributed flood inundation models using remote sensing observations" *by* F. Pappenberger et al.

F. Pappenberger et al.

Received and published: 30 October 2006

Referee 1: Neil Hunter

Thanks a lot to Neil, whose comment are (as usual) constructive and challenging. They greatly helped to improve this paper. I always really appreciate him reviewing.

In this response we have summarized his initial outline:

. . . it is difficult to ascertain any direct benefit from adopting the more-computationally intensive FST approach.

The approach in this paper does acknowledge the uncertainty in the observations the traditional approaches do not allow this – this seems to be unfortunately not clear enough. Additional sentences have been added to the introduction and text.

Neither the FST or GSA aspects are explored in sufficient detail to justify their future inclusion in a practitioner's modelling toolbox.

The GSA method has been extensively documented elsewhere. The GSA analysis in this paper is only used to compare the results achieved with traditional measures with the methodology proposed here. Additional explanation was added to make this more clear.

This paper does not aim to propose a method which can be included into a practitioner's toolbox. The research into ways to incorporate observational uncertainties 'properly' into model calibration is still in its infancy.

To better demonstrate the value of the FST approach over traditional comparison measures (a central aim of the paper), a simplified calibration problem - for example, lumped channel and floodplain roughness coefficients only - could have been adopted.

True, however, the data of this exercise have been readily available and the methodology introduced in this paper is independent from the model used. Moreover, using a realistic example for the demonstration of the methodology avoids the criticism of over-simplifying and puts the problem in a wider context, which should increase paper's value.

The main criticism of the paper relates to the extremely limited discussion presented, particularly in relation to establishing the utility/efficacy of FST-GSA for flood model calibration

This is a real shame and something that I know the authors have remedied in later publications (Pappenberger et al., accepted for Journal of Hydrology). I would therefore encourage interested readers to look beyond the obvious limitations and consider the potential of these techniques for more robust uncertainty estimation of spatial flood

HESSD

3, S1289–S1300, 2006

Interactive Comment

predictions in the near future.

The quoted paper is in press and can be downloaded online. The discussion was extended and we hope that introduced changes made the paper's goals clearer to the reader.

Specific comments

1. Predicting inundation patterns versus flood depths. For this initial investigation, both the model predictions and the remotely-sensed observations against which they are compared are assumed to be two-dimensional and binary (i.e. wet/dry) in nature. Whilst this is standard practice in many flood inundation studies (e.g. Horritt, 2000; Aronica et al., 2002; Yu and Lane, 2006; Bates et al., 2006), it may be argued that comparing modelled and observed spatial patterns is of limited use for determining flood risk/hazard. Unlike a simple wet/dry classification of model results, water depth is a continuous variable and will present many more problems for generating useful hazard assessments than alluded to here. It should also be remembered that good pattern predictions often do not equal good depth predictions because of the laterally constrained nature of many floodplains. Some additional comment/discussion to this effect would be very useful.

I do think that this issue has been already partially addressed. We fully agree with Neil that there is a problem with models which only predict a good pattern (and we indeed refer directly to several references about this issue). Indeed, the study by Thieken et al. (2005) does demonstrate that water level, flood duration, and contamination are the most influential factors for building and for content damage. However, "a key element of any evaluation process is that model performance should be calibrated or conditioned on criteria that are closely linked to the purpose of the modelling exercise" (quote from the paper). And indeed, for a full hazard assessment we would not only need water depth but also flow velocities. Nevertheless, many flood maps are still presented as inundation outlines. For example, when you buy a house in the UK, the main worry for

HESSD

3, S1289–S1300, 2006

Interactive Comment

your mortgage lender is whether you house is within a flood prone area (as indicated by the outline of the Environmental Agency flood maps). Moreover, whether you can insure your house or not against flooding depends again on this outline (at least for some of the insurers in the UK). Thus comparing a model only against spatial patterns is indeed of limited use, but even when a comparison against the variables mentioned by Thieken et al. (2005) is conducted an additional comparison with the outline has to be undertaken (we only introduce a method which is promoting this). This discussion has been added to the paper.

2. Calibration strategy outlined in Section 2.4. With the exception of inflow representation, I really can't see how the sampling ranges and distributions for the various calibration parameters are justified.

The explanation of the sampling strategy was added in section 2.4. It was chosen following the preliminary sensitivity analysis, with the aim to explore the parameter space in the most effective way.

I believe there are several inter-related issues causing confusion/problems here. Firstly, though the roughness parameters are obviously effective (although this is not stated), the upper bound on the range specified for the channel is too high to be sensible within the model structure. In LISFLOOD, such high values will almost certainly swamp the influence of the other parameters (possibly with the exception of Qin), and are probably being used to compensate erroneously for errors either in the floodplain topography or in the channel bathymetry parameterisation strategy.

We argue that model parameters are always effective see references quoted in the paper.

