

# Rebuttals to the interactive comment on “Modelling stream flow with a discrete rainfall–runoff model and 37GHz PDBT microwave observations: the Xiangjiang River basin case study” by Haolu Shang et al.

Anonymous Referee #1

Received and published: 10 August 2016

In this paper, the authors develop the new rainfall-runoff model. They introduce the newly developed inundation data product into the semi-empirical hydrological model. They improve the skill of simulating streamflow using the observed groundwater storage and the inundation data. Unfortunately, I recommend the editor not to include this discussion paper in the HESS journal. Although the topic of this paper is suitable to HESS, I think the design of this study has several major deficiencies which may not be fixed in a short period of time.

**Q1:** The most important conclusion of this paper is that the retrievals of Water Saturated Soil (WSS) and inundated area improve the model accuracy. However, I cannot agree with it. The model performance should be evaluated in the calibration/validation process. The authors should not conclude that they can improve the accuracy just by the results in the calibration period. Figure 12 clearly shows that there is no improvement from their first implement to second implement in their model validation. Their validation clearly rejects their hypothesis.

**Answers:** Thanks a lot for this comment. In the revised manuscript, we clarify one of our goals is to evaluate whether using the retrieval of Water Saturated Soil and inundated area can reduce the required duration of antecedent precipitation in the discrete rainfall-runoff model. The duration determines the number of the model parameters. Long duration will lead to the problem of overfitting, thus a relatively short duration is preferred in the implementation of the discrete rainfall-runoff model. A duration = 80 days was applied to calibrate the second and third implementation. A similar model performance was achieved as that of the first implementation but using a duration = 150 days. The validation results showed that the second and third implementation required a duration of 10 days to achieve a similar model performance as the best one of the first implementation that required a duration of 90 days. Above all, we concluded that applying the WSS and inundated area can reduce the required duration of antecedent precipitation, without degradation of model performance. Accordingly, we modified the text as follows:

Page 1, Line 23-28, in the abstract, we illustrated the results of validation and stated our conclusion as:

*“All model implementations were evaluated against observed discharges in 2001. The second and third implementations performed better than the first one, with  $NSE \approx 0.83$  and  $RRMSE \approx 34\%$ , when using model parameters averaged over 2002 and 2005 and the durations of antecedent precipitation of 10 and 20 days. The first implementation needed much longer duration of antecedent precipitation to achieve a similar performance. The calibration and validation proved that the retrievals of WSS and inundated area can reduce the required duration of antecedent precipitation, i.e. to achieve a satisfactory performance with fewer model parameters..”*

Page 16 Line 18 – Page 17 Line 2: We added a new subsection in the section of Discussion to compare the calibration and validation results of the three implementations. The comparison in calibration results shows that, the second and third implementation required a duration = 80 days to achieve a similar model performance as that of the first one using a duration = 150 days. The validation results show that the second and third implementation required a duration of 10 days to

achieve a similar model performance as the best one of the first implementation, which required a duration of 90 days. Thus we concluded that applying the WSS and inundated area can reduce the required duration of antecedent precipitation, without degradation of model performance.

**Q2:** The authors may need to pay more attentions to the overfitting issue. For example, the accuracy of their rainfall-runoff model is improved when they increase the longest durations of antecedent precipitation. The authors interpret this result in terms of the basin characteristics. However, when the authors increase the longest duration of antecedent precipitation, the performance of the linear regression should be always improved since the number of weights (unknown parameters) increases. It is possible that the improvement of the rainfall-runoff model in this paper is not related to any real rainfall-runoff processes. The authors may need to check if their statistical model does not overfit to the data using **the cross-validation analysis**. In the second and third implementation, the model has better performances with the larger durations of antecedent precipitation in the calibration period (Figures 6 and 7) while their performances are degraded by increasing the duration in the validation period (Figure 12). In the case of second and third models, increasing the duration causes the overfit to the data.

**Answers:** Thanks a lot for this useful comment about overfitting. We applied one of the cross validation methods, i.e. the leave-one-out method, for each implementation and included the results in the revised manuscript.

