The authors sincerely thank the reviewers for their elaborate review and their useful comments. First we will give a reaction to the general comments of each reviewer. Also, the fact that the method is deemed innovative and of high interest strengthens us in our future research aims.

The specific comments that we have replied to will hopefully give an insight in how we will change these in the next version of the paper. For the specific comments that we have not replied to we will change the texts as asked.

GENERAL COMMENTS

Comment I (Matgen):
"First, we regret that the authors made the choice to present an operational service rather than a scientific method. This means that arguably there are some parts in this paper that are not very relevant in the context of HESS (e.g. production of Google maps, discussion on the handling of current and future data amounts, speed of image processing etc.) and that blur the paper's significant scientific message."

Answer to comment I:
We will focus on presenting the scientific method and leave out most of the operational service information, which are not relevant in the context of HESS.

Comment II (Matgen):
"It also means that in our opinion the authors did not carry out a convincing quantitative evaluation and sensitivity analysis of their scientific method. This is a pity as in our opinion this would be of much higher interest for the community than the more technical aspects of the operational service (at least in the context of HESS). At one point there is a visual assessment of the flood probability maps. This purely qualitative evaluation does not provide much insight into the performance of the method. The title of section 10 ("Validation of the method : : :") is misleading and does not seem appropriate. At no point the authors compute performance indices. Is there no data set available that would allow for a quantitative evaluation of the algorithm's performance? Would it not be possible to compare the maps to those obtained with other methods?"

Answer to comment II:
In an earlier stage we ahd already discussed this issue. In the project in which this research was performed we have done validation work, but we chose to focus on the NRT possibilities of the operational service instead of the validation. As both reviewers have let us know that a validation study is more relevant to the HESS journal, we will add this validation. This also means that we add another author (C. Bishop) who has done most of this validation.

Comment III (Matgen):
The third concern relates to the way the authors approximate the prior knowledge that a pixel within the SAR scene is covered with water or not (see Equations 1 & 2). In fact the authors state that they have no prior knowledge and that therefore the probability is

set on 0.5. Clearly, this prior largely overestimates the probability of a pixel being covered by water (basically this would mean that half of the Earth's terrestrial surface is covered by water). A much better prior would be in our opinion to consider the number of permanent "water" pixels (e.g. using the same mask that was used to build the empirical histogram of water surfaces) and to divide it by the number of pixels covering a tile. Under these assumptions the simplifications that lead to equation 7 would be no longer valid and, as a result, the posterior probabilities of the individual pixels being covered by water would change. Moreover, the fact that the authors compute histograms of water and dry pixels on different areas needs more explanation and discussion. Indeed, the fact that these two histograms are not computed on the same population (global scale for water and tile scale for dry land) could render the Bayes' law not applicable. The statistical distribution of water is probably stable over the globe but then authors have to demonstrate that it is the case.

Answer to comment III:

A: We have tried to translate our existing algorithms in a straightforward Bayesian approach. We believe that by referring to a Bayesian approach, we can better explain the probability calculation. We will try to follow this approach taking into your comments (here and in the general comments) and change this. If this turns out to be a problem, we will go back to our original approach of defining the calculation of the water probability.

Comment IV (Matgen):
In general, it would be also beneficial for the manuscript if the authors would be more critical about their own approach. What are the inherent weaknesses of their method if it is compared to other established SAR-based flood mapping methods? For example, we would hypothesize that it is difficult to apply this method to very high resolution SAR imagery (because of the difficulty to have a sufficiently large samples of images and because of the speckle). Does this method enable the detection of water within urban settlements or under vegetated canopies? Or when wind roughens the water surface? For all these reasons, we recommend accepting this contribution, subject to some moderate revisions.

Answer to comment IV:
The authors think that by adding a validation paragraph and an extra discussion on this validation, where another SAR based classification will be used, will help for a more critical analysis on our approach. In the discussion, we will also focus on answering the reviewer's questions on urban settlements, canopies and wind.

