

Responses to comments

We, the authors of the manuscript, appreciate the valuable and constructive comments from the Referee, D. Klotz. We will thoroughly revise the manuscript according to these comments. The detailed responses to the comments and questions are as follows.

General comments

The present study proposes a bootstrap based quantification of the spatial rainfall uncertainty for rainfall-runoff modelling. The main body of the work uses aforementioned bootstrap method is used to select between different numbers of stations and different groups of stations. The influence of the number and grouping is then quantified via the Nash Sutcliffe Efficiency and exponential fit is used to derive a heuristic for the estimation of the optimal densities of rainfall measurement stations. The whole analysis is based on three manually selected rainfall events from 2001, 2002 and 2006, with data from of the Qingjian basin. Additionally, rainfall data from a second basin (Longxi) is used to provide a basic comparison about the scale and uniqueness of the conclusions.

The topic as such is certainly relevant for the study of the rainfall-runoff relationship as such and would fit a HESS. I do also conceive that the endeavor of the authors poses major challenges. Rainfall process can be extremely heterogeneous and as the authors state a quantification of the uncertainties depends multiple interplaying phenomena (such as the position of the meteorological stations, the unique futures of the studied events or the elevation profiles of a catchment). So, the authors did probably a good job, if we consider the possible difficulties that are inherent in such a study. I am nevertheless left with many concerns. The most important wants are listed in the following points, minor comments and questions are given later.

[Answer] We would like to thank the referee for giving us valuable and constructive comments, which have encouraged us to view our work with much greater insight than before. We will thoroughly revise the manuscript based on these comments. We hope the revision will improve the completeness and accuracy of the results. Please see our responses to all the comments in the following for details.

1. First and foremost, I believe that considerable attention should be given to improve the introduction in special and the readability of the manuscript in general. The whole manuscript needs a clearer structure, the aims must be defined in a clearer fashion (maybe less focus on different aspects and more focus on the main topics) and the manuscript as a whole should meander less between topics. Arguments about the computational expense, for example, can be found throughout the script, but no chapter is devoted to it.

[Answer] We appreciate the referee's constructive comment. In the revision, we will improve the introduction part through citing more related papers and deleting the

irrelevant ones. We will also reorganize the text with a clearer structure and further highlight the main topics of this manuscript. As suggested by the referee, we will add some sentences to discuss the computational expense of this method in the revised manuscript.

2. In reference to my last point, I have to say that I truly believe that it is worth to discuss the computational expense of the method as such. Currently hints about the computational efficacy are spread throughout the manuscript, but I miss a dedicated chapter and quantitative arguments. The method seems quite computationally expensive as the hydrological model needs to be executed approximately $N \times K$ times. Now what I would like to know is what it does mean quantitatively. How long does a run of the used r-r model take, and how long does an execution of the bootstrap need. Especially the last point provides a plethora of possible topics/experiments: Influence of shorter and longer time series, usage of the bootstrap without and with parallelization, usage of the model with and without parallelization, on a cluster or a normal pc, etc.

[Answer] The selected upper Qingjian River basin has a moderate size. The number of hillslope-channel unit in this river basin is about 8000, and a serial simulation using the DYRIM costs about 5 minutes. The selected basin is only for the demonstration of the proposed methods in this manuscript. Further application of the methods to larger river basins and to calibrate more parameters will of course cost more computational time, and the proposed methods will keep effective.

The model simulation was implemented on the Microsoft Windows Azure cloud computing platform, allocated with a total number of 80 processor cores. With the Windows Azure configuration, the simulation time of the DYRIM with 4 processor cores for one event in the upper Qingjian River basin was 130 s.

3. The actual time-resolution of the modelling needs some further discussion and explanations. I think the main problem here is that the descriptions are not specific enough. The authors say that “most of the data” was at a resolution of 2 hours and of 6 minutes respectively. It is difficult for readers to infer what that means exactly. A table that indicates which stations have fine/coarse resolutions and sparse data would be very helpful. In current manuscript also forces readers infer the magnitude of the distortion that is induced by the different resolutions. How exactly is the coarse scaled data disaggregated? How does the comparison work between aggregated (hence smoothed) fine scale data and disaggregated data? For many readers (including myself) it will seem strange that a 6 minutes interval for simulating the runoff data is used. The authors claim that the reason for that is to emphasize the main hydrological process. They do however not explain how works if the input data has a resolution of two hours only.

