Responses to comments

We, the authors of the manuscript, appreciate the valuable and constructive comments from Anonymous Referee #1. We will thoroughly revise the manuscript according to these comments. The detailed responses to the comments and questions are as follows.

General comments

The authors set out to establish a simple method for estimating the uncertainty of areal rainfall estimates on hydrologic simulations. While such approaches have been broadly considered (and are well cited in the introduction), the authors suggest that existing approaches lack an ability to be extrapolated to other places. Presumably the authors believe their approach addresses this, though that is not explicitly clear to the reviewer. Even if this were the case, unfortunately the modeling study suffers from fatal flaws that prohibit further interpretation of the results. Namely, the authors rely on running a 6-min simulation model with 2-hour forcing (precipitation). This causes multiple problems with model behavior (see specific comments). Additionally, the model clearly suffers from misspecification as a result of overfitting – demonstrating that the curve can be matched but at the expense of process fidelity.

[Answer] We would like to thank the anonymous 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.

Yes, this study aims to propose a general methodology that will not be limited by data, model and river basin. For example, the Qingjian River basin and Longxi River basin are selected as two different cases. With reference to the simulation results, we have to recognize that they are not very satisfying due to several reasons, especially the quality of the rainfall data. 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 results in the revised manuscript. Moreover, the overfitting problem will also be carefully investigated in the revision.

Details can be found in the following responses to the specific comments.

Specific comments

1. P3.L14-16: Is the argument for this study that a more straightforward methodology is needed? Or simply that this type of study needs to be performed at more sites than it has been to date?

[Answer] The objective of this study is to propose a more straightforward methodology. We will make this clearer in the closing part of the Introduction in the revised manuscript.

2. P3.L24-25: In L12-14 on this page, you argue that results are not reliably extrapolated to other locations, so can your study really accomplish this, as stated?

[Answer] Thanks for the referee's comment. It is worth noting that this study aims at a general method for managing rainfall station density rather than a uniform conclusion on the "representative area of rainfall stations". We believe that the proposed method can be used to extrapolate the results to other locations, e.g., in this study, the analysis of the Longxi River basin is an extension and supplement to that of the upstream Qingjian River basin. This indicates that the proposed method would not be limited by data, model and river basin; however, the results would be. In the revised manuscript, we will make the statement in a less strong tone.

3. P4.L30-31: Why use Thiessen? Why not use something more sophisticated?

[Answer] The calculation of the average basin rainfall needs to be performed hundreds of times in the bootstrap method, which makes the performability as the first factor to be considered. Thiessen polygon method is a classic method and can be easily achieved for batch loops with the ArcGIS. Furthermore, Thiessen polygon method is only used as one representative method to estimate the average basin rainfall, and other methods can also be selected to accomplish this. We can try more interpolation methods in the revision.

4. P6.L1-3: Why are you using 6-min time steps here, when DYRIM performance above (P5.L26-29) is noted as satisfactory for daily to monthly time steps?

[Answer] 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 of short duration in this study, the minimum time interval of the observed runoff data can be only 6 minutes. That is why we have used 6-min time step rather than the daily or monthly time step here. For hydrological simulations of longer durations, the time step will still be 6-min but the original discharge outputs will be used to calculate the daily or monthly discharge for comparison (e.g., Shi et al., 2015, 2016).

5. P6.L6-8: How many parameters are in DYRIM, how many need to be calibrated?

[Answer] The parameters of DYRIM can be divided into two types (Wang et al., 2015; Shi et al., 2016; Zhang et al., 2016). The first type includes invariant parameters that are used to describe the properties of the land use and soil types, including the field capacity of the topsoil and subsoil layers. The invariant parameters can be determined from field measurements and handbooks and have less influence on the simulated basin runoff. The other type includes all of the sensitive and adjustable parameters, which must be calibrated before model application using the observed rainfall and runoff data. Among them, the most sensitive parameters need to be calibrated are the vertical and lateral saturated conductivities (Table 2). The hillslope runoff-yield model and the main parameters are showed clearly in the following figure. Moreover, the parameters related to the river routing processes such as river manning coefficient should also be concerned.