Regarding the hypothesis on application on high resolution SAR we actually think that our approach could result into larger bandwidth of the histograms, which could result into lower probabilities in areas that are flooded. But this will also reflect the ability the quality indicator, which in our vision is the first step of putting an uncertainty on a flood classification. We think this method used on HR data – and especially the ability of defining the uncertainty - will open the door to use of these maps in other applications, such as hydrological forecasting.

Comment V (Matgen):
BTW: Another recommendation would be to make two papers based on this material: the first paper could deal with the scientific method and its critical evaluation and the second paper could deal with the operational service.
Answer to comment V:
We are indeed aiming for another paper in which we explain the operational service part (It-infrastructure and Google Earth part). But this will not be in the HESS paper, so we propose to leave this out of the discussion for now.

Comment VI (Wagner):
"Yet, I agree with the review by Patrick Matgen that the authors probably would have done better to concentrate more on the scientific aspects of this service rather than discussing system characteristics. My recommendation is to shorten the discussions of the system properties (e.g. drop sections 8 and 12) and add a much more detailed discussion of the strength and weaknesses of the approach. This shall be done by carrying out additional validation activities."

Answer to comment VI:
The authors are happy that the reviewers share the vision on the moderations needed. As said, we had chosen for the operational part and intentionally left out the validation, but this has shown to be the wrong approach for a HESS publication, which we agree upon at this stage. We will drop most of the information stated in sections 8 and 12, as long as it does not hinder the scientific aspects. Also we will shorten the discussion on system properties and add more discussion on strengths and weaknesses by adding additional activities.

Comment VII (Wagner):
Also, the authors may consider to compare their results with the ones obtained by other published algorithms. I realize that it is quite difficult to obtain reliable reference data for validating SAR derived water maps, yet one advantage that the authors may make use of is the fact that their system allow global processing of the complete ENVISAT ASAR archive. This should allow them to identify a number of floods, for which reference data and flood maps have been published.

Answer to comment VII:
The authors hope that by adding additional validation from another SAR based method (which we will refer to), but validated in more than one way with different thresholds for water probability and quality indicator will be sufficient.

Comment VIII (Wagner):
The paper is generally well written and easy to follow. However the appearance of the figures should be improved in my view. Overall, I recommend a major revision of the manuscript

Answer to comment VIII:

We assume that the comments on the appearances have been mentioned in the specific comments and of course will handle these accordingly.

SPECIFIC COMMENTS

(Matgen) Introduction: the state of the art of existing SAR-based flood mapping methods is not exhaustive. Some important papers seem to be missing (e.g. Mason et al., 2010, Pulvirenti et al., 2011, Martinis et al., 2009, 2011).

A: Thanks for this, we will go through these papers and add them where possible,

(Matgen) p.7808 l.25 we cannot understand the range of values given here. The values in dB have to be much lower (i.e. negative values for backscatter on open water). This seems to be a mistake.

A: We think you mean p7807? If so, there indeed seems to be an error here, as indeed some lower values for water seem to exist and we have probably looked only at land. We will look this up and give clarification in the texts.

(Matgen) p.7809 l.20 We think it is no justifiable to have the same priors. Your water mask (p.7807) enables you to compute more adequate priors per 1 degree tile.

A: We have tried to translate our existing algorithms in a straightforward Bayesian approach. We believe that by referring to a Bayesian approach, we can better explain the probability calculation. We will try to follow this approach taking into your comments (here and in the general comments) and change this. If this turns out to be a problem, we will go back to our original approach of defining the calculation of the water probability.

(Matgen) p.7810 l.13 Do you have to parameterize the pdfs of the land and water empirical pdfs? The subplots in figure 2 do not seem to support your assumption of a normal distribution.

A: They are not normal distributions. We made a typo here and will correct for this.

(Matgen) p.7811 How did you select the threshold value of the HAND index? Since you are not carrying out a quantitative evaluation of your output maps, it is difficult to evaluate the correctness of your value. At the very least a sensitivity analysis should be carried out.