[Answer] Thanks for the useful comment. Actually, the time step used in the DYRIM is determined by the temporal scale of the observed runoff data. For the event-based hydrological simulation in this study, the minimum time interval of the observed runoff data can be only 6 minutes, and therefore, we have used 6-min time step rather

than the larger time step. We have to select the simulated discharge time nearest to the observed runoff time for the NSE evaluation in order to reduce the time error. In that case, most of the observed runoff times can be matched by the simulated discharge times; only a very few observed runoff times should be treated as above.

Actually, disaggregating sparser rainfall data into smaller time steps (i.e., 2 hours) will have little effect on the precision of hydrological simulation. In contrast, aggregating fine time step rainfall data into larger time steps will affect the precision of hydrological simulation since short-duration and high-intensity rains may be homogenised. However, this study focuses on the spatial uncertainty of rainfall rather than the temporal uncertainty, and thus, the measured rainfall data should be pre-treated to have the same time step (i.e., 2 hours).

We fully agree with the referee that this study will be stronger if the input data can fit the desired framework. However, unfortunately, we did not have the measured rainfall data with finer time steps during the preparation of the original manuscript. Most (up to 95%) of the time steps of the measured rainfall data are 2 hours, and the measured rainfall data with finer time steps are archived by the local hydrology bureau with some confidentiality. This is the main reason why we have to adjust the time steps of all the measured rainfall data to 2 hours. Now, we have obtained high-temporal-resolution rainfall data for some of the stations, and we will try to include those new data to get better simulation results in the revised manuscript.

4. The used method need further underpinning. Why are Thiessen Polygons used and not a more sophisticated method (e.g. Kriging)? Especially when the authors conceive that the method can cause problems (Page 9, Lines 14-16). I have similar concerns regarding the use of the genetic algorithm. Why are not more modern and conventional techniques used. In the domain of rainfall-runoff modelling Shuffled Complex Evolution and Dynamically Dimensioned Search come to mind. In general, there also exist more modern versions of genetic/evolutionary algorithms.

[Answer] We appreciate the referee's comment. The calculation of the average basin rainfall needs to be performed hundreds of times in the bootstrap method so that the performability should be regarded as the first consideration. Thiessen polygon method can be easily achieved for batch loops with the ArcGIS, and it is worth noting that Thiessen polygon method is only adopted as one representative way to estimate the average basin rainfall, which can also be realized by using other methods.

In this manuscript, GA is selected as the example optimization algorithm that employed in the double-layer parallelization (Zhang et al., 2016), mainly because GA is simple and widely-accepted. For hydrological model optimization, SCE-UA and DDS are considered better than GA, and the double-layer parallelization plus SCE-UA (or DDS) will be better than plus GA. However, the two layers of parallelization are independent from each other, and then the upper layer is capable of incorporating other optimization algorithms, including the SCE-UA (or DDS). Therefore, the adoption of GA will not largely influence the contribution of this manuscript on hydrological uncertainty.

5. Finally, I would be ready to be convinced otherwise, but I am not sure if the NSE is a well-chosen objective criterion for this sort of study. The low NSE values (that results, according to the authors, from the high resolution) make it difficult to use conventional intuition about the objective measure. Additionally, the chosen events differ largely in from, magnitude, runoff-coefficient and error structure. The amount of evaluated data with low runoffs (before and after the events) also seems to vary. Nonetheless, I believe that in this particular use of the NSE is suboptimal, since information about magnitude of the events and errors are hidden by its bounded. For this particular case I would therefore propose to use the MSE instead. This would directly express the goodness of fit and the improvements due to the respective calibrations can be seen as relative improvements. Also no additional explanations would be needed to report the jumps of improvements reported at page 9, line 25. Alternatively, it would be useful to provide a sensitivity measure with regard to the individual results.

[Answer] The Nash-Sutcliffe coefficient of efficiency (NSE) has been used as one of the very popular indices to assess the performance of hydrological models, especially for the event-based simulation. The peak time and peak value can be well evaluated by using the NSE. However, we may also use the MSE as another objective criterion and compare the results against those obtained by using the NSE. Moreover, the three events with different rainfall-runoff responses were selected in this study because they are partially caused by different spatial rainfall patterns, and they cannot be accurately simulated by using the same set of parameters. That is why we have tried to calibrate the parameters for these three events independently.