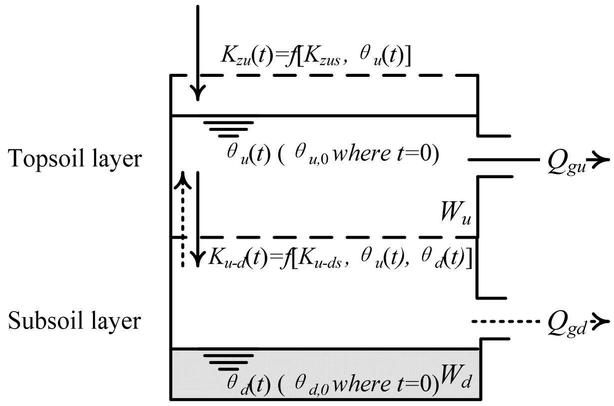

Figure. 1 The hillslope runoff-yield model and the main parameters that are used in DYRIM. In this figure, t is time, W_u is the water storage of the topsoil, Q_{gu} is the topsoil drainage, W_d is the water storage of the subsoil and Q_{gd} is the subsoil drainage. K_{zu} is the vertical conductivity of the topsoil layer, which is a function of K_{zus} (the vertical saturated conductivity of the topsoil layer) and $\theta_u(t)$ (the topsoil moisture). $K_{u\text{-}d}$ is the vertical conductivity between the topsoil and subsoil, which is a function of $K_{u\text{-}ds}$ (the vertical saturated conductivity between the topsoil and subsoil), $\theta_u(t)$ and $\theta_d(t)$ (the subsoil moisture) (From Zhang et al., 2016)

6. P6.L28-29: Why would you use this rather than something like SCE-UA or DDS, which are recognized to be the "best" global optimization algorithms for hydrological applications?

[Answer] Thanks for the referee's kind suggestion. Yes, for hydrological model optimization, SCE-UA and DDS are better than GA, and to use SCE-UA (or DDS) may make the results better. However, we have proposed a method for hydrological model calibration which is parallelized with a double-layer structure on HPC systems and it is proved to be valid (Zhang et al., 2016). This method has drawn the attentions of other researchers (e.g., Huang et al., 2016; Kuchar et al., 2016; Nourani and Partoviyan, 2017; Gelleszun et al., 2017). In this study, we just applied this method. It is worth noting that the two layers of parallelization are independent from each other, and then the upper layer is capable of incorporating other optimization algorithms, including SCE-UA (or DDS). Therefore, following the referee's kind suggestion, we may conduct further studies on evaluating the performances of this method by using other methods (e.g., SCE-UA or DDS) in our future work.

7. P9.L3-4: But, if you don't have 6-min forcing data, what does the 6-min runoff simulation get you? I am also concerned by the number of ad hoc adjustments being made in the model implementation (disaggregating rainfall data to smaller time steps, selecting the nearest observed runoff time for NSE evaluation, etc).

[Answer] As mentioned in the answer to Comment 4, the time step used in the DYRIM is determined by the temporal scale of the observed runoff data. We have used such fine time step (i.e., 6 minutes) in order to match the exact timing of the observed runoff data, which are instantaneously sampled and have about only

6-minute time steps during the main flooding processes.

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 (i.e., 2 hours) 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).

Moreover, 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. However, how to disaggregate the rainfall data in the stations without new data into the 6-minute time step will be a new problem.

In addition, we believe that selecting the simulated discharge time nearest to the observed runoff time for NSE evaluation is acceptable because fine time step (i.e., 6 minutes) is used in this study and the largest time error will be only 6 minutes. In fact, 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.

8. P9.L13-14: Exactly, you can't really simulated 6-min runoff with 2-hour rainfall. My question, then, is what value do these results have? The study would be stronger if the data fit the desired framework…

[Answer] 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. 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. However, how to disaggregate the rainfall data in the stations without new data into the 6-minute time step will be a new problem.