The results of cross validation for the three implementations were illustrated for each implementation in the section Results. They showed that: 1) when the duration  $\leq 100$  days, overfitting did not occur in all three implementation; 2) when the duration  $> 100$  days, overfitting would occur in the second and third implementation. It is clear that within this duration range, model performances of three implementations in the calibration period were improved by increasing the duration of antecedent precipitation, especially for the second and third implementations. The cross validation results proved that when the duration  $\leq 100$  days, these performance improvements were due to antecedent precipitations contributing to stream flow production, not simply related to the increment of model parameters. In the validation period, the model performance of the first implementation was slightly improved by increasing the duration, while the performances of the second and third implementations were overall degraded. This difference in the responses of model performance to the durations between the first and the other two implementations is probably due to that the weight range of antecedent precipitations in the first implementation is much smaller than that in the second and third implementations, i.e. the weights of the first implementation is not so sensitive to the durations as the second and third implementations. Accordingly, we modified the text as follows:

Page 1, Line 22 – 24, in the abstract: “*Cross validation results showed that when the duration  $\leq 100$  days, the model improvement from the first to the second and third implementations was due to the application of the WSS and inundated area.*”

Page 11, Line 10 – 16: We introduced one of the cross validation method, i.e. Leave-One-Out, in the section of Method.

Page 13, Line 18 – 26: The results of cross validation for the first implementation were illustrated in the section Results. They showed that the overfitting problem did not occur in the first implementation.

Page 14, Line 25 to Page 15, Line 2: The results of cross validation for the second implementation were illustrated in the section Results. They showed that the overfitting problem did not occur in the second implementation, when the duration  $\leq 100$  days.

Page 15, Line 33 to Page 16, Line 6: The results of cross validation for the third implementation were illustrated in the section Results. They showed that the overfitting problem did not occur in the third implementation, when the duration  $\leq 130$  days.

Page 18, Line 4 – 26: We summarized the evaluation of model performance as: “*In the calibration experiments, the performance of the three implementations improved when increasing the duration of antecedent precipitation, especially for the second and third implementation. Cross validation results proved that when the duration  $\leq 100$  days, these performance improvements were due to antecedent precipitations contributing to stream flow production, not due to the increment of the number of model parameters. In the validation experiments, the performance of the first implementation was slightly improved by increasing the duration, while the performances of the second and third implementation were overall degraded.*” It is explained that the difference in the responses of model performance to the durations between the first and the other two implementations was probably due to the weight range of all applied durations in the first implementation being much smaller than that in the second and third implementations, i.e. the weights of the first implementation are not so sensitive to the durations as the second and third implementations.

**Q3:** In addition, I believe that the number of unknown parameters also increases when the authors upgrade their model from the first to second implementation. The authors split the weight ( $w$ ) into two other parameters ( $\beta$ ). It is possible that the improvement of the model is just caused by increasing the number of parameters and overfitting to the data. In validation, the second and third implementations cannot significantly outperform the first implementation. It indicates the second and third implementations overfit to the calibration data. I am not convinced that the WSS and inundated area data positively impact to the model performance.

**Answers:** Thanks a lot for this useful comment. The overfitting problem has been discussed in the answers to Q2. Cross validation results clearly showed that when the duration  $\leq 100$  days, the model improvement from the first to the second and third implementations was due to the application of the WSS and inundated area, not due to the overfitting. In the calibration experiments, the performance of the three implementations improved when increasing the duration of antecedent precipitation, especially for the second and third implementation. Cross validation results also proved that when the duration  $\leq 100$  days, these performance improvements were due to antecedent precipitations contributing to stream flow production, not due to overfitting. In the validation experiments, the difference in the response of model performance to the durations between the first and the other two implementations is probably due to that the weight range of antecedent precipitations in the first implementation is much smaller than that in the second and third implementations, i.e. the weights of the first implementation is not so sensitive to the durations as the second and third implementations. Accordingly, we modified the text as follow:

Page 1, Line 22 – 24, in the abstract: “*Cross validation results showed that when the duration  $\leq 100$  days, the model improvement from the first to the second and third implementations was due to the application of the WSS and inundated area.*”

Page 16, Line 27 – 30: “*The results of cross validation for the three implementations showed that, when the duration  $\leq 100$  days, overfitting did not occur in all three implementations. Thus within this duration range, the model improvement from the first to the second and third implementation was due to the application of the retrievals of WSS and inundated area.*”

Page 18, Line 4 – 7: “*In the calibration experiments, model performances of three implementations were improved by increasing the duration of antecedent precipitation, especially for the second and third implementations. Cross validation results proved that when the duration  $\leq 100$  days, these performance improvements were due to antecedent precipitations contributing to stream flow production, not due to overfitting.*”

**Q4.** I recommend that the authors to use the model which let us to interpret the results more easily. The physically-based hydrological models is more appropriate since they do not need a heavy calibration.

**Answers:** Thanks for this comment. Our discrete rainfall-runoff model is also a physically-based hydrological model that takes the redistribution of antecedent precipitation and the water balance into account. The weights vector clearly shows the catchment response to precipitation, by producing fast and slow water flows and by the interaction between water flows and water storage. In fact, to determine which duration is appropriately applicable is also one of our goals. According to the best model performance in validation (Table 1), the duration of antecedent precipitation should be no more than 9 time steps for the first implementation, i.e. with the number of unknown parameters  $\leq 11$ , and less than 2 time steps for the second and third implementation, i.e. with the number of unknown parameters  $\leq 6$ . The number of unknown parameters that can be applied to the three implementations is similar to most hydrological models. It is, however, much easier to calibrate the parameters with a linear regression method in our model than in physical-based hydrological models.

Accordingly, we modified the text as follows:

Page 3, Line 9 – 32: We introduced hydrological processes that related to the redistribution of precipitation in time and explained the relationship between the discrete rainfall-runoff model and other hydrological models.

Page 4, Line 7 – 9: “*One advantage of the three implementations is that they can be calibrated with a linear regression method, much easier than the calibration methods used by other conceptual hydrological models, e.g. automatic calibration (Gupta et al., 1999; Madsen, 2000).*”

Page 5, Line 17 – 29 : We express the redistribution of precipitation in time as the summation of the water storages in soil and ground water due to antecedent precipitation at certain time interval and the cumulated evapotranspiration and water flow due to that precipitation in a certain period. This expression is used to derive and to explain the discrete rainfall-runoff model.

Page 17, Line 20 to Page 18, Line 2: We evaluated the model performance of the three implementations with the averaged parameters and chosen durations against stream flow in 2002 and 2005. The result shows that the first and second implementation with the mean parameters and a duration of 10 days (i.e. model parameters of 4) can be applied to different years.

Page 18, Line 27 – Page 19, Line 31: We explained the relationship between the weights in the first implementation and those in the second and third implementations. Through interpreting the weights evolution in the second and third implementations, we can understand the catchment response to precipitation better.

I give my specific comments below. I hope it helps.

Major Points:

**Q5** Page 2, Line 1: I think this Introduction can be significantly improved. In the Introduction section, the authors should explain their problem to be solved and review the previous works related to this problem. Then, the authors should briefly summarize the increment of their research from the previous papers. The author should not include the detail of their methodology in the Introduction section. I believe that in this paper, both previous contributions and the author's original approaches coexist in the Introduction. I recommend the authors to move some paragraphs to the following method section and focus more on citing the previous works and explaining the context of their problem.

**Answers:** Thanks a lot for this useful comment. The paragraphs about the water balance equation and the discrete rainfall-runoff were all moved to the section 2.1. We improved the Introduction by adding a review of previous studies about the application of microwave observations in the hydrological models, so that to describe the advantages of applying the retrievals of WSS and inundated area to the discrete rainfall-runoff model. Accordingly, we modified the Introduction as follows:

Page 2, Line 6-18: We reviewed some hydrological models which used water saturated soil to describe the catchment response to precipitation. We clarified that the problem is to apply the retrieval of WSS and inundated area into conceptual hydrological models.