Also, in Table 1, what is a 'log' distribution? Do you mean log-normal? If so, how is this justified? It would assume some prior knowledge of roughness that you don't appear to have here.

HESSD

3, S1289–S1300, 2006

Interactive Comment

We applied log-uniform distribution as it is skewed towards the lower parameter values. This was done in order to focus on more physically realistic (i.e. lower) roughness values while maintaining their wide spread.

Second, it is not clear how the error sampling on the cross section data is a useful or meaningful addition to the calibration problem. Sure, there will be errors in the data but fudging this assessment to work with a kinematic wave model (i.e. resampling to maintain a positive channel gradient) is not, in my opinion, the way forward.

We could have chosen to use a different routing approach (this probably should be even included into the uncertainty analysis) – However, it is not relevant for this paper and the reader is referred to Pappenberger et al. (in press-a – available online) for a more detailed discussion of this issue.

We have chosen the one which is simple, does not introduce too many parameters and is physically realistic. We think that the negative gradient for a 50 m grid would not be physically acceptable. The explanation for the choice of this distribution was added in section 2.4 (page 10).

3. Limited discussion of parameter (in)sensitivities. No conclusions drawn over the relative significance of parameters within the calibration problem. Is this not one of the intended goals of GSA?

An additional explanation in section 2.9 and a discussion in the conclusions were added. However the paper did state: This table mainly illustrates that the new fuzzy performance measure compares well to the traditional measures and thus gives us further assurance regarding its adequacy. This confusion may have partially resulted from the problem of understanding table 4.

I also don't understand Table 4.

A more detailed description has been added to the table.

4. No comparison map for Figure 7 using traditional cell-by-cell objective functions.

HESSD

3, S1289–S1300, 2006

Interactive Comment

A inundation likelihood map constructed using one of the standard performance measures would be a useful benchmark for the FST approach.

The comparison to the traditional measures has been extensive – we have compared them in their sensitivities and point patterns. Different performance measures will lead to different hazard maps of inundation patterns. This has been discussed extensively in Pappenberger et al. (in press-a – available online). This map has been produced to demonstrate that it is possible to derive flood hazard maps from the fuzzy performance measure introduced in this study.

5. LISFLOOD-centric discussion of results in Section 4.1. Many of the results are interpreted primarily in terms of the numerical model used (LISFLOOD-FP) in section 4.1. However, Section 2.2 states that it is the methodology, and not the model, that is the significant development in this paper. This problem is compounded by a number of incorrect statements about the LISFLOOD-FP code - e.g. the depth threshold for flow calculation is 1 mm and not 100 mm as stated in the text.

If the map would have been shown without explanation, why it looks like it looks, then any reader would have rightfully questioned the map. An additional discussion in the conclusions was added in order to better integrate different sections of the paper.

6. The "disconnected" areas of high inundation hazard observed in Figure 7 are probably due to large ponds/storage areas on floodplain (i.e. topography-based) rather than any model-based explanation.

Quote from the paper: " ...or possibly as a result of inundation from different directions in different model parameterisations, such that any particular pathway leading to flooding of that area does not have a particularly high hazard, ..." Added the word 'depression'.

Technical corrections

p2245, l11: Consider using "relatively" instead of "rather". Matt Horritt's work on the

HESSD

3, S1289–S1300, 2006

Interactive Comment

topic would suggest otherwise. Paragraph has been shortened and changed

p2246, l17-20: Consider replacing existing sentence with "The methodology applies the 2-D LISFLOOD-FP model within a Generalised Likelihood Uncertainty Estimation (GLUE) framework to derive the possibility distribution of inundation extent for an 8 km reach of the River Alzette, Luxemburg."

Done

p2247, l24: Consider adding "to generate inundation predictions".

Done

Second paragraph of Section 2.2 is poorly written, particularly the fifth sentence (˜6 lines long). Rephrase. . . . sorry my German slipped . . .

Done

Section 2.3 is also poorly written & structured. Section has been rephrased *The first four sentences need rearranging to make a coherent argument. What is an "acceptable level of performance"?*

It is beyond the scope of this paper to discuss this – a reference to this particular subject has been included.

More specifically: p2248, l16: Consider replacing "factors (particularly distributed parameters e.g. frictional coefficients)" with "parameters (e.g. frictional coefficients)".

Disagree, the word factors means more than merely the parameters of a model.

p2248, l19: Pappenberger & Beven reference required. p2248, l25: "framework."

Done

p2248, l26-27: Consider replacing "MC approaches consists of running repeated simulations of a model using a range of values for each uncertain input parameter."

HESSD

3, S1289–S1300, 2006

Interactive Comment

Sentence deleted p2249,

l1: Replace "more normally" with "typically".

Done

P2249, l21: What do you mean by the "the error due to the raster size"? Rephrase.