6. Figure 4 needs more explanations and better quality. For some reasons the observation-dots seem not to be equidistant! It is not explained why the model reacts faster than the real system (at least for event 2001 and 2006). At least the strange fit for the event of 2001 seems to be related to this issue, since the “independently calibrated” hydrograph might only be so low because of the large errors at the raising limb. The current resolution of the figure is not good and the depiction is difficult to read. For me it was, for example, not possible to discern what the cluster of observations means at the start of the raising limb of the 2002 event.

[Answer] Thanks for the comment. In this study, the minimum time interval of the observed runoff data is only 6 minutes during the flood, while the time intervals are much larger before or after the flood. That is why the observation-dots are not equidistant.

For the precision of the simulation results, with the scale ranging from monthly to daily, former studies of the DYRIM have achieved pretty good results and the NSE could easily reach a satisfactory value, for example, higher than 0.85 (Wang et al., 2015). What we are endeavoring to do in this manuscript is to achieve the simulation of flood events with an hourly scale. In the DYRIM, the temporal resolution of runoff simulation is 6 minutes, but each rainfall data point with a 2-hour step is uniformly assigned to corresponding simulation time steps. The time step of rainfall may have a considerable impact on the results, resulting in the flatter shape of simulated flow

discharges processes (Fig. 4). To better simulate such short-duration high-intensity rainfall-runoff events, rainfall observations with higher spatial and temporal resolutions are badly needed. When finer rainfall records are obtained from concurrent meteorological stations and satellite remote sensing, e.g., with a time step as fine as 1 minute and a spatial resolution at 8 km, the time step of hydrological models will be even shorter than 6 minutes, and the hillslope-channel units will be smaller than 0.1 km². In such a situation, the computational consumption of hydrological model calibration will increase dramatically, and the proposed double layer parallelism will be one of the necessities.

Specific comments

S1) This might be nitpicky, but the authors state that the other methods for spatial rainfall uncertainty quantification are not applicable to other basins. Why is that and why does the proposed method not exhibit this problem? The test is only conducted for one basin.

[Answer] Former studies mainly focused on the certain river basins or rainfall events with specific method such as the conditional simulation, and this may lead to different results of spatial rainfall variability, such as the rules proposed by Villarini et al. (2008) (e.g., more than 15 stations over an area of approximately 135 km² were necessary to estimate the true areal rainfall at a three-hour scale with an error less than 20%).

To the best of our knowledge, this paper is an innovative work to propose such a bootstrap method for uncertainty estimation. This contribution primarily aims at the construction and realization of the method architecture, rather than an uncertainty case study. Furthermore, the hydrological model simulation and the bootstrap method proposed in this paper are independent from each other. The capability of the proposed system may incorporate different hydrological models and other techniques beyond the GA and Thiessen Polygon. This is also an important advantage of the proposed approach.

S2) Bootstrapping is a very, very robust method, so the following concern might be of less importance: As far as I know bootstrapping does (still) assume independence of samples. Is this given for rainfall stations? The description of the cluster analysis seems to suggest otherwise. What influence does this have?

[Answer] Data series from any rainfall station are recorded independently at the point where the rainfall station locates, and data from all rainfall stations are indispensable for providing areal rainfall estimates. What we have done for cluster analysis is positive that a certain rainfall in an area has the inherent connection because of the weather condition and physical climate effects. What we want to do is to achieve the more accurate simulation results and complete understanding of the rainfall spatial uncertainty by using the cluster analysis and the bootstrap method based on the available data.

S3) For all Figures: Please always mention the used basin explicitly in the figure

captions. I understand that it is somewhat redundant because the evaluation takes only place with regard to Qingjian river basins. But, it is very useful for readers who want to get an overview.

[Answer] We will revise this in the revision. Thanks.

S4) Page 2, Line 9f: I think rules should be in plural here. Villarini et al. (2008) use a catchment of the size of 125km², but seem to compare the gauging stations with satellite images with a resolution of 200km². I therefore believe that the “rules of thumb” should approximate the size of the satellite pixel. I am not sure however, the argument can get finicky here. Lastly, I think one should mention that the study takes place in England.

[Answer] We will revise this in the revision. Thanks.

S.5) Page 6, Line 28: What does “adopted” mean in this context? What was changed from the original one. Why is a genetic algorithm used and not a more prevalent method such as the shuffled complex evolution? Why is not a newer evolutionary/genetic algorithm?