Page 2, Line 19 – Page 3, Line 5: We reviewed previous studies to solve this problem, e.g. by using microwave observation to assimilate regional soil moisture using a soil-canopy-atmosphere model. This method complicates the application of conceptual hydrological models. We also reviewed previous studies on the direct but lagged relationships between the WSS and inundated area and the river discharge or stage. These studies suggested that there may be a simple method to apply the retrieval of WSS and inundated area into conceptual models.

Page 3, Line 6– Line 23: The physical meaning of the WSS and inundated area was introduced and physical basis of the discrete rainfall runoff model, i.e. the redistribution of precipitation in time, was explained.

Page 3, Line 24 – Page 4, Line 3: We explained how other hydrological models, i.e. physical or statistical approaches, consider the redistribution of precipitation in time. From them, we introduced our discrete rainfall-runoff model.

Page 4, Line 4 – 18: The advantages of our model were explained: 1) the convenience in model calibration and fewer input data; 2) most of the input data can be derived from satellite observations.

Page 4, Line 19 – 35: The objective and the overall structure of the paper were introduced.

**Q6:** Page 5, Line 1-23: The authors may move these paragraphs to the following method section.

**Answers:** Thanks a lot for this comment. We moved it to a new section, i.e. section 2.1 from Page 5, Line 16 to Page 6, Line 26.

**Q7:** Page 5, Line 17-22: I think the upgrade from the second to third implementation is not related to the use of satellite WSS data. Is the third implementation really needed considering the scope of this paper?

**Answers:** Thanks a lot for this comment. The third implementation is more complicated than the second one, due to the introduction of the potential subsurface flow. In this paper three implementations were developed by increasing complexity and evaluated separately. Our aim was to understand whether higher model complexity would yield higher accuracy. In both calibration and validation experiments, the implementation that has shorter duration and better performance is preferred. The calculation of the potential subsurface flow needs the WSS and inundate data, thus the upgrade from the second to third implementation is also related to the WSS data. Accordingly, we modified the text as follow:

Page 1, Line 12 – 14: *“The model was implemented at three levels of increasing complexity with precipitation, ground water table depth, and the WSS and inundated area, which are designed to reduce the duration required to achieve a reasonable performance.”*

Page 10, Line 2 – 6 : *“The second implementation partitions precipitation into overland and infiltrated flows. A fraction of the infiltrated flow will result in subsurface flow when it encounters deeper saturated soil layers (Beven and Kirkby, 1979; Beven et al., 1984; Sivapalan et al., 1987), which we define as potential subsurface flow. The remaining infiltrated flow is stored into soil and ground water and mainly used for evapotranspiration. Model performance might be improved by specifying the potential subsurface flow, instead of the infiltrated flow in the second implementation.”*

**Q8:** Page 6, Line 16: Is the “discrete” hydrological model identical to the “distributed” hydrological model? Please define the “discrete” model.

**Answers:** Thanks a lot for this comment. The discrete rainfall-runoff model is different from the distributed model. Distributed models mean stream flow calculated applying spatially distributed parameters / model structure, while, the discrete model means stream flow is redistributed in time. The term “discrete” means the contributions of precipitation in a certain period can be assigned to time intervals of several days or longer. Accordingly, we modified the text as follow:

Page 4, Line 21 – 23: *“The discrete rainfall-runoff model is a lumped hydrological model, developed according to the water balance equation and the redistribution of antecedent precipitation in time. The term “discrete” means the contributions of precipitation in a certain period can be assigned to time intervals of several days or longer.”*

**Q9:** Page 10, Line 16-17: Although I understand that their WSS product has been explained in Shang et al. [2014], please explain some details of their product in this paper. I believe that this information is helpful to interpret the author’s results. For example, how is the spatial resolution of the product? Is their 37GHz product affected by atmospheric vapor and clouds? In addition, there are some contradictions between precipitation time series and WSS time series in Figure 2 (e.g., sharp peak in DOY 250 of 2005 with no strong rainfall). Please clarify the capability and limitation of their data.