A: We assumed that we had explained that on p7815, line 9. We selected these by trial and error with an analysis of different values. Since the entire part of selection of the HAND resulted in several pages, we tried to shorten this into the compilation as it is now. We chose to do this, because we thinks it is important to filter out the more extreme topography: we think it is important not to underestimate the HAND – as flood plains will be filtered out then - but it is less important if the intermediate values are still in: values in between flood plains and mountains will not be filtered out, but probably not

classified as water too. We propose to explain that a rough sensitivity analysis has been done to be sure that we are not filtering out flood plains.

(Matgen) p. 7812 l.5 What happens if smaller tiles are considered? Is there any advantage to consider smaller tiles? Does it have any impact on the results (other than computation time)?

We mainly chose 1 x 1 degree for the sake of computation time. Besides computation time, considering smaller tiles will result in a longer 'training time', or put otherwise a longer building phase of the training histograms. It could however result in better classification. We have also thought of building the histograms not for cells of x by x degree, but for land cover, using a global land type map. However, by doing this we will add another data source that can also contain errors which we cannot quantify.

(Matgen) P.7813 l.2 and in general: in the context of a systematic SAR-based flood mapping application it might be useful to indicate the number of images that are available across the globe. We assume that in some areas the sample size is too small to carry out the statistical analysis. Is this something one has to take account of? You mention that during the flooding event ESA "switched on" the image mode. But do you have enough historical data to build the training data set for this image mode?

Excellent question and suggestion. We do log the amount of images for each ASAR mode used in the training histogram. As in Thailand flooding is more frequent a training histogram already existed. But is has to be said that it could form a noise problem when there is a flood: the IM mode is switched on that and pdf (our training histogram) will be affected by that. In other words, we get more 'noise' on our land training set. In general, for GM and other modes, that will cause a small amount of noise (as also mentioned For IM mode for example areas that are trained as land were actually flooded during the making of this training histogram. This causes the training histograms to be more distorted, to have more noise. Ideally, this would have to be corrected by having someone manually leave out historical flooding periods during training. And for GM mode, if there is a flood it cannot be classified, as there is no GM image.

Nonetheless, we will add an extra paragraph on this source of noise and the relation to other sources of noise.

(Matgen) p.7813 l.11-18: in our opinion, these observations cannot be considered as "validation".The same comments could be made with respect to many other SAR-derived flood inundation maps. The fact that some elevated features can be identified on the map and that according to the algorithm's result the city of Bangkok did not suffer extensive flooding hardly validates the method. Do you not have any other study area with more useful ground information, aerial photographs or very high-resolution satellite imagery that could be used for a more meaningful evaluation?

The description was meant as general description of things to be seen on the classification image. As said, we will add a validation.

(Matgen) p.7814 The authors conclude that their algorithm is useful to detect floods in cloud covered areas. However, we do not see any hard evidence that would support this conclusion.

We will try to explain why we think there is evidence in the moderated version. Summarized this will be something like this: the structures to be seen make it hydrologically more explainable that the SAR image is right than that the MODIS image is right. We have proof of large areas being cloud covered in the area by MODIS (which we will show), so we conclude that the MODIS image is cloud covered.

(Wagner) One assumption made by the authors is that temporarily flooded areas have the same backscatter distribution as permanent water bodies. However, it should be expected that the extracted water class histograms are heavily influenced by large water bodies. Wind probably has a stronger impact on large water bodies than on smaller ones so therefore the histograms derived for those water bodies are probably not representative for small flooded areas. While the authors made the decision to leave out oceans from the training dataset they used measurements from large lakes which arguably are more similar to coastal waters of oceans than to relatively small scale flood events. Because of these reasons, it would be imaginable to derive water histograms for different climate zones and to mask out very large water bodies.

In the article, we show that by classifying systematically without too much prior knowledge derived from other data sources (also possibly containing noise, which is most of the times unquantifiable). As rightly suggested by the reviewer, further improvements are indeed possible by adding other water histograms (deep water, shallow water, etc.). Also, adding local water histograms per climate area would also be an excellent idea and we thank the reviewer for the suggestion. However, we chose to prove that a systematic approach without too much detailing per climate area (meaning without manual and expert knowledge with unquantifiable errors) already can be successful. Also, we can derive that area that has just been water-logged has a significant different backscatter characteristics than water or land: it is somewhere in between. The algorithm will thus give the water probability somewhere in between (like 50%).