Deleted, didn't add any value/

P2251, l13: Remove the comma after "thus" and the plural of "formula" is "formulae".

comma removed. Formula is a loan word and formulas as well as formulae is correct (see e.g. Cambridge Dictionary). The 'correct' form is the one that sounds better in context (for a comprehensive explanation see e.g. http://en.wikipedia.org/wiki/English_plural). Changed to formulae.

P2252, l12: Consider replacing "remote" with "remotely sensed".

Done

P2252, l16: Consider replacing "no" with "zero" onwards.

Not changed: to me 'no' confidence sounds better

P2254, l22: Consider replacing "sub-heterogeneity" with "sub-grid scale heterogeneity".

Done

Figure 3. Pixels are shown at 50m resolution?

Doesn't matter here, but good for orientation, thus inserted into figure legend and graph.

Figure 4. Wrong graphic inserted.

Terrible sorry about that!!!!!!!!!!

HESSD

3, S1289–S1300, 2006

Interactive Comment

Figure 7. A standard/benchmark graphic would be beneficial to demonstrate clearly the advantage of the FST approach advocated here.

See above

Referee 2: Anonymous

Thanks to the anonymous reviewer for the comments.

General comments

results are little discussed and there is almost no comparison with traditional methods.

See comment above. We disagree that there is not enough comparison to the traditional measures. They are substantial part of the discussion.

Additional discussion and explanations were added to the text in the response to similar comments of the 1^{st} reviewer.

Specific comments

Section 2.4: The parameter ranges and distributions given in table 1 are poorly explained and discussed in the paper.

See comments above The discussion of the choice of parameter ranges and prior distributions was added in section 2.4

Section 3.1: What do you mean by "are as expected" (p. 2258, l12) and "behave well' in comparison to traditional approaches" (p. 2258, l13)? Maybe a more detailed analysis will be useful here. What is the potential of fuzzy performance measure with respect to classical performance measures?

Clarification added

Section 4.1: Fig. 7 is analysed with respect to the model used although the methodology is not supposed to focus on a particular model. . .

HESSD

3, S1289–S1300, 2006

Interactive Comment

See comments above

Technical corrections

p. 2248, l19: reference Pappenberger and Beven is missing.

Corrected

p. 2258, l5: is it necessary to number section 3.1 as there is no other section in part 3?

p. 2260, l15: is it necessary to number section 4.1 as there is no other section in part

Sorry, Microsoft word mixup – I should have checked more carefully – corrected

p. 2268, Table 2, 4=Odds: precise that f is False alarm rate.

Done

p. 2270, Table 4, "Only the roughness of the channel exhibits sensitivity" doesn't match what is written in the text: (p.2260, l5) only two parameters exhibit any sensitivity (Standard deviation for cross-section error and Roughness channel).

Corrected

Fig. 3: axis labels are not very legible

Disagree – makes no difference for understanding

Fig. 4 is missing.

Corrected

Referee 3 Sylvain Néelz

These are non-exhaustive comments that do not duplicate comments made by other referees:

1) The river is crossed by at least one bridge in the studied area as clearly shown in

HESSD

3, S1289–S1300, 2006

Interactive Comment

Figure 2. A road leading to this bridge runs through the adjacent floodplain and might well be set on a raised embankment (although this is unknown). Such features (bridge + embankment if it exists) will have a critical impact on both the river flow and flooding patterns on the floodplain. The paper does not mention how these critical controls were included in the model, if at all. In addition, it is very surprising that no parameter related to these was included in the author's choice of uncertain parameters. It would also have been useful to read whether the "disconnected high risk" area (p2261-l25) immediately upstream from this road is related to these features in any way.

The controls were not included as this was not the main aim of the paper. Therefore also the parameters corresponding to the mentioned features were not specifically distinguished. However, the proposed method has the potential to account for higher uncertainty of the model predictions in the specific areas on the floodplain. The research on this subject would be interesting but was beyond the scope of this paper.

2) The paper is also yet another publication where the potential of SAR images as calibration data may be overestimated. Many SAR images are very difficult to interpret typically in areas of high vegetation, and misclassification errors are rife (due to vegetation, but not only). While the "snake" algorithm in itself is not to blame, it is only as good as SAR images are. The worst case scenario would be for example a floodplain with vegetation in higher areas only. A calibration based on SAR images of such a floodplain (with misclassification errors) would artificially force water levels down and result in totally unrealistic parameter values. The best case scenario is a vegetation free floodplain. It is not clear whether the case reported in the paper can be placed in relation to these two scenarios, but the possibility of significant inaccuracies in the observation data should perhaps have been considered more specifically.

We agree with the reviewer, there is a range of real life scenarios which are difficult to take into account using a general (not event specific) model. There is also a question of the level of accuracy required by the model in the presence of very uncertain observations of flood inundation extent. The aim of the paper was to show that these inaccuracies can be modelled using fuzzy set approach and incorporated into the process of the derivation of risk maps.

Others:

On figure 7 the scale is wrong. It should probably say 1000m, not 100m.

Corrected

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 3, 2243, 2006.