**Answers:** Thanks a lot for this useful comment. We added the explanations about the retrieval of WSS and inundated area, data product and its spatial resolution in the section Data and Study area. The atmospheric influence on PDBT at 37 GHz has been removed with a time series method as described in Shang, H., Jia, J., and Menenti, M.: Modeling and Reconstruction of Time Series of Passive Microwave Data by Discrete Fourier Transform Guided Filtering and Harmonic Analysis, Remote Sensing, 08, 970, 2016. We checked the raw data and found that there are several pixels that miss the brightness temperature at horizontal polarization probably due to the heating problem in the radiometer (this kind of missing data also occurs at other scan paths). Thus we remove these errors and plot new images in Fig. 2. We modified the text as follow:

Page 11, Line 30 to Page 12, Line 12: the information about the retrieval of WSS and inundated area was added, including the removal of atmospheric influence. The information about the SSM/I radiometer, the used gridded brightness temperature data and its spatial resolution was now introduced.

Figure 2: We plotted the new images by removing the error observations of SSM/I in 2005.

**Q10:** Page 11, Line 29-30: As described above, whenever the number of weights (unknown parameters) increases, the linear regression performs better. I believe that the improvement should not be interpreted in this way unless the authors confirm that the improvement is robust by the calibration/validation process shown in section 4.5.

**Answers:** Thanks a lot for this useful comment. The analysis about the number of weights has been explained in the answers to Q 3 in detail. In the revised manuscript, we illustrated the results in the calibration and validation experiments for each implementation at first and obtained the conclusions later on in the section Discussion. According, we modified the text as follow:

We moved the corresponding text to the section Discussion at Page 20, Line 6 – 8: *“When the duration was increased from 1 to 15 time steps, the reduction in RRMSE was significant in the autumn and winter for the first implementation (Fig. 19a). This indicates that the stream flow after June is influenced by longer antecedent precipitations, for example those in the spring.”*

**Q11:** Page 12, Line 23-24: It is hard for me to interpret the negative weights. The authors explain that negative weights are caused by water storages. It looks a reasonable explanation. However, if the weights are negative, streamflow decreases with increase of rainfall. How can the authors explain this? Do the authors have some previous works which interpret the negative weights in this way? Do the authors think they need a constraint of non-negative weights in the linear regression?

**Answers:** Thanks a lot for this useful comment. The negative weights in the discrete rainfall-runoff models indicate the interaction between stream flow and various water storages. This interaction has not been accounted for in current hydrological models, but has been studied in many cases. For example, a large fraction of precipitation will be stored in a flooded forest plain and be released regularly later on in Amazon, according to Vörösmarty, et al.: Analyzing the discharge regime of a large tropical river through remote sensing, ground-based climatic data, and modeling. In the hydrogeology, there are many studies about the interaction between surface water and ground water, e.g. Sophocleous, M.: Interactions between groundwater and surface water: the state of the science. Thus a negative value is a very important factor in our model, and the model does not need a constraint on non-negative weights. Stream flow integrates the volumetric contribution of precipitation in a certain period, i.e. combines a set of positive and negative weights. So we could not interpret the relationship between stream flow and precipitation from the negative weight at one time step, without taking the contributions of

precipitation at other time steps in account. That is why we interpret the weight set from the point of view of the evolution in the contributions of precipitation to stream flow at various time steps. Accordingly, we modified the text as follows:

Page 3, Line 11 – 23, in the Introduction section: We introduced the interaction between stream flow and various water storages in a catchment at first.

Page 5, Line 17 – Page 6, Line 26: We formulated the redistribution of precipitation in time and explained the meaning of weights in the discrete rainfall-runoff model.

Page 8, Line 5 – 27: We formulated the redistribution of component flows in time and explained the meaning of weights in the discrete rainfall-runoff model.

Page 10, Line 9 – 17: We explained that the weights of the discrete rainfall-runoff indicate the evolution in contributes of precipitation and component flow to stream flow at various time steps. The relationship between stream flow at one time step and precipitation in a certain period varies at different time steps, thus it is difficult to analyze the pattern of this relationship.