We will try to elaborate on the issue in the discussion and thank the reviewer for new inspiration.

MINOR COMMENTS
We will also correct and explain these minor issues without further explaining in the moderated version:

(Matgen):

Fig. 2 A legend is missing. Moreover, we doubt that on the y-axis these are backscattering

values in dB.

Fig. 3 A legend is missing here to understand the colour coding.

Fig. 4 The units are missing.

Fig. 7 The font size should be increased.

Minor comments:
p.7802 l.7 please reformulate this sentence as there can be no "probability distribution function of a pixel" p.7802 l.23 I would add "accurately". In general I would put more emphasis on "accuracy" and "reliability" than on the fastness of the service (also in the conclusion). p.7804 l.6-7 The meaning of this sentence is not very clear. It would be useful to add some examples of how the maps can be used in the context of flood inundation modelling (see for example: Schumann et al., 2008 for an extensive review on the subject). p.7804 l.1-16 It is also worth mentioning the problem of shadow and layover hampering the detection of water (see for example Mason et al., 2010) p.7805 please reformulate "an army of human operators" p.7805 l.13 The acronym NRT should be introduced here rather than in l. 17 p.7806 l.18 distribution p.7806 l.23 "degree" is missing here p.7806 l.27 "permanently dry" does not seem to be an appropriate term in this context. These pixels can be flooded. p.7807 l.2 the processing of smoothing needs to be described somewhere. p.7808 l.10 Could you state that the underlying working hypothesis of lumping all freshwater values is that local effects are neglected (e.g. some areas are more exposed to wind than others, some water bodies have more or less sediments or contain more or less salt – all these factors may influence the shape of the backscatter histogram). p. 7810 l.3: it might be useful to formulate this condition as this helps the comprehension of Eq. 6 p.7813 l.6 This is not clear as in Fig. 9 all the water probabilities are provided. In general we do not understand why the authors use a 70% probability for computing the binary map. Should it not be 50 % (i.e it is more likely that the pixel is flooded than not flooded) ? p.7813 l. 9 delete "is" p.7813 l.20 it might be useful to introduce the "Global Flood Observatory" and its acronym (is given later on p. 7814) p.7814 l.22 What is the "complementarity" of the two mapping methods?

(Wagner):

p. 7805, l. 1: "unless an army of human operators" is too colloquial
p. 7806, l. 11: only altimeters looks straight down, so this statement does not seem to be relevant here.
p. 7806, l. 27.: The authors state that empirical distribution functions were estimated for "pixels within a 1 x 1 degree tile, which are permanently wet". This contradicts p. 7808, l. 9-10 according to which one global histogram for the freshwater class is made.
P. 7807, l. 15: "in between 1.55 and 4.25": are these numbers correct, or maybe I do not understand the meaning of this sentence.
p. 7810, l. 9 ff.: The histograms are derived for each 1x1-degree tile. Does this mean there is only one quality indicator q per tile and polarisation? In fig. 8 (right) q is

presented for each pixel. Please make this clearer.
p. 7813, l. 19 ff.: Is this supposed to validate the HAND index or to show the plausibility of the resulting flood maps? Validation of HAND is not the focus of this paper, what is missing is rather a proper validation of the flood mapping algorithm.
p. 7813, l. 15: Did you expect to be able to detect floods in urban areas? Although the moderate resolution hardly permits mapping floods here flooded urban areas would be expected to show higher backscatter values due to double-bounce effects (Mason et al. 2010). The algorithm does not explicitly address this possibility.
Fig 8 left: a more linearly spaced colour table would be desirable here. The colour table in use here is actually more of a threshold.
Fig. 9: From a scientific point of view it would be more interesting to show the flood probability than the thresholded map. Also the quality indicator would be useful.