9. P9.L14-16: So why use Thiessen method?!

[Answer] We appreciate the referee's comment. Please see the answer to Comment 3 for details.

10. P9.L16-19: This is a clear flaw in the study design. Any interpretation of these results will naturally be influenced by this.

[Answer] Yes, we have to recognize that this will influence the simulation results to some extent because short-duration and high-intensity rains may be homogenised. However, such negative impacts are mainly from the aspect of temporal uncertainty rather than spatial uncertainty, and this study aims to estimate the influence of rainfall spatial uncertainty in hydrological simulations. The rainfall temporal uncertainty in hydrological simulations may be further investigated in our future work.

11. P9.L21-22: Why should the vertical saturated conductivity of the topsoil change so much? Shouldn't this value theoretically be constant? Aren't these changes

representative of compensation of other errors in the modeling?

[Answer] Thanks for the constructive comment. The vertical saturated conductivity of the topsoil layer (K_{zus}) controls the surface infiltration rate and primarily influences the infiltration-excess runoff. For example, surface roughness and vegetation of a hillslope affect the residence time and infiltration rate of water on its surface, and finally influence the total amount of infiltration. Yes, the vertical saturated conductivity of the topsoil should not change so much for a river basin. In this study, we first calibrate this parameter based on the observed data of all the three selected events and find that the simulation results are not so good. Therefore, we further calibrate this parameter independently for the three selected events in order to demonstrate the potential performance of the DYRIM. You may see that the simulation results using the same vertical saturated conductivity of the topsoil for the three selected events are also listed in Table 1.

12. P9.L25-26: But it doesn't prove that it achieves this performance for an appropriate reason. This is just curve fitting, isn't it?

[Answer] Yes, we agree that it is curve fitting, which is a way for evaluating the performance of the model. The results show that the topsoil vertical saturated conductivity K_{zus} is sensitive in this case, and using the same K_{zus} value may make the model performance unsatisfactory.

13. P10.L6: How would you utilize these results to improve rainfall gauging density or placement? Can you predict these locations without already having measured rainfall? If not, then what is the benefit?

[Answer] We appreciate the referee's insightful comment. We have to recognize that locations of rainfall stations may not be predicted or determined from this method without the measured rainfall data. However, this study aims at a general method for managing rainfall station density, which would not be limited by data, model and river basin. Moreover, the proposed method can give the suggestion of the proper D_S in a river basin using the proposed fitting of Equation (2). For example, this study indicates that the controlling area of each rainfall station should be about 40 km^2 for hydrological simulation at the hourly time scale in the middle Yellow River basin (see Section 4.5). However, a uniform conclusion on the "representative area of rainfall stations" cannot be obtained by using this method because the results would be influenced by rainfall patterns in different river basins.

Technical corrections

1. P2.1-2: "more" reliable…

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

2. P3.3: Change "good fit" to "goodness of fit"

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

References:

Gelleszun, M., Kreye, P., Meon, G., 2017. Representative parameter estimation for hydrological models using a lexicographic calibration strategy. Journal of Hydrology, doi: 10.1016/j.jhydrol.2017.08.015.

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.

Kuchar, S., Podhoranyi, M., Vavrik, R., Portero, A., 2016. Dynamic computing resource allocation in online flood monitoring and prediction. IOP Conference Series: Earth and Environmental Science, 39, 012061.

Nourani, V., Partoviyan, A., 2017. Hybrid denoising-jittering data pre-processing approach to enhance multi-step-ahead rainfall–runoff modeling. Stochastic Environmental Research and Risk Assessment, doi: 10.1007/s00477-017-1400-5.

Shi, H.Y., Li, T.J., Liu, R.H., Chen, J., Li, J.Y., Zhang, A., Wang, G.Q., 2015. A service-oriented architecture for ensemble flood forecast from numerical weather prediction. Journal of Hydrology, 527, 933-942.

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.

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.