[Answer] We are sorry for the misleading in this part. The genetic algorithm (GA) was firstly proposed by Holland (1975), but the GA implementation using binary and real coded variables with the newer C++ version that we adopted in this study was developed by Deb (1997). This implementation employed in this paper treats the parameters to be optimized as real numbers with simulated binary crossover and real-parameter mutation. This technique promises the hydrological model parameters independent of the GA. We will further revise this part in the revision.

For the reason of choosing the GA rather than other methods, please refer to our response to **Point 4** of the General comments for details.

S.6) Page 6, Line 8: What does it mean that the “influence of topography on rainfall is negligible”?

[Answer] As is known, the topography has impacts on the rainfall, particularly for the mountainous regions. It should be taken into account in the hydrological simulation. However, this kind of influence is not a major problem in this study and has not been considered. Nevertheless, we will remove this statement in the revision.

S.7) Page 7, Line 1-2: Please specify the “stop criterion” explicitly (number of generations, NSE, ?) How many generations did it take on average?

[Answer] During the fitness evaluation of the GA, a new generation of the model parameter is proposed to explore more parameter combinations until the stop criterion or the maximum number of generations is reached. We want to explore the fluctuation range of the NSE in uncertainty estimation, so we set the maximum number of generations (e.g., 10 generations in this study) as the stop criterion.

S8) Page 9, Line 9: What kind of independent calibration are we talking about here? Why are parameter interactions not seen as a problem in doing this? Why is only the

K_{ZUS} optimized and not the most sensitive parameter group (e.g. all soil hydraulic conductivity parameters)?

[Answer] There are two kinds of schemes for the model calibration in this study. One is the independent calibration on each rainfall-runoff event using different sets of parameter combinations, and the other is the calibration on all the events sharing the same set of parameter combination. When the model is calibrated on the events independently one by one, more precise results would be obtained to show the potential performance of the model. However, in such a scheme, the optimized parameters will adapt to each described distribution of rainfall by the stations. Therefore, to insulate the effect of rainfall spatial distribution from model parameters, at least in an average sense, the three events should be considered comprehensively through using the same set of parameter combination.

S9) Page 9, Lines 25-26: I would disagree with the statement that the individually calibrated model runs prove that DYRIM is able to represent the rainfall-runoff events in a sufficient way, as long as no evidence is provided that the results are not just due to overfitting (Maybe evaluate the individually derived parameters for the other events to?). I would propose to see the NSE values as hints (if anything). One might also be able to argue that the generated hydrographs can be seen as some sorts are upper boundaries or best case scenarios for the DYRIM simulations. Additionally, would it not be possible to use this information to determine the possible upper bound for the relationship between NSE and measurement side density (Equation 2 and Figure 9)?

[Answer] We have to recognize that the model performances are not so good for certain events and that is why we have conducted the independent calibration on each rainfall-runoff event using different sets of parameter combinations. You may regard the latter results as the upper boundaries or the best scenarios for the simulations of these events. However, we are sorry to say that this information has not been used to determine the possible upper bounds for the relationships presented in Equation 2 and Figure 9 yet. We may conduct further study on this in our future work.

S10) Page 11, Lines 1-11: In my opinion this should be part of the method section and not of the results.

[Answer] We will add a new subsection (i.e., 3.3 Uncertainty quantization of basin rainfall and simulated runoff) and move this part to that subsection. Thanks.

S11) Page 13, Lines 29-31. Here it is argued that the large errors of the hydrological simulations of the 2006 rainfall event are most likely due the structural and parameter deficiency. Is it possible to plot the cumulated rainfall alongside the cumulate measured runoff for this event (or for all of them)? On basis of the low runoff coefficient and the hydrographs I would (perhaps naively) assume that it there is a bias in the input or the runoff measurements.

[Answer] Thank for the insightful suggestion. A new figure comparing the cumulated rainfall and measured runoff will be plotted in the revised manuscript.

Minor Remarks

Page 1, Line 25: Can you provide additional sources here? Beven, 2001 is a large tome, while Cibin 2014 seems to be focused on ungauged basins (which are not even mentioned as application example).

[Answer] We will cite more papers (e.g., Singh and Prevert, 2002; Fares et al., 2014) in the revision. Thanks.

Page 1, Line 25-29: For me the first sentence of this passage is intertwined. Could you divide the sentence into two? One point is that hydrologist try to improve the accuracy of simulations and predictions and another point is how this can be achieved (i.e. improving the model structure, better calibration method, better measurements of input data). From there on I think it would be worth to add an additional sentence, arguing why the precipitation is seen important factor for improving simulations and forecasts.