Page 18, Line 27 to Page 19, Line 31: We interpret the weights of three implementations.

**Q13:** Page 13, Line 6-7: Please explain this statistical assessment more deeply and discuss the model performance quantitatively.

**Answers:** Thanks a lot for this comment. We illustrated the changes in model performance with increasing the durations from 1 to 10 time steps, and assessed the model performance at the 10 time step. The model performances of duration > 10 time steps are not reliable, due to the overfitting problem. Accordingly, we modified the text as follows:

Page 14, Line 17 - 24: *“When the duration of antecedent precipitation was increased from 1 to 10 time steps, NSE increased from 0.78 to 0.94 in 2002, from 0.83 to 0.96 in 2005, and RRMSE decreased from 34 % to 17 % in 2002, from 27 % to 12 % in 2005. The improvement rate was also different in different stages: in 2002, large improvements occurred when the duration was increased from 1 to 3 time steps and from 7 to 9 time steps; in 2005, large improvements occurred when the duration was increased from 2 to 3 time steps and from 6 to 9 time steps. When the duration = 10 time steps, the model performance of the second implementation was better than 90% of current hydrological models in the calibration process, according to the statistical assessment by (Ritter and Muñoz-Carpena, 2013). The performance improvement when the duration > 10 time steps is probably due to the overfitting, as shown below.*

Page 14, Line 26 to Page 15, Line 2: We illustrated the results of cross-validation. The results showed that overfitting occurred in the second implementation when the duration > 10 time steps.

**Q14:** Page 13, Line 8-11: As discussed above, this interpretation is valid after the authors perform the validation and confirm the robustness of the model’s improvement.

**Answers:** Thanks for this comment. We removed this interpretation in the revised paper.

**Q15:** Page 13, Line 32-35: At the third time step, the weight goes negative. However, it goes positive at the sixth time step again. How can the authors interpret it?

**Answers:** Thanks a lot for this comment. The negative weight in component flows mean that a fraction of this component flow is stored in various hydrological elements, e.g. lake and soil, or is used by human activities, e.g. irrigation. The stored and consumed water will be released in a lagged period, e.g. at the sixth time step. Our modifications are summarized in the answers to Q11.

**Q16:** Page 14, Line 32-33: I believe that the authors also use ground water table depth data in the first implementation. Please clarify it.

**Answers:** Thanks a lot for this useful comment. We modified the text as follow:

Page 20, Line 1 – 16: The text about the recharge period of ground water was moved to the section 5.4. The analysis of RRMSE in the first implementation was added to identify the recharge period of ground water in the study area

**Q17:** Page 15, Line 6: As discussed, I strongly believe that this model validation rejects the authors' hypothesis.

**Answers:** Thanks a lot for this comment. Please see the answers to Q1.

Minor Points:

**Q18:** Page 7, Line 13 - Page 8 Line20: I recommend the authors to move this sub-sections to the end of this method section. The model calibration and the evaluation metrics are applied to all of implementations so that it should not be included only in the first implementation section.

**Answers:** Thanks a lot for this comment. We modified the text as follow:

From Page 9, Line 23 to Page 10, Line 28: Model calibration and metrics to evaluate performance were explained in the new section 2.5

**Q19:** Page 9, Line 32: Beta 21 and Beta22 may be Beta31 and Beta32, respectively.

**Answers:** Thanks a lot for this useful comment. They have been modified in Page 9, Line 17

**Q20:** Page 10, Line 15: Please show the locations of 9 wells in the river basin.

**Answers:** Thanks a lot for this comment. Due to the data policy in China, the detailed location of the 9 wells is not accessible for us. Alternatively, we show the location of the pixel in Fig. 1 that includes the 9 wells. Accordingly, we modified the text as follow:

Page 11, Line 25 : *“We used observations of the ground water table depth at 9 wells in Changsha (at the pixel numbered 1 in Fig. 1).”*

**Q21:**Page 14, Line 5: rives!rivers

**Answers:** Thanks a lot for this comment. This sentence was removed in the revised paper.