We thank reviewer #2 for his/her valuable comments. We have implemented the recommended suggestions. Doing so the quality of the paper has greatly improved. Point-to-point answers to the comments are given here below.

Comment 1 :

The paper has a very unclear structure and non-informative naming of sections. There seems to be a mixture of general information, methods and results present in each of its sections. A strict editing for sequence of information and content would make it much more readable. For example, section 2.2 promises an overview of RS in modeling of inundation, titles of subsections suggest it is a description of methods/data used in the paper, while the contents of subsections delivers a combination of both. Section 4 contains elements of methods, data and results. Section 5.1 suggest description of available data, while in fact it is a description of methods used in the paper. I would suggest sections 4 and 5 to be reworked into a clear and succinct materials and methods, with parts relating to "preanalysis of hydrological data" moved to the result section.

The paper also contains several unnecessary elements. For example, RMSE and MAE and Nash-Sutcliffe coefficient are indices that do no have to be defined, It will be sufficient if the authors describe with adequate precision what model variables were these indices determined for.

NB, is seems to me that AME is an inappropriate name for the index the authors use, and should be corrected, see one of the comments below.

Answer 1:

As suggested by reviewer 2 and the other reviewers, we have restructured the paper in a more "classical" manner, which has greatly improved its clarity. Similarly, the naming of sections has been edited to better reflect the content of each section.

The paper now has the following structure:

1/ Introduction: Wetlands are important ecosystems under threat. Lumped hydrological models, in combination with satellite data, can be used to characterize the flooding events for poorly gauged wetlands. The objectives of the study are then given, followed by a rapid presentation of the structure of the paper.

2/ Presentation of the study site (the Tana River Delta). We first present the social context, then the hydrological, climatic and pedological characteristics that are important to undertake hydrological modelling.

3/ Material and methods. We first present the general strategy used to combine hydrological modeling with remote-sensing techniques, then develop these two sections. 4/ Results. This section follows a similar sequence as Material and methods. We first focus on the remote-sensing results (use of water index, time series of flood extents, then the spatial extent of the floods), then detail the results related to the hydrological model (calibration and validation of the two parts of the model, then quantification of the hydrological variables of interest).

5/ Discussion and conclusion. We discuss the results pertaining to the remote-sensing data, then the hydrological modelling, then we discuss the utility of the results of this study for decision makers.

We have defined RMSE, MAE and the other likelihood functions (all indexed with epsilon) in Table 4. We consider it important as the epsilon used in this study is slightly different from the usual error used for these indices.

Indeed, we had mistermed in the text AME as Absolute Mean Error instead of Absolute Maximum Error, which we have now corrected.

Comment 2:

The concept of flood (inundation) frequency as authors use it seems to be somewhat weak. Inundation frequency is not precisely defined in literature, but a common

understanding would allow for the following definition: a number of times an area changes status from non-inundated to inundated during a certain period of time (a year, perhaps). According to eq. 3, the authors express inundation frequency as a fraction of images in which a pixel was flooded (inundated). Considering the temporal characteristics of flooding in the studied system, where in most of the area there are principally two flooding events during a year, it is possible, or even likely, that a single inundation event is counted more than once. Under assumption of uniform temporal distribution of images from which inundation is derived, fi of Eq. 3 expresses percent of time inundated and not inundation frequency. Additionally, whether the images are uniformly distributed in time. Although they state that the images are well-distributed throughout the year, they repeatedly stress that the time series of MODIS images is "discontinous". Data are not presented, but there are 48 8- day composites throughout a year, which gives 405 images for the 9-year period. The authors analyse only 67 of these. Considering the cloudiness-rainfall-flood association, it is difficult to believe that the flood/no flood periods are indeed uniformly sampled by MODIS flood maps. In fact, the statement from page 11289: " The total surface with a frequency fi >0.01 is of 450km2, which means that each corresponding pixel was flooded over four times in 2002-2011.", which is impossible to interpret, only confirms the problems with the approach adopted by the authors. Perhaps the term "flood frequency" should be properly defined in the beginning, or a different, more appropriate term should be used instead.

Answer 2:

We agree with Reviewer 2 that the term "flood frequency" is ambiguous and can be misleading (understood as the number of times an area passes from a non-flooded to a flooded state, instead of the fraction of times a pixel was in a flooded state). We have therefore used a more appropriate term "empirical frequency of flooded state", which has been defined in the text "This term expresses the fraction of times a pixel was classified in a flooded state, considering the observations of the pixel when it was classified as non-clouded", and in equation 2.

The sentence "The total surface with a frequency fi >0.01 is of 450km2, which means that each corresponding pixel was flooded over four times in 2002–2011." was re-phrased within the results and discussion sections.

Comment 3:

In section 4.3.2 the authors describe a measure of uncertainty of inundation mapping, and this measure is subsequently used to assess the quality of the hydrological model (section 5.3). Table 5 shows that this measure is determined as a mean percentage of dry pixels that were classified as inundated, and a mean percentage of inundated pixels that were classified as dry in classification of three images. The authors equal these to underestimated and overestimated inundated area. These two are, however, not the same, and the percent indices derived in this way cannot be used in the context presented in fig. 4, i.e. to set upper and lower bounds of derived inundated area. Additionally, these uncertainty bounds are entirely arbitrary, i.e. there is no clear relationship between them and characteristics of distribution of the error such as standard deviation of inundation size (if one assumes error to be normally distributed) or usually used percentiles (e.g. 10th and 90th). Because of that, all errors (AME, MAE, RMSE) calculated from RE are essentially meaningless in absolute terms. The concept of error of inundation classification has to be revised.

Additionally, for the sake of consistency values of "% underestimation" in table 5 should be calculated accounting for errors of classification for 25 May 2009, even though there was no flooding during that period (e.g. % underestimation for th=0.09 should be around 10% and not 15%.

Answer 3:

The authors agree with this essential comment of Reviewer 2. Percents of misclassification can be calculated from the error matrix derived from comparing the field data and the classified MODIS images (as 100 minus the user's or producers accuracy). However, the percent of misclassification (be it calculated from the user's or producer's accuracy) is representative of the population of pixels sampled, and not of all the pixels within the study area for 2002-2011.

We therefore used a different methodology to estimate the uncertainty due to the classification of the pixels. A numerical simulation was carried out: each pixel classified as flooded was sampled 1000 times, and attributed a flooded or non-flooded state with a probability corresponding to the percent misclassification of flooded points using the user's accuracy. The same was done with the non-flooded pixels. For the clouded pixels, we used the probability corresponding to the ratio of observed pixels classified as flooded to the pixels classified as non-clouded. From these simulations, we calculated 10th and 90th percentiles to delimit an uncertainty range relative to the classification of the pixels. In the process, we improved Table 5 and the explanations relative to the classification error. We used the appropriate terms of producer's, user's and overall accuracy of the classification as defined by Congalton (1991). We no longer use "underestimation" or "overestimation" but "upper and lower uncertainty boundaries".

Comment 4:

The model has been run on hourly basis, which seems absolutely not compatible with the nature and amount of data used, and the nature and variability of hydrological processes in the study site. It would be prudent to provide justification for this hydrological modeling "overkill".

Answer 4:

We agree that the time-step at which the model's results can be analyzed is constrained by the time-step of the input variables. In our study, the main variables (e.g. MODIS, discharges with a decadal floating mean) are approximately on a 10-day time-step basis. Therefore, the results are averaged over this time-step.

The downscaling of evapotranspiration and rainfall at Garissa is justified a posteriori by the fact that their contribution to the water-balance is minor. Hence, even if the temporal distribution within the month was slightly different, this would not affect the results of the study. Finally, the downscaling of monthly values to daily values is justified in our case by the fact that no daily or decadal evaporation or rainfall (at Garissa) data were available.

The time-step at which the results were analyzed is different from the time-step at which Equation 5 was solved. The latter was solved at an hourly basis, so that the numerical resolution of this differential equation (by RK4) was stable. Hence, the equation was solved at an hourly basis, but the model is run on a 10-day basis.

Specific comments:

1/ p. 11268, line 21: the model is not based on GLUE, the model is calibrated using GLUE approach

Ans:

This sentence was rephrased.

2/ p 11269, line2: what's the hypothesized reason of wetland decline? Line 10: flood extent, timing, frequency etc. are not "hydrological processes in the adjacent river". Rather, these are characteristics of flooding or hydrological conditions.

p. 11270 line 22 "Important environmental questions..." what questions? In what context?

Ans: Explanations concerning the decline of wetland area have been added in the introduction, and the sentences rephrased for clarity.

3/ p. 11257, line 22 "precipitation would usually occur before flooding extent" seems to be a truism, while "precipitation does not induce the floods" sound counterintuitive, and perhaps need explicit explanation. It seems that distinguishing between "local", i.e. precipitation over the wetland itself, and "upstream" precipitation would make the above statements much more precise.

Ans: These sentences have been restructured, incorporating the distinction between local and upstream precipitation, as suggested by the reviewer.

4/ p. 11281, first paragraph: "to determine threshold . . . that best differentiated. . ." - it is not very clear how the "best" value was obtained. It is also not clear what the "upper and lower uncertainty range" pertain tom and how they were determined. One would assume that the upper and lower uncertainty bounds describe a range of estimates of inundated area, obtained by using a certain, non-optimal values of the threshold. Or does the uncertainty pertain to the threshold itself? How exactly were these bounds calculated? Also "the upper uncertainty range was increased" – by how much? The authors have to be more specific here.

Ans: "to determine threshold . . . that best differentiated. . ." was rephrased as "The NDWIGao,M2–5 value that maximized the percent of correctly classified pixels (i.e. the overall accuracy of Congalton, 1991) was retained as the threshold value." The other questions have been commented in Answer 3.

5/ p. 11281, line 19: "frequency of floods corrected for cloud cover" – it is an awkward formulation. The flood frequency is not corrected for cloud cover per se. Rather, the method of calculation of flood frequency takes into consideration discontinuity of inundation maps that is caused by cloud cover.

Ans: this sentence was deleted during the restructuring of the paper

6/ p. 11283, line 7: "the ratio of flood extent to the visible zone was identical to the ratio of flooded extent to the clouded zone" – perhaps what the authors meant was that "ratio of inundated to non-inundated area in the visible zone was identical with that ration in the clouded zone"

p. 11287, line 8: "Parameter sets which did not allow a full resolution of differential equations. . ." what did the authors mean exactly? Parameter sets for which solution could not be achieved?

Ans: these suggestions were incorporated into the text.

7/ p.11283, line 1: "temporal frequency of flood extent" – this is a confusing term. Equation 4, as explained in the text does not give any expression of inundation frequency. Rather it gives a mean inundated area for times depicted in flood maps only.
Ans: We agree that the term "temporal frequency" is inadequate. We have replaced it by "empirical probability of flood extent" (which is also called the relative frequency). Equation 4 has been split into two equations to better reflect the calculation procedure.

8/ p. 11284, line 15: "..coefficient a was expressed as a function of initial condition corresponding to the surface of permanent lakes'. It seems to me that the "reference" for coefficient a is "the surface of permanent lakes", and not the "initial condition". That initial condition is arbitrary and just happened to correspond to the stage at which only permanent lakes are filled with water. Expression " coefficient was expressed as a function" gives a wrong impression, it suggests a is a dynamic variable. It seems to me a is static and had to be calculated from volume of water stored in permanent lakes and the area of permanent lakes.

p. 11287, line 25: ".. each parameter gives a percentile for every time step." – what does it mean?

Ans: We agree with the reviewer, that a is a constant, and that it should be calculated

from the surface of permanent lakes rather than the more arbitrary initial condition. We therefore rephrased, so that "a was calculated from the surface of permanent lakes S_L (4 km₂), the corresponding water height Z_L , S_{max} and r. Initial flood extent was set to S_L ". The sentenced "... each parameter gives a percentile for every time step." was deleted in the new version of the paper to avoid confusion.

9/ p. 11288, equation 14: conditions describe in that equation are not exhaustive and mutually excluding. It is not clear how RE is calculated from REÂn⁴ RE5 and RE6. p. 11288, As we set the relative error to nil each time the respective uncertainty ranges overlap, the indicators will have a higher value.

p. 11289 , line 23: "The total surface with a frequency fi >0.01 is of 450km2, which means that each corresponding pixel was flooded over four times in 2002–2011. ". It is difficult to understand how fi of 0.01 translates to 4 times in 9 years. Some additional explanations are needed.

Ans: Equations (15) and (16) have been checked so that they are mutually exclusive. The text has been re-arranged and a longer explanation on the error is given (section 3.4 of the new version).

The other sentences have been rephrased or deleted.

11/ p. 11306, table 4. The formula for "Absolute Mean Error" suggest that this term should rather be called "Maximum Absolute Error"

p. 11293, section 6.3.4: perhaps terms "flow" and "inflow" can be replaced with "flux" and "input"? Otherwise the sentence: "A mean absolute difference of 0.14 % between the incoming and outgoing flows for all parameter sets was observed and attributed to the use of numerical resolution and its related approximations. " can be interpreted as meaning that components of water balance other than river flows, i.e. rainfall, infiltration and evaporation were negligible.

Use of terms: I suggest replacement of the terms listed below with alternatives:

p. 11268, line 9: "water height' – "water level" or "water depth" (also elsewhere in the document); line 20: "primordial" – perhaps "a prerequisite"?

p. 11270, line 3: "rapidly" – perhaps better "easily", line 4 "parsimonious" – perhaps "simple"

P. 11271. Line 23: "uniqe" – "single" (also elsewhere in the document); line 24: "calibrated with . . .data" – "calibrated against ... data"

p. 11276, line 5: "here and there, small islands" – " small island present throughout the system.."; line 10 "inundations" – "flood water"

p. 11276 line 23 "chronic" – "time series", "missing data were completed" – "gaps in record were filled"; line 24 – "coinciding with flood event" – "when flood events occurred" p. 11277 line 1: "lacunary" – "patchy" or "discontinuous"; line 6: "stage board" – "gauge plate"; "rating curve is under construction" – "measurements are carried out in order to establish rating curve"

p. 11278 line 3 "reconstituted" – "reconstructed"; line 7: "varying propagation trends" – "differences in propagation trends"; line 11: "floods" – "flood events"

p. 11279, line 1: "discontinuous flood surfaces" - "flood surfaces"

p. 11281, line 20: "discontinuous time series" – "irregular time series" or simply "time series" (also elsewhere in the document)

p. 11282, line 16: "for each non-masked pixel" – "for each pixel classified as non-clouded" $% \left({{\left[{{{\rm{D}}_{\rm{c}}} \right]}_{\rm{cl}}} \right)$

Ans:

We thank the reviewer for these definitions, and have incorporated the modifications in the text.