[Answer] In the revision, we will divide this sentence into two and also add an additional sentence to argue why precipitation is seen important factor for improving simulations and forecasts. Thanks.

Page 2, Line 1: Please rethink this sentence. The comparison between radar and rainfall station is strangely formulated.

[Answer] We will revise this in the revision. Thanks.

Page 2, Lines 11-13: Please provide additional sources. Four studies are not numerous.

[Answer] We will provide more details and cite more papers (e.g., Bedient et al., 2000; Smith et al., 2004; Casper et al., 2009; He et al., 2013) in the revision. Thanks.

Page2, Line 14: Please recalculate the units from square-miles to square kilometers and use that unit consistently throughout the manuscript.

[Answer] We will revise this in the revision. Thanks.

Page 2, Line 18f. You described what Moulin et al. (2009), but not their results/conclusions. Readers will wonder why not, as they are provided for the previously mentioned studies. Could you expand on that?

[Answer] We will provide more details about the study of Moulin et al. (2009) in the revision. Thanks.

Page 3, Lines 19-20: I do not understand what the citations are referring to. Are all these authors proving the advantages of simplicity and high-accuracy?

[Answer] The citations here are referred to the analysis or application of the bootstrap method in the hydrological model, and several of them have proved the advantages (e.g., simplicity and high-accuracy). We will further revise this sentence to make it

clearer. Thanks.

Page 3, Lines 7: Can somehow you remove the double and in the sentence (maybe use “as well as”? It would improve the readability of the sentence

[Answer] We will use “as well as” instead of “and” in the revision. Thanks.

Page 4, Lines 8-9: I find that argument a bit difficult, in my eyes the bootstrap is a generalization of the jackknife. Maybe compare it to another resampling technique?

[Answer] The jackknife is thought of as a linear expansion method for approximating the bootstrap. Bootstrap methods are more widely applicable than the jackknife and also more dependable. We will further revise this sentence to make it clearer. Thanks.

Page 4, Line 13: Is it wise to put this argument forward like this. As far as I know, bootstrapping assumes independence of samples. That is not necessarily a property of the population per se, but the sentence could lead to misunderstandings.

[Answer] We will remove this sentence in the revision. Thanks.

Page 4, Line 19-22. You might want to split the sentence to improve readability.

[Answer] We will revise this sentence in the revision. Thanks.

Page 5, Line 13: Change to “are obtained”.

[Answer] We will revise this in the revision. Thanks.

Page 5, Line 16: Can you make the following statement more explicit: “..., the bootstrap method is used to traverse most of the combinations of rainfall stations, ...” (emphasis is my own). What does most mean?

[Answer] In Section 2.1, we have explained the repetition of runoff simulation and estimation for the average basin rainfall by using the bootstrap method, respectively. *“To reduce the repetition of simulations, for a certain combination, a certain number of stations in a certain group are randomly selected only once.”* The loop traversed all the combinations of station groups in rainfall estimation, but a few combinations in runoff simulation were skipped based on the station classification.

Page 5, Line 30f: What exactly does “unit” mean in this context?

[Answer] The “unit” means the basic hillslope-channel unit. This unit is an independent rainfall runoff module since the DYRIM is a penalization hydrological model. We will further revise this sentence to make it clearer. Thanks.

Page 6, Line 5: Remove “Then”.

[Answer] We will remove “Then” in the revision. Thanks.

Page 6, Line 8: Remove “End”.

[Answer] This sentence will be removed in the revision. Thanks.

Page 6, Line 31-31 I do not understand the sentence “This technique promises the parameters independent of the GA and easy to be optimized”

[Answer] The parameters from the hydrological model are independent of the GA, which means these parameter combinations are input and output of the optimization programs. We will further revise this sentence to make it clearer. Thanks.

Page 8, Line 1: Write “One rainfall event, which occurred on ...”.

[Answer] We will revise this in the revision. Thanks.

Page 9, Lines 11-12: I do not understand this sentence.

[Answer] The main reason of poor calibration results is the hourly resolution of runoff results using rainfall input with lower temporal resolution. Better precision of simulation can be found in daily to monthly results, with satisfactory NSEs higher than 0.8 (Shi et al., 2016; Wang et al., 2015). In order to demonstrate the potential performance of the DYRIM, the topsoil vertical saturated conductivity K_{zus} was calibrated on the three events independently one by one (as explained in the following paragraph).

Page 9, Lines 32-24: This statement needs a citation.

[Answer] We will cite one paper (Huang et al., 2016) in the revision. Thanks.

Page 10, Lines 3-4: Shouldn't this be part of the results and discussion?

[Answer] We will move this sentence to the Section 5. Thanks.

Page 10, Lines 27-31.: Please reformulate.

[Answer] We will revise this in the revision. Thanks.

Page 11, Lines 27-28: Sentence is unclear.

[Answer] We will provide some explanation in the revision. Thanks.

Page 12, Line 10: Please reformulate.

[Answer] We will revise this in the revision. Thanks.

Page 13, Lines 3-4: Write “that even if”

[Answer] We will revise this in the revision. Thanks.

Figure 2: Please rework the plot. The legends are hard to read. It is difficult to grasp the extends of the basins.

[Answer] The legends of this figure will be reworked. Thanks.

Figure 9: Why is the “prediction curve with upper limit” in plot c, higher than the prediction curve without upper limits. The former appears to be fitted for less data. Also, is it possible to provide uncertainty bounds?

[Answer] In the third case, the NSE value was low, and the higher D_5 values (i.e., 6 to

11) were selected for curve fitting in Figure 9c (the dash lines). We found that the low NSE value of the third case might be caused by the hydrological model structure and parameters rather than the spatial uncertainty of rainfall records. However, in terms of the other two cases, the proposed fitting of Equation (2) would be capable to reveal the possible improvement of simulation. Moreover, the uncertainty analysis for rainfall input combined with model structure and parameters could provide more complete and accurate uncertainty bounds.

References:

- Bedient, P.B., Hoblit, B.C., Gladwell, D.C., Vieux B.E., 2000. NEXRAD radar for flood prediction in Houston. *Journal of Hydrologic Engineering*, 5(3), 269-277.
- Casper, M.C., Herbst, M., Grundmann, J., Buchholz, O., Bliefernicht, J., 2009. Influence of rainfall variability on the simulation of extreme runoff in small catchments. *Hydrologie Und Wasserbewirtschaftung*, 53(3), 134-139.
- Deb, K., 1997. Genetic algorithm in search and optimization: the technique and applications. *Proc.of Int.workshop on Soft Computing & Intelligent Systems*, 58-87.
- Fares, A., Awal, R., Michaud, J., Chu, P.S., Fares, S., Kodama, K., Rosener, M., 2014. Rainfall-runoff modeling in a flashy tropical watershed using the distributed HL-RDHM model. *Journal of Hydrology*, 519, 3436-3447.
- He, X., Sonnenborg, T.O., Refsgaard, J.C., Vejen, F., Jensen, K.H., 2013. Evaluation of the value of radar QPE data and rain gauge data for hydrological modeling. *Water Resources Research*, 49(9), 5989-6005.
- Huang, P.N., Li, Z.J., Chen, J., Li, Q.L., Yao, C., 2016. Event-based hydrological modeling for detecting dominant hydrological process and suitable model strategy for semi-arid catchments. *Journal of Hydrology*, 542, 292-303.
- Shi, H.Y., Li, T.J., Wang, K., Zhang, A., Wang, G.Q., Fu, X.D., 2016. Physically-based simulation of the streamflow decrease caused by sediment-trapping dams in the middle Yellow River. *Hydrological Processes*, 30(5), 783-794.
- Singh, V.P., Prevert, D.K., 2002. *Mathematical models of large watershed hydrology*. Water Resources: Highlands Ranch, Colo., 891.
- Smith, M.B., Koren, V.I., Zhang, Z., Reed, S.M., Pan, J.J., Moreda, F., 2004. Runoff response to spatial variability in precipitation: an analysis of observed data. *Journal of Hydrology*, 298(1), 267-286.
- Wang, G.Q., Fu, X.D., Shi, H.Y., Li, T.J., 2015. Watershed sediment dynamics and modeling: a watershed modeling system for Yellow River. In Yang C.T. and Wang L.K. (eds), *Advances in Water Resources Engineering, Handbook of Environmental Engineering, Volume 14, Chapter 1*, 1-40. Springer Cham Heidelberg New York Dordrecht London.
- Zhang, A., Li, T.J., Si, Y., Liu, R.H., Shi, H.Y., Li, X., Li, J.Y., Wu, X., 2016. Double-layer parallelization for hydrological model calibration on HPC systems. *Journal of Hydrology*, 535, 